Investigating the Structure-Function Relationships Between Flagellins and Toll-Like Receptor 5

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Michael En Wei (恩惠) Bell

aus Katherine, Northern Territory, Australien

Tübingen

2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	19.07.2024
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatterin	Prof. Dr. Ruth E. Ley
2. Berichterstatter	Prof. Dr. Thilo Stehle

Acknowledgements

I would not have been capable of completing this thesis without substantial support from a great number of people, who have helped me scientifically, professionally, and personally. I cannot name everyone, but I hope those who are unnamed know that I appreciate them.

First and foremost, I give my heartfelt thanks to Ruth Ley for allowing me the opportunity to undertake this thesis. I joined as a starry-eyed 23-year-old, with ideas for a project in a field that both I and the lab had little practical background in. The freedom and constant support that she provided were invaluable for me to build the confidence and experience needed to become, may I say, a competent biochemist. My TAC committee (Ruth Ley, Thilo Stehle, Andrei Lupas, John Weir) were a bulwark of support. When I was doubtful, their positivity and advice kept me motivated and on-track. Their recommendations were always fair and insightful. I thank them all, and I hope my work is something that they have looked upon fondly. I would particularly like to thank John, who endured me intermittently knocking on his door and complicating his day without complaint. His feedback on almost all aspects of my project was thorough, and unconditional.

I thank Sara Clasen, with whom I worked extensively. She always gave valuable advice and feedback, and I hope to one-day match her work-ethic. Felipe Merino gave me far more assistance than I was entitled to. I cannot thank him enough, as without him my skills in cryo-EM data processing, structure building, structure analysis would not be close to my current ability, and much of my dissertation would not have been possible.

My students Miriam Haag and Marieke Böcker were both fantastic and effective. Kateryna (Katya) Maksymenko beat me to graduating, but still had time to help with my SPR work. Veronika Altmannova taught me to manage insect cells and optimise my proteins. Katharina Hipp and Iris Koch helped with negative staining and Cryo-EM imaging. Jan-Hannes Schäfer and Arne Möller, from the Universität Osnabrück, have been great collaborators to work with. Lastly, Marcus Hartmann and Andrei Lupas were a great help in analysing and interpreting structural data. I thank all of them for their help and support.

Without Karin Klein, Ursula Schach, and Heike Budde I would have had far more administrative migraines. I hope they didn't have too many on my behalf. Similar thanks must go to the Researcher Support Team and George Deffner, without whom I would be lost under a pile of German bureaucracy, and very probably homeless.

I thank Silke Dauser for the German practice and making the lab far more enjoyable to work in. YiHua Liu, thank you for keeping me positive, you're impact on my mobility cannot be understated. I thank Anna Bryson, Jan Kisker, Meghna Chatterjee, Christian Kirk, and others in the lab, for both distracting me or leaving me be, as I needed. I want to thank Jason and Edmund for inspiring me. My batchmates Veysi Piskobulu and Anupam Gautam have been constant friends, and I'm thankful to have had them alongside me throughout. My office was a fun place to be thanks to Isabella Casini, YiHua Liu, XiaoYing Liu, Mirabeau Mbong Ngwese, and Sofia Esquivel. Linda Chen, Saskia Funk, Julia Hagauer, Volker Soltys, and Magda Firlej unofficially adopted me into the FML, which was my safe harbour during stressful periods, so I thank them and their couch.

I have had a fantastic network of friends away from work who have kept me going. Heidi Lai has been a true friend. I thank her for making the effort despite never seeing me, and for being there for me when I've needed it. Adrian Weghofer, David Maulat, and my other friends from TSG Tübingen volleyball H1 have made my time in Tübingen much more enjoyable; it wouldn't have been the same without them.

I want to thank Annapurna Interactive and Daniel Keyes for giving me some much-needed perspective at an important time, and The Weekly Planet for keeping me constantly entertained.

Finally, I want to thank my family. Gemma Pitotti was there for me through thick and thin. She has kept me going through the rough patches, and I cannot thank her enough for her love, support, and culinary prowess. Thomas Bell should have charged me for his advice; he's been more helpful than he realises. I thank him for giving me confidence in my writing. I wouldn't be here without my Mum or Dad, and thank them for supporting my decisions. Lastly, PorPor and KongKong have always loved me unconditionally, and I'm truly sorry that I didn't finish soon enough for both of them to see this. I hope that I have made them all proud.

Table of Contents

Common Abbreviations

Cryo-EM	Cryoelectronic Microscopy
ECD	Extra-cellular domain
IBD	Irritable Bowel Disease
IRAK	Interleukin-1 receptor associated kinase
LRR	Leucine Rich Repeat
MAMP	Microbe Associated Molecular Pattern
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NLR	NOD-like Receptor
PIM	Primary Interface Mutant of bacterial flagellin
ΔD0	Flagellin mutant truncated to remove the D0 domain
PRR	Pattern Recognition Receptor
*Rh*FlaB	*Roseburia hominis* FlaB
*St*FliC	*Salmonella typhimurium* FliC
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SPR	Surface Plasmon Resonance
TIR	Toll/interleukin-1 receptor homology domain
TLR	Toll-Like Receptor
TLR5	Toll-Like Receptor 5
$TLR5^{N14}$	Truncated TLR5 containing the first 14 LRRs
TLR5-FL	Full-length TLR5
TLR5-ECD	Complete TLR5 ectodomain
TM	TLR5 Trans-Membrane Domain
WT	Wild-Type flagellin

Summary

The gut microbiome, a complex community of trillions of archaea, fungi, viruses, and bacteria in the gastrointestinal tract, has a well-established role in host health and homeostasis. The balance of microbial species plays a crucial role in disease and pathogenicity, making understanding host-microbe interactions of great therapeutic interest. Toll-like receptor 5 (TLR5), is an innate immune receptor, which specifically responds to bacterial flagellin. TLR5 overactivity is linked to pro-inflammatory diseases like IBD, while disrupted activity increases susceptibility to pathogens. This dissertation explores how flagellins from the microbiome modulate TLR5 responses, providing insights essential for developing targeted therapeutics.

Through a screen of interactions between TLR5 and flagellins from common commensal bacteria, we identified "silent" flagellins, primarily from the *Lachnospiraceae* family, which bind to TLR5 but elicit a weak immune response, suggesting a widespread mechanism of immune avoidance. TLR5 signalling was modulated by interactions at the flagellin D1 and D0 domains, with the latter containing an allosteric binding site thought to induce conformational change in TLR5. A novel kinetic analysis revealed distinct binding profiles between the D1 domain binding $TLR5^{N14}$ and the stimulatory *S. typhimurium* FliC (*St*FliC) and silent *R. hominis* FlaB (*Rh*FlaB) flagellins. *St*FliC formed longer-lasting complexes despite a weaker binding strength and association rate. *Rh*FlaB relied solely on a well-defined TLR5 epitope for binding, while *St*FliC association instead depended on a secondary interface. A novel structure of *Rh*FlaB revealed distinct surface characteristics, absent from *St*FliC, which were likely causal to its unique kinetic binding profile.

I further explored the role of conformational change in TLR5 signalling, and produced two TLR5-TIR dimers capable of partially activating the TLR5 signalling pathway, though less effectively than *St*FliC, which warrant further investigation. I also produced and purified high concentrations of functional recombinant full-length TLR5 and TLR5 ectodomain constructs for structural analyses. Preliminary negative staining and Cryo-EM identified TLR5 monomers in the expected conformation, but high-resolution density maps have not yet been resolved.

My work reveals how microbes modulate TLR5 immune responses and the structural mechanisms affecting binding kinetics, and lays a solid foundation for further research. These findings underscore the crucial role of microbial interactions in host health and homeostasis, and highlight the significant scientific and medical benefits of continued research in this field.

Zusammenfassung

Das Darmmikrobiom, eine komplexe Gemeinschaft von Billionen von Archaeen, Pilzen, Viren und Bakterien im Magen-Darm-Trakt, spielt eine bedeutende Rolle für die Gesundheit und die Homöostase des Wirts. Das Gleichgewicht der mikrobiellen Spezies ist relevant bei Krankheiten und Pathogenität, so dass das Verständnis der Interaktionen zwischen Wirt und Mikroben von großem therapeutischem Interesse ist. Der Toll-like Rezeptor 5 (TLR5) ist ein angeborener Immunrezeptor, der spezifisch auf bakterielles Flagellin reagiert. Eine Überaktivität von TLR5 wird mit entzündungsfördernden Krankheiten wie IBD in Verbindung gebracht, während eine gestörte Aktivität die Anfälligkeit für Krankheitserreger erhöht. In dieser Dissertation wird untersucht, wie Flagelline aus dem Mikrobiom die TLR5-Reaktionen steuern und damit wichtige Erkenntnisse für die Entwicklung gezielter Therapeutika liefern.

Durch ein Screening der Interaktionen zwischen TLR5 und Flagellinen aus verschiedenen weit verbreiteten kommensalen Bakterien haben wir „stille“ Flagelline identifiziert, vor allem aus der Familie der *Lachnospiraceae*. Diese binden an TLR5, lösen aber nur eine schwache Immunantwort aus, was auf einen weit verbreiteten Mechanismus der Vermeidung einer Immunantwort hinweist. Die TLR5-Signalwirkung wird durch Interaktionen an den Flagellin-D1- und -D0-Domänen gesteuert, wobei die letztere eine allosterische Bindungsstelle enthält, von der man annimmt, dass sie eine Konformationsänderung in TLR5 auslöst. Eine neuartige kinetische Analyse zeigte unterschiedliche Bindungsprofile zwischen der D1-Domäne, von den stimulierenden Flagellinen *S. typhimurium* FliC (*St*FliC) und silent *R. hominis* FlaB (*Rh*FlaB), und $TLR5^{N14}$. *St*FliC bildete trotz einer schwächeren Bindungsstärke und Assoziationsrate länger anhaltende Komplexe. Für die Bindung von *Rh*FlaB war ausschließlich ein gut definiertes TLR5-Epitop nötig, während die Assoziation von *St*FliC zusätzlich von einer sekundären Schnittstelle abhing. Eine neuartige Struktur von *Rh*FlaB zeigte deutliche Oberflächenmerkmale, die bei *St*FliC fehlten und die wahrscheinlich für das einzigartige kinetische Bindungsprofil verantwortlich waren.

Um die Rolle der Konformationsänderung bei der TLR5-Signalübermittlung weiter untersuchen zu können, habe ich zwei TLR5-TIR-Dimere hergestellt, die den TLR5-Signalweg teilweise aktivieren können, wenn auch weniger effektiv als *St*FliC, was weitere Untersuchungen rechtfertigt. Außerdem habe ich rekombinante TLR5- und TLR5-Ektodomänen-Konstrukte in

voller Länge für Strukturanalysen hergestellt und in hohen Konzentrationen aufgereinigt. Vorläufige negativ Färbungen und Cryo-EM Bilder, identifizierten TLR5 Monomere in der erwarteten Konformation, aber hochauflösende Dichtekarten sind noch nicht erstellt worden.

Meine Arbeit zeigt, wie Mikroben die Immunantwort von TLR5 modulieren und welche strukturellen Mechanismen die Bindungskinetik beeinflussen. Diese Ergebnisse unterstreichen die entscheidende Rolle der mikrobiellen Interaktion für die Gesundheit und die Homöostase des Wirts und verdeutlichen den bedeutenden wissenschaftlichen und medizinischen Nutzen der weiteren Forschung auf diesem Gebiet.

Chapter 1: General Introduction

1.1. The gut microbiome

The human body is host to a vast number of microorganisms, collectively referred to as the microbiome. These microbial communities, encompassing bacteria, viruses, fungi, and archaea, cohabit with human cells at an approximate 1:1 ratio [1,2]. The microbiome plays a pivotal role in human health, influencing a myriad of physiological processes, including digestion, metabolism, immune function, and even behaviour [3–5]. The gastrointestinal tract contains the highest abundance and diversity of these microbes, particularly within the large intestine. This community forms what is known as the gut microbiome [6–8]. The vast majority of the trillions of microorganisms which make up this community come from the bacterial phyla *Bacteroidetes* and *Firmicutes* [6,9,10].

The composition of the gut microbiome has direct impacts on gut homeostasis and overall health, directly affecting metabolism, nutrient absorption, the synthesis of essential vitamins, and immune responses against pathogenic bacteria. This composition is highly variable and dynamic, subject to numerous influences including diet, age, genetics, and environment [11–16]. Understanding how microbes interact with the host is therefore of great importance in advancing our understanding of human health, preventing and treating diseases, and developing new therapeutic strategies to maintain homeostasis [16,17].

1.2. Microbes and the immune system

Advancements in large-scale sequencing, mass-spectrometry, and structural analyses have facilitated a substantial increase in research exploring host-microbe interactions [18,19]. Such interactions primarily occur through the innate and adaptive immune systems. The innate immune system is the first line of defence against pathogens, consisting of physical barriers such as the skin and mucous membranes, and a network of cellular receptors. The adaptive immune system provides more specialised and long-lasting responses, acting primarily through lymphocytes such as T and B cells [20–23]. Cross-talk between the innate and adaptive immune systems results in a robust response, with cytokines produced by the former causing the differentiation and proliferation of adaptive immune T and B cells [24,25]. Consequently, initial interactions between microbes and the innate immune system are essential in health outcomes.

Microbes interact with the innate immune system through the production of numerous products, including metabolites, proteins, and lipids. These products directly target cellular receptors and instigate a pro-inflammatory immune response, promoting a microbial ecosystem beneficial to their survival. The innate immune system works conversely, acting against specific microbes in the microbiome to create an environment suitable for beneficial symbionts and hostile to potential pathogens. These interactions ultimately allow microbes and the host to maintain gut homeostasis, which is crucial for host health and metabolism [18,19,26].

In vertebrates, the innate immune system relies on pattern recognition receptors (PRRs) including the Toll-like receptor (TLR) and nucleotide-binding oligomerization domain (NOD)-like receptor (NLR) families [19,22,27]. Both respond to microbe-associated molecular patterns (MAMPs), which are highly conserved sequences or motifs in bacterial products which are essential for microbial fitness and pathogenicity. The conservation of such MAMPs between species allows the innate immune system to recognize and act against a broad spectrum of microbes. Notable examples of these are lipopolysaccharides (LPS) from Gram-negative bacteria, peptidoglycan from Gram-positive bacteria, and flagellin from motile bacteria [28–31].

1.3. Toll-Like receptors

TLRs are the first discovered, largest, and most well-studied family of innate immune receptors, and are conserved throughout the animal kingdom [32]. These Type-1 transmembrane proteins are widely expressed in a variety of cells, including phagocytes and epithelial cells. Most TLRs are expressed at the plasma membrane, and detect microbial products in the cell's local environment, while the rest are expressed in vesicles, and detect bacterial infiltration [33–35].

All TLRs contain an MAMP-recognising leucine-rich repeat (LRR) domain, an anchoring transmembrane (TM) domain, and an intracellular Toll/IL-1 receptor (TIR) domain responsible for activating pro-inflammatory signalling pathways (Fig. 1.1) [34,36]. Through these pathways, TLRs play a critical role in maintaining gut homeostasis. Imbalances in TLR activity can cause dysbiosis, which has implications for diseases such as inflammatory bowel disease (IBD) and cancer [19,37–39]. Consequently, TLRs are promising targets for therapeutic interventions [32].

TLR signalling is initiated when ligands bind TLR dimers, which either facilitate dimerisation or induce conformational changes, allowing the TLR5-TIR dimer to recruit adaptor proteins such as MyD88 or TRIF [40–42]. These assemble members of the Interleukin-1 receptor associated kinase (IRAK) family, generally IRAK-1/2 and IRAK-4, forming a complex known as the Myddosome [42,43]. Several subsequent downstream pathways can then be activated, depending on the specific TLR, which ultimately lead to the production of transcription factors, including nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), which instigate pro-inflammatory cytokine or type-1 interferon responses [44,45].

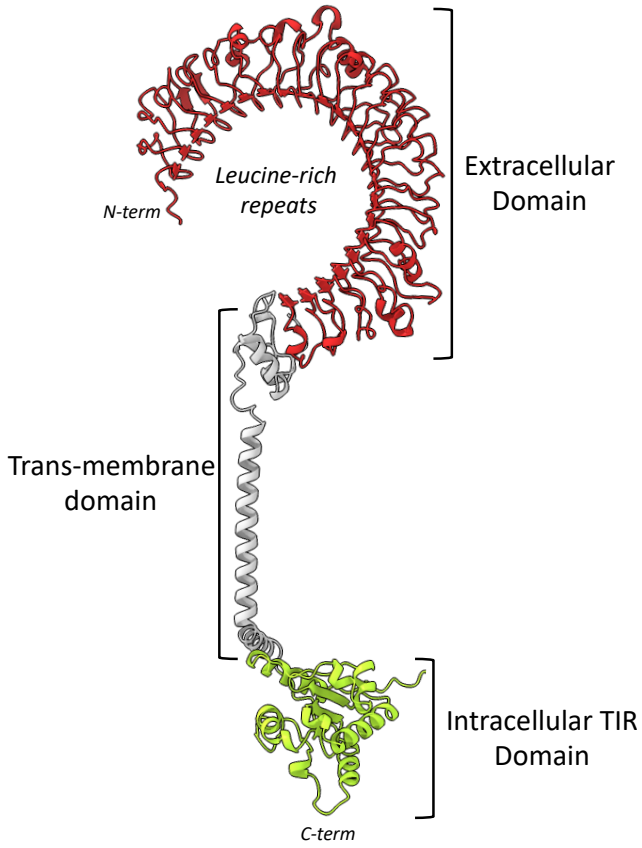


Figure 1.1 – Annotated structural model of Toll-Like receptor 5.
Toll-Like receptors have a conserved structural profile. The MAMP-binding extracellular domain, composed of leucine rich repeats, is generally located at the cell surface. It is linked by a trans-membrane α-helix to the intracellular TIR domain, which recruits MyD88 and triggers the TLR signalling pathway. The model was generated with Alphafold2 [105].

1.4. Toll-Like receptor 5 and flagellin

The Toll-like receptor 5 (TLR5) has come into focus due to its medical relevance. TLR5 overactivity can cause chronic inflammation, impacting conditions including IBD, arthritis, and cancer [46–50]. Conversely, mutations in TLR5 can impair immune responses, increasing susceptibility to infections [51–53]. Understanding the mechanisms essential to TLR5 signalling are therefore critical in understanding its role in disease, and developing targeted therapies against TLR5.

The TLR5 signalling pathway has been well characterised (Fig. 1.2). TLR5 specifically recognises and responds to highly conserved MAMPs in bacterial flagellin. Once bound, flagellin-TLR5 heterodimer complexes are able to recruit MyD88, leading to the formation of the Myddosome complex with IRAK4 and 1. The Myddosome interacts with TRAF6 and TAK1, stimulating the IkB and MAP kinase pathways, which respectively induce the production of NF-κB and AP-1 transcription factors, and initiate a pro-inflammatory cytokine response [54–57]. Exactly how the TLR5-flagellin complex initiates signalling is unclear, though it is often implied in the literature that flagellin binding is essential for TLR5 hetero-dimerisation, thus facilitating MyD88 recruitment [58–61]. However, the existence of unliganded asymmetric TLR5 dimers challenges this assumption. Instead, ligand-induced conformational change, as seen in TLR9, likely facilitates MyD88 recruitment by TLR5-flagellin heterodimers [41,62]. A more complete understanding of TLR5-flagellin binding is essential for clarifying this mechanism.

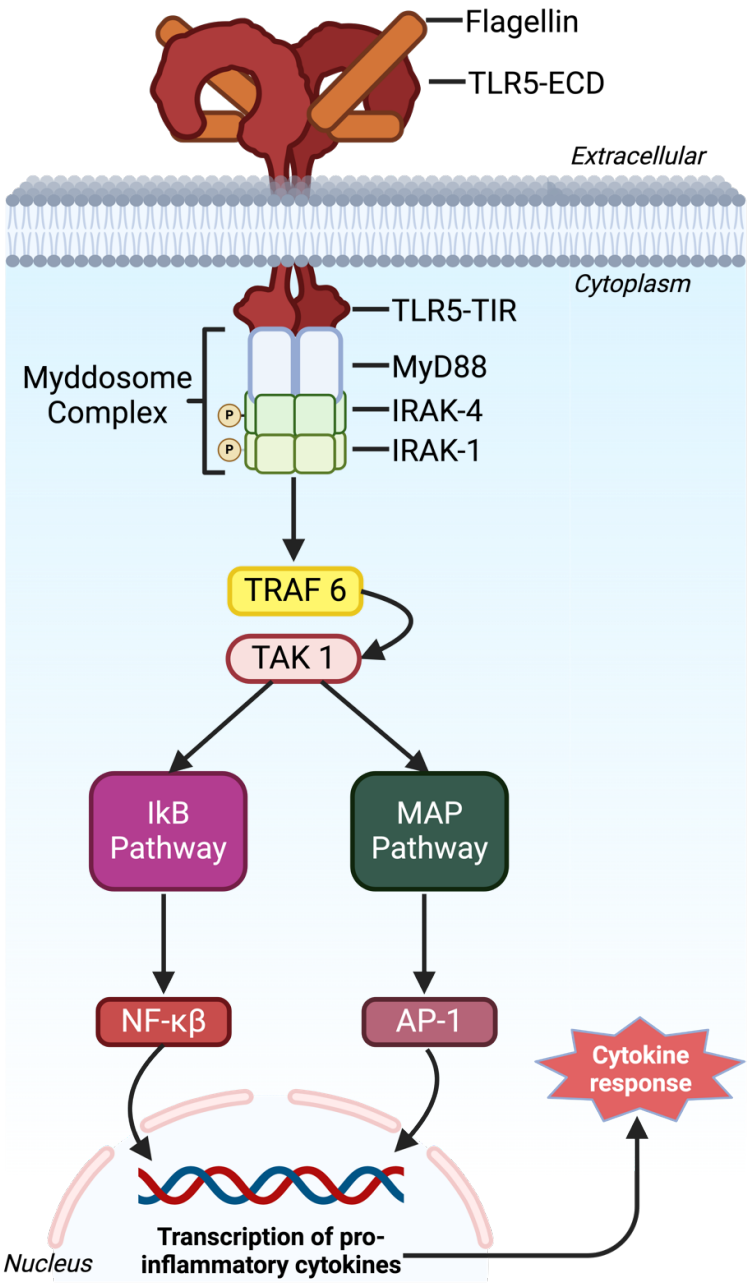


Figure 1.2 – The Toll-like receptor 5 signalling pathway.

TLR5 signalling is initiated by flagellin binding at the extracellular domain (TLR5-ECD). MyD88 is recruited, which forms the Myddosome complex. Through TRAF6 and TAK1, the IkB and MAP kinase pathways are triggered, activating the NF-κB and AP-1 pathways and promoting cytokine production [54-57]. This figure was created using BioRender.

Flagellins are the principal protein component of the bacterial flagellum, an organelle essential for swimming motility and chemotaxis. Flagella also play roles in adhesion, biofilm formation, and are known to be essential for pathogenesis in certain species [63–65]. Accordingly, they are widely expressed throughout a diverse range of bacteria, and approximately half of flagellated bacteria possess two or more flagellin genes within their genome [66–68]. The diversity between bacteria has shown benefits for evading innate immune recognition, with flagellin heterogeneity limiting the specificity of TLR5 binding [69]. Flagellin diversity within single bacteria improves bacterial motility and adaptive immune evasion through mechanisms such as phase variation, where such bacteria alternately express different flagellin genes [70,71].

Despite the substantial sequence variation between flagellins, the overall structure is remarkably well conserved (Fig. 1.3). Flagellins contain two highly conserved domains (D0 and D1), and a number of hypervariable domains. The D0 and D1 domains are primarily composed of alpha-helices, and polymerise to form the core of the bacterial flagella. The hypervariable domains primarily contain β-sheets, and comprise the flagella surface [67,72–74]. The hypervariable domains have roles in adaptive immune avoidance and motility, and may interact with TLR5 in specific cases [73,75,76]. However, TLR5 primarily targets highly conserved MAMPs in the D0 and D1 domains, which are essential for flagellin polymerisation, and consequently flagellum formation and motility.

As these MAMPs are normally buried in the core of the flagellum, TLR5 can only bind monomeric flagellin that has been ‘shed’ from the flagellum complex [58,77]. Some bacteria, including *Helicobacter* and *Vibrio* species, generate a protective ‘sheath’ around the flagellum to prevent flagellin shedding and reduce MAMP detection [78,79]. Conversely, pathogens including *Salmonella typhimurium* and *Pseudomonas aeruginosa* intentionally shed flagellins under certain conditions, provoking an immune response which disrupts the gut epithelium and facilitates invasion [80,81].

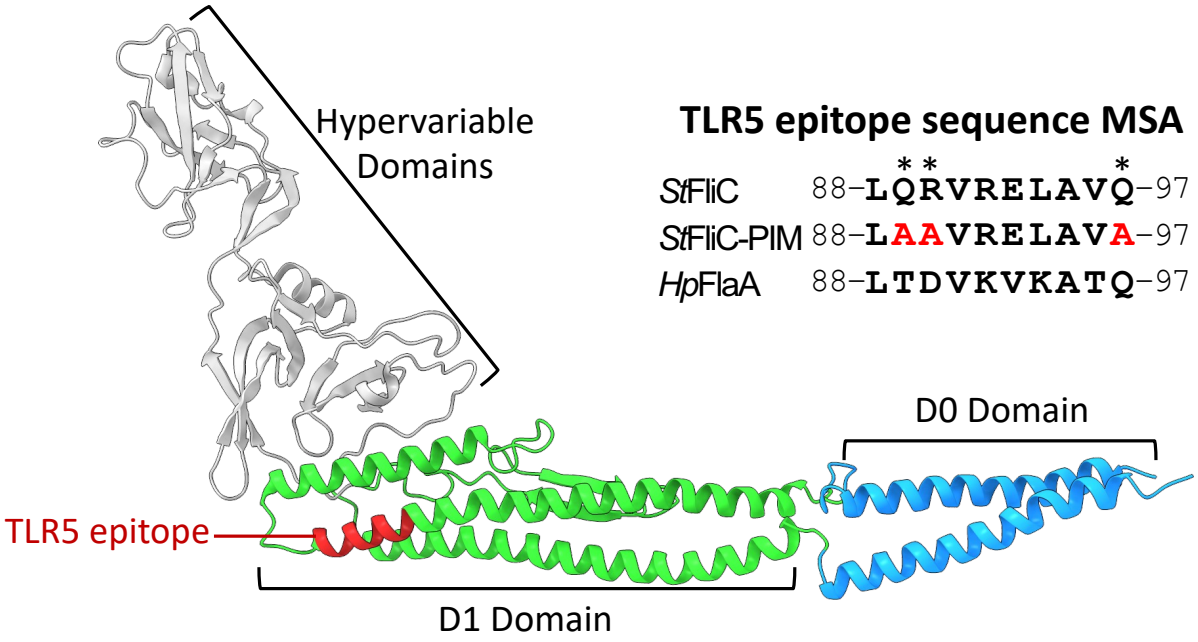


Figure 1.3 – The conserved structure of bacterial flagellin.
The Cryo-EM structure of *S. typhimurium* FliC (*St*FliC; PDB ID #1UCU), annotated to display the shared domain layout, and the highly conserved TLR5 epitope [106]. A multiple sequence alignment of the TLR5 epitope in *St*FliC and the evader *H. pylori* FlaA, made with Clustal O (1.2.4), is also shown. The three alanine substitutions made in PIM mutants are shown in red, and indicated by '*'.

1.5. Experimental studies on TLR5-flagellin interactions

Mutagenic studies used to identify flagellin MAMPs have primarily explored flagellins of pathogenic origin, including the canonical TLR5 agonist *S. typhimurium* FliC (*St*FliC). In *St*FliC, a set of conserved residues known as the TLR5 epitope were identified (Fig. 1.3). The TLR5 epitope is considered essential for TLR5 binding and signalling, as primary interface mutants (PIM), which substitute three key residues within this epitope (*St*FliC: Q89A, R90A, Q97A), display a dramatically reduced TLR5 response [77,82,83]. The TLR5 epitope is notably absent from TLR5 evading flagellins, such as *Helicobacter pylori* and *Campylobacter jejuni*, which have extensively modified their flagellin sequences and flagellum structures to evade TLR5 detection [84,85]. More recent studies have identified additional binding residues in the flagellin D0 domain, which are predicted to bind the C-terminal end of the TLR5 ectodomain [58].

The bulk of structural insights into TLR5-flagellin binding come from a crystal structure of zebrafish TLR5b (zTLR5b) and truncated FliC from *Salmonella enterica* subspecies *enterica* serovar Dublin (*Sd*FliC) in a 2:2 complex. This study identified two D1 domain binding interfaces, which were involved in both 1:1 and 2:2 complex stability. 'Primary interface-A' consisted of residues in the central cD1-domain, which interacted with TLR5 LRR1-6. 'Primary interface-B' contained the TLR5 epitope, and primarily bound a binding pocket in the highly conserved TLR5 LRR9 loop, inducing a conformational change [40,86,87]. The three residues mutated in 'PIM' mutants were shown to form hydrogen bonds in this binding pocket in both *Sd*FliC and *Bacillus subtilis* flagellins, emphasising their importance in TLR5-flagellin complex formation and stability [40,86].

A notable caveat of the structural data is the use of zebrafish-derived TLR5. TLR5 structure and function are dependent on their species-specific context. Chimeric TLR5s with domains from different species are less capable of responding to infection [88,89]. Zebrafish specifically uniquely expresses two TLR5 proteins which must heterodimerise to signal, but corresponding modifications impair the functionality of human TLR5 [90]. Finally, differences in TLR5 response to specific flagellins have been shown in human, mouse, chicken, and bovine contexts [82,89,91–94]. This species-specificity is unsurprising given the host-specificity of bacterial pathogens, but does limit reliability of transferring such findings to a human context [95,96].

1.6. Scope of the dissertation

Despite the extensive research described, significant gaps in our understanding of TLR5-flagellin interactions remain. Interactions between TLR5 and commensal bacteria, which generally contain 'conventional' *St*FliC-like flagellins, are poorly studied. How such species persist in the gut without triggering a TLR5 immune response is unknown. The mechanisms behind TLR5-flagellin complex formation, especially the dynamics of flagellin binding at both D1 and the D0 domains, are still largely unclear. Finally, to fully understand the TLR5 signalling pathway, it is essential to clarify if and how flagellin-induced conformational changes in TLR5 facilitate MyD88 recruitment at the TIR domain. In this dissertation, I describe four separate projects which aim to address these open questions.

My first project (Chapter 2) investigates TLR5 interactions with abundant flagellins in the gut microbiome through a large-scale binding and activity screen. We identified a new category of interactions termed 'silent' recognition, characterised by strong TLR5 binding but a weak immune response. This was caused by a specific interaction between TLR5 and flagellin at the D0 domain, which was present in *St*FliC but absent in the representative silent flagellin *Roseburia hominis* FlaB (*Rh*FlaB). This interaction facilitated flagellin binding to preformed TLR5 dimers, significantly enhancing TLR5 signalling. Silent flagellins were primarily found in the phyla *Lachnospiraceae*, and silent recognition was observed in both human organoids and mice. These findings provide a mechanism by which the innate immune system can tolerate commensal flagellins in the gut microbiome, while remaining vigilant against pathogens.

The second project (Chapter 3) explores how D1 domain binding facilitates silent interactions. Using Surface Plasmon Resonance (SPR), the binding kinetics of interactions between a truncated flagellin D1-domain binding TLR5 construct ($TLR5^{N14}$), and the stimulatory *St*FliC or silent *Rh*FlaB flagellins were examined. Despite a stronger overall binding strength and comparatively rapid association rate, *Rh*FlaB only formed short-lived complexes when compared to *St*FliC. Distinct roles for the PIM residues were identified, which were essential for association in *Rh*FlaB but only required for stability in *St*FliC. The structure of *Rh*FlaB was solved and compared to *St*FliC, revealing distinct surface characteristics which were likely causal to *Rh*FlaB's kinetic binding profile. These findings evidence a novel temporal mechanism by which silent flagellins could evade immune detection, and alongside the structural insights have potential applications in the design of agonists and antagonists targeting TLR5.

My third project (Chapter 4) examines the rotational specificity of the TLR5-TIR, to clarify how conformational changes in the extracellular domain (ECD) could permit MyD88 recruitment at the TIR. I generated artificial TIR dimers at seven distinct rotational orientations, which were transfected into HEK 293 reporter cells to observe their impact on the TLR5 signalling pathway. No construct triggered a strong TLR5 response when compared against a positive control stimulated with *St*FliC, but two displayed a consistently higher signal than non-stimulated cells, potentially indicating partial activation of the signalling pathway. Troubleshooting experiments were performed to help guide future research. If an active TLR5-TIR dimer is identified, it would evidence the importance and specificity of conformational changes required for MyD88 recruitment, and TLR5 signalling.

Lastly, my fourth project (Chapter 5) tackles the problems of species-specificity and the absence of structural data covering binding at the flagellin D0 domain. I aimed to achieve this by generating complete human TLR5 and TLR5-flagellin complexes for Cryo-electron microscopy (Cryo-EM) analyses. Full-length TLR5 (TLR5-FL) and TLR5 ectodomain (TLR5-ECD) constructs were produced and purified at high concentration, and several assays verified their ability to bind *St*FliC. Negative stain and Cryo-EM imaging of the TLR5-ECD resolved 2D-classes in the expected conformation. However, 2D classes with clear evidence of TLR5-ECD and *St*FliC in complex were not apparent in either co-incubated or crosslinked samples. Clear plans for continued work are outlined. A complete structure TLR5 and the TLR5-flagellin complex would allow the study of binding interactions at the flagellin D1 and D0 domains in a human context, and potentially show evidence of flagellin-induced conformational changes in TLR5. These findings would be invaluable for designing competitive binders to both antagonise and stimulate TLR5.

Cumulatively, this dissertation tackles significant gaps in our understanding of the structural and functional dynamics underlying the TLR5 immune response. By advancing our knowledge of TLR5 signalling, this work broadens our understanding of other TLRs and the complex interactions between microbes and our innate immune system, which are essential for maintaining gut homeostasis. My hope is that this research will eventually pave the way for the development of novel treatments for inflammatory diseases, cancer, and other conditions linked to TLR5.

Chapter 2: Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5

This chapter is adapted from a paper published in Science Immunology:
https://doi.org/10.1126/sciimmunol.abq7001
Only changes to formatting have been made.

Authors are listed in publication order:
Sara J. Clasen[1], Michael E. W. Bell[1], Andrea Borbón[1], Du-Hwa Lee[2], Zachariah M. Henseler[1], Jacobo de la Cuesta-Zuluaga[1], Katarzyna Parys[2], Jun Zou[3], YanLing Wang[3], Veronika Altmannova[4], Nicholas D. Youngblut[1], John R. Weir[4], Andrew T. Gewirtz[3], Youssef Belkhadir[2], Ruth E. Ley[1,5].

[1] Department of Microbiome Science, Max Planck Institute for Biology, Tübingen 72076, Germany.
[2] Gregor Mendel Institute (GMI), Austrian Academy of Sciences, Vienna BioCenter (VBC), Dr. Bohr-Gasse 3, Vienna, Austria.
[3] Center for Inflammation, Immunity and Infection, Institute for Biomedical Sciences, Georgia State University, Atlanta, GA, USA.
[4] Friedrich Miescher Laboratory of the Max Planck Society, Max-Planck-Ring 9, Tübingen 72076, Germany.
[5] Cluster of Excellence EXC 2124 Controlling Microbes to Fight Infections, University of Tübingen, Tübingen, Germany.

Manuscript contributions: This project was conceptualised by myself, alongside SJC, ATG, YB, and REL. I designed the methodology for the comparative TLR5-flagellin screen based on work by ZH, DL, and YB. I produced all recombinant FC-tagged TLR5^{N14} used in this study. Alkaline-phosphatase tagged flagellins were produced by myself and DL. I performed the comparative TLR5-flagellin binding screen, and validated these results. I analysed these results alongside SJC, AB, JCZ, and REL. I wrote text included in the methods section of the original draft, and reviewed the original manuscript.

Abstract

Flagellin, the protein subunit of the bacterial flagellum, stimulates the innate immune receptor Toll-like receptor 5 (TLR5) after pattern recognition or evades TLR5 through lack of recognition. This binary response fails to explain the weak agonism of flagellins from commensal bacteria, raising the question of how TLR5 response is tuned. Here, we screened abundant flagellins present in metagenomes from human gut for both TLR5 recognition and activation and uncovered a class of flagellin-TLR5 interaction termed silent recognition. Silent flagellins were weak TLR5 agonists despite pattern recognition. Receptor activity was tuned by a TLR5-flagellin interaction distal to the site of pattern recognition that was present in *Salmonella* flagellin but absent in silent flagellins. This interaction enabled flagellin binding to preformed TLR5 dimers and increased TLR5 signalling by several orders of magnitude. Silent recognition by TLR5 occurred in human organoids and mice, and silent flagellin proteins were present in human stool. These flagellins were produced primarily by the abundant gut bacteria *Lachnospiraceae* and were enriched in non-industrialized populations. Our findings provide a mechanism for the innate immune system to tolerate commensal-derived flagellins while remaining vigilant to the presence of flagellins produced by pathogens.

2.1. Introduction

Innate immune responses are initiated by pattern recognition receptors (PRRs) that evolved to detect conserved microbe-associated molecular patterns (MAMPs) [1]. The Toll-like family of receptors (TLRs) are membrane-bound PRRs, widely expressed in many cell types, that activate pro-inflammatory pathways following MAMP-binding to their horseshoe-shaped ectodomains [2]. Since MAMPs are not unique to pathogens, a question that has persisted for decades is whether TLRs respond differently to ligands derived from beneficial or commensal microbiota, relative to those produced by potentially pathogenic microbes [3]. This question is especially relevant for TLRs that interface with the intestinal microbiota such as Toll-Like Receptor 5 (TLR5), which is highly expressed by epithelial cells that line mucosal surfaces [4].

TLR5 is plasma membrane bound and binds extracellular flagellin, the protein subunit of the bacterial flagellum [5]. Phylogenetically diverse bacteria produce structurally similar flagellins that consist of conserved N- and C-terminal D0-D1 domains separated by a hypervariable region [6]. The MAMP recognised by TLR5 is located in the N-terminal D1 (nD1) and referred to as the TLR5 epitope [7,8]. Studies on the FliC flagellin derived from the human pathogen *Salmonella enterica* serovar *Typhimurium* show that mutating key residues in this region (*St*FliC-PIM) reduces ligand potency and abolishes bacterial motility [7]; crystal structures of *St*FliC in complex with *Danio rerio* TLR5 confirm a direct interaction between these residues and leucine-rich repeat 9 (LRR9) in the N-terminal region of the receptor ectodomain [9,10]. Furthermore, flagellins that do not stimulate TLR5, like FlaA from the human pathogen *Helicobacter pylori* (*Hp*FlaA) have different amino acids in their TLR5 epitope site [8,11]. TLR5's inability to respond to *Hp*FlaA is characterized as "evasion" and is presumed to occur through loss of TLR5 binding. Together, these studies demonstrate that robust TLR5 signalling requires the receptor ectodomain to bind the flagellin TLR5 epitope.

Commensal bacteria produce flagellins with TLR5 epitopes identical to those of *St*FliC; however, these flagellins induce a range of TLR5 activity [12–14], raising the question of how the TLR5 response is tuned. Here, we addressed this question by investigating how TLR5 interacts with flagellins from commensal bacteria. We identified members of the *Lachnospiraceae* family as the major producers of flagellin in the human gut and measured TLR5 recognition and response to 40 of the most prevalent flagellins in the human gut. This approach led us to identify a type of flagellin that we termed "silent" due to the observation that these flagellins retained binding to TLR5 yet poorly activated TLR5 signalling. Last, we described how silent flagellins circumvented TLR5 stimulation, thereby providing additional insights into the mechanism of TLR5 activation.

2.2. Results

2.2.1. Commensal bacteria produced flagellins that decoupled TLR5 binding from activation

To examine the TLR5 response to commensal-derived flagellins, we first searched for flagellins commonly encoded by the human gut microbiome. Flagellin diversity was vast: Of the 10 million proteins encoded by the human gut microbiome, more than 5000 different proteins were classified as flagellins [15]. Most flagellin in the human gut is produced by *Lachnospiraceae* [16], a prevalent and abundant family of *Firmicutes* that includes beneficial bacteria such as the butyrate producers of the *Roseburia* and *Eubacterium* genera [17]. Selecting from the most abundant flagellins observed in 270 samples from individuals without inflammatory bowel disease (IBD) [18] (Supp. Fig. 2.1), we expressed an initial 40 recombinant flagellins (34 belonging to *Lachnospiraceae* species) and screened these for both TLR5 signalling and recognition (Supp. Figs. 2.2 and 2.3).

To quantify TLR5 recognition of flagellin, we measured the relative binding strength between the receptor and the TLR5 epitope (Fig. 2.1A, magenta region) using a truncated form of the human ectodomain, $TLR5^{N14}$ (similar to the one used in the crystal structure complex in [9]) (Fig. 2.1B). This construct contained the first 14 of the 22 LRRs that comprise the ectodomain, including the flagellin binding site identified in the crystal structure (LRR9), followed by a C-terminal adaptor sequence tagged to immunoglobulin G (IgG)–Fc. Binding was quantified by incubating $TLR5^{N14}$ with flagellins expressing C-terminal alkaline phosphatase (AP) and measuring AP activity. Consistent with its TLR5 epitope directly interacting with $TLR5^{N14}$ [9], *St*FliC bound strongly (Fig. 2.1C). In contrast, *St*FliC-PIM, which lacks three conserved residues in the epitope, showed a ninefold reduction in binding compared with *St*FliC. Similarly, *Hp*FlaA, with its altered TLR5 epitope, failed to bind $TLR5^{N14}$. The D0 domain of *St*FliC is unnecessary for binding to TLR5 [9,19], and we observed strong binding of *St*FliC-ΔD0 to $TLR5^{N14}$ (Fig. 2.1A,C). Together, these results showed that the flagellin-$TLR5^{N14}$ interaction required the TLR5 epitope and thus reflected pattern recognition by TLR5 [11].

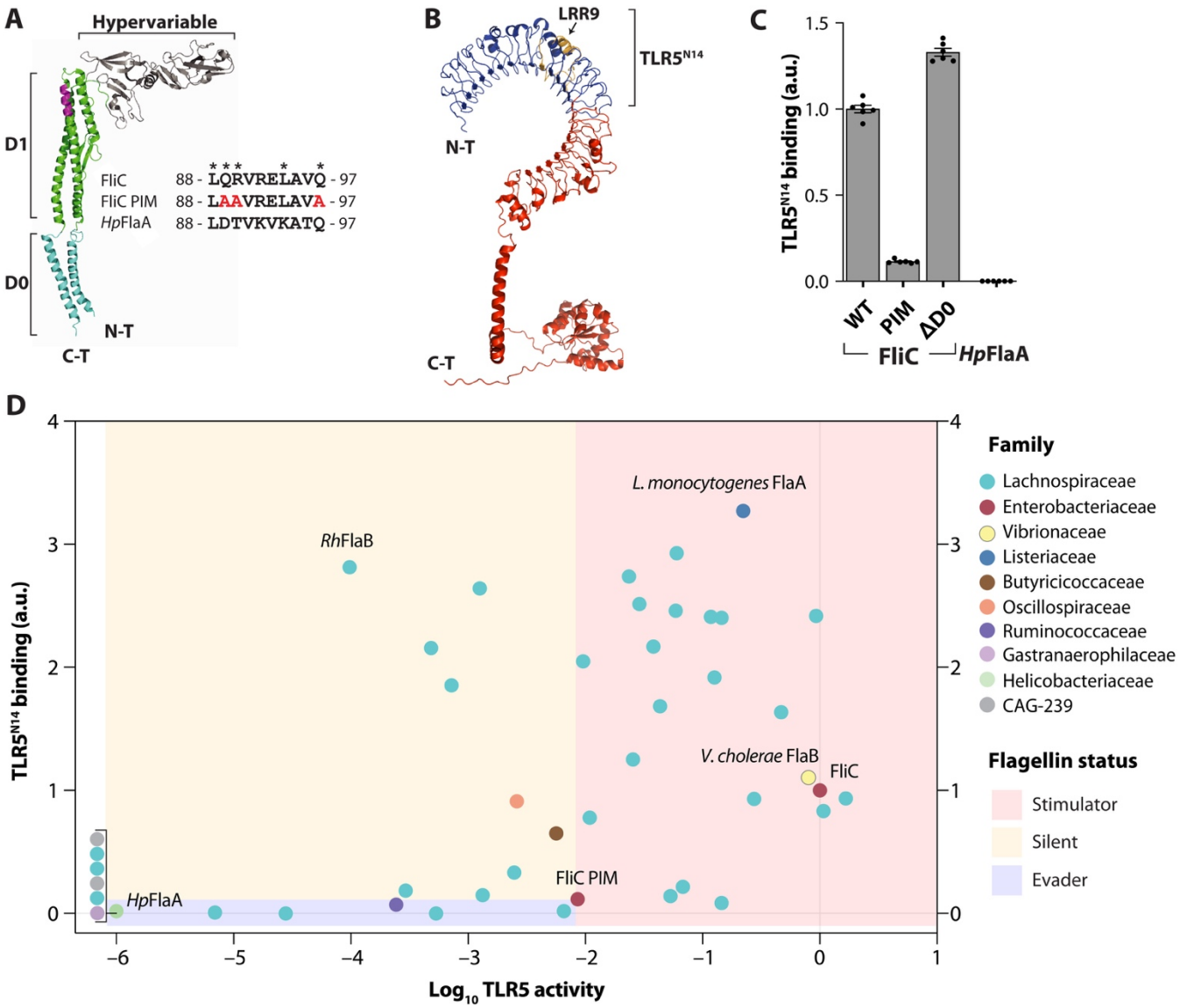

A
Hypervariable
D1
D0
N-T
C-T
FliC 88 - LQRVRELAVQ - 97
FliC PIM 88 - LAAVRELAVA - 97
HpFlaA 88 - LDTVKVKATQ - 97
B
LRR9
TLR5N14
N-T
C-T
C
TLR5N14 binding (a.u.)
1.5
1.0
0.5
0.0
WT
PIM
ΔD0
FliC
HpFlaA
D
TLR5N14 binding (a.u.)
Log10 TLR5 activity
L. monocytogenes FlaA
RhFlaB
V. cholerae FlaB
FliC
FliC PIM
HpFlaA
Family
Lachnospiraceae
Enterobacteriaceae
Vibrionaceae
Listeriaceae
Butyricicoccaceae
Oscillospiraceae
Ruminococcaceae
Gastranaerophilaceae
Helicobacteriaceae
CAG-239
Flagellin status
Stimulator
Silent
Evader

Figure 2.1 – Flagellins from human gut commensals were silently recognised by TLR5.

(A) *St*FliC crystal structure (PDB ID #3A5X from [6]) and multiple sequence alignment of nD1 TLR5 epitope (magenta) from *Salmonella* and *H. pylori* flagellins. Asterisks denote residues in *St*FliC required for TLR5 recognition; red: residues mutated in *St*FliC-PIM. FliC = *St*FliC (B) Predicted structure of human TLR5 [10]. Blue: Region present in $TLR5^{N14}$. Yellow: TLR5 epitope binding site, LRR9. (C) Flagellin binding to TLR5 ectodomain: absorbance units (a.u.) are relative to *St*FliC. Error bars are SEM for n = 6. (D) TLR5 activity versus $TLR5^{N14}$ binding for 40 flagellins abundant in gut microbiome (see Supp. Fig. 2.1) and from pathogens. Circles represent individual flagellins; colours indicate family-level taxonomy (GTDB). $TLR5^{N14}$ binding described in (B); data represent mean for n ≥ 2. TLR5 activity represents negative EC_{50} normalized to flagellin expression in lysates. Data represent mean from at least three independent experiments; values normalized to *St*FliC. Flagellin status is defined relative to *St*FliC-PIM: Stimulators are more active (EC_{50} < 10 nM), silent flagellins are less active (EC_{50} > 10 nM) with increased binding to $TLR5^{N14}$, and evaders are less active (EC_{50} > 10 nM) with reduced binding to $TLR5^{N14}$. FliC = *St*FliC**.**

In our screen, most of the 40 flagellins from commensals have TLR5 epitopes whose key residues either are identical to those of *St*FliC (21 of 40) or differ at only one position (16 of 40) (Supp. Fig. 2.4A). In addition to flagellins from commensals, we included three flagellins from pathogens (*St*FliC, *Vibrio cholerae* FlaB, and *Listeria monocytogenes* FlaA) and two negative controls (*Hp*FlaA and *St*FliC-PIM). We generated AP-tagged flagellins to assay $TLR5^{N14}$ binding (Supp. Fig. 2.2) and separately expressed N-terminal Myc-tagged flagellins to quantify TLR5 activation (Supp. Fig. 2.3). Flagellins were incubated with nuclear factor κB (NF-κB) reporter human embryonic kidney (HEK) cells engineered to express TLR5, and NF-κB–dependent AP activity was measured as a readout for TLR5 activation.

Consistent with the notion that binding TLR5 leads to its activation, we generally observed a positive correlation between $TLR5^{N14}$ binding and TLR5 activity (Fig. 2.1D). Flagellins that induced a greater response than that of *St*FliC-PIM (corresponding roughly to a median effective concentration (EC_{50}) of <10 nM) were categorized as “stimulators” (red region in Fig. 2.1D) regardless of their ability to bind $TLR5^{N14}$; this described nearly half the flagellins in our screen. “Evaders,” in contrast, bound and stimulated more weakly than *St*FliC-PIM (blue region). This group included *Hp*FlaA and 11 commensal-derived flagellins. A TLR5 activity score of ≤−3 corresponded broadly to an EC_{50} of >100 nM; these weak agonists showed high

variability in our activity assay (Supp. Fig. 2.4B). The remaining flagellins (9 of 40) resembled evaders with respect to TLR5 activation (stimulated worse than *St*FliC-PIM) but acted like stimulators with regard to TLR5^{N14} binding (stronger interaction compared with *St*FliC-PIM; yellow region). We termed these unexpected ligands silent flagellins in reference to their inability to induce signalling despite intact TLR5 recognition.

2.2.2. Silent flagellins lacked an allosteric activator in D0 domain

We further investigated how silent flagellins decoupled TLR5 ectodomain binding from agonism. We selected FlaB from *Roseburia hominis* (*Rh*FlaB) as our representative silent flagellin because it bound TLR5^{N14} the strongest among the silent flagellins from our initial screen (Fig. 2.1D). *R. hominis* is of wide interest, as it is a common gut commensal species belonging to the *Lachnospiraceae* and generally thought to be anti-inflammatory and thus beneficial to host health [20]. *R. hominis* is motile and expresses *Rh*FlaB in vivo [20,21]. We purified recombinant *Rh*FlaB and observed that, in addition to binding TLR5^{N14}, it also bound full-length human TLR5 tagged with hemagglutinin (HA) (Fig. 2.2A, lane 5). *St*FliC bound more full-length TLR5 compared to *Rh*FlaB, whereas the evader *Hp*FlaA showed an equal lack of binding to both truncated and full-length TLR5 (Fig. 2.2A, lane 4). We also validated that *Rh*FlaB was a weaker TLR5 agonist than *St*FliC-PIM, despite its intact TLR5 epitope (Fig. 2.2B; Supp. Table 2.1).

Next, we tested whether *Rh*FlaB bound TLR5 through its TLR5 epitope. We constructed the flagellin *Rh*FlaB-PIM, which carries the same mutations as *St*FliC-PIM that resulted in loss of binding to TLR5^{N14}. *Rh*FlaB-PIM failed to bind the full-length receptor, consistent with TLR5 binding requiring the TLR5 epitope (Fig. 2.2C, lanes 5-6). However, unlike *Rh*FlaB-PIM, *St*FliC-PIM showed no reduction in binding to full-length TLR5 (Fig. 2.2C, lanes 3-4). This result was unexpected, because *St*FliC-PIM did not bind TLR5^{N14} (Fig. 2.1C) [9].

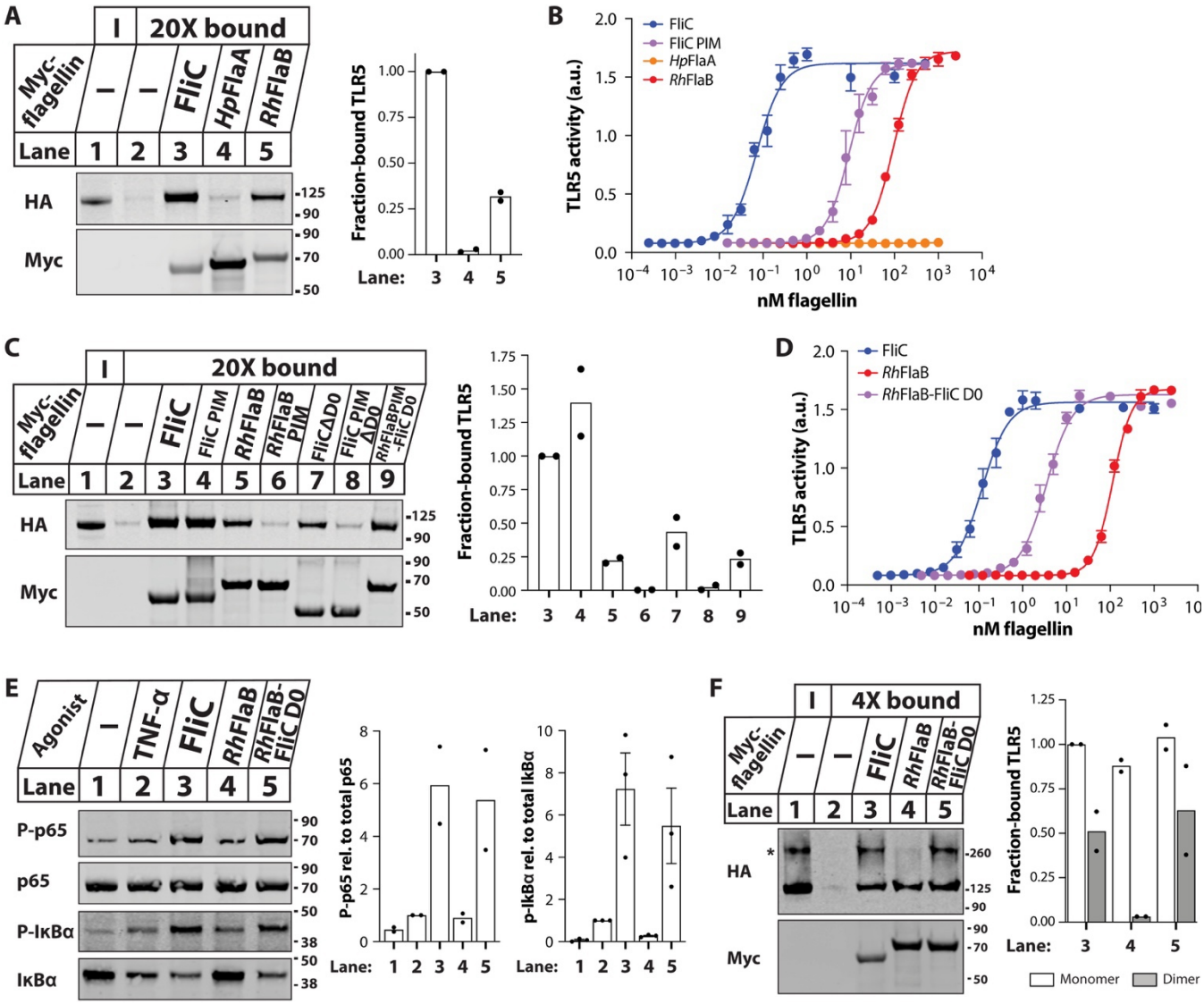

A
I
20X bound
Myc-flagellin
–
–
FliC
HpFlaA
RhFlaB
Lane 1 2 3 4 5
HA
Myc
125
90
90
70
50
Fraction-bound TLR5
1.00
0.75
0.50
0.25
0.00
Lane: 3 4 5
B
FliC
FliC PIM
HpFlaA
RhFlaB
TLR5 activity (a.u.)
2.0
1.5
1.0
0.5
0.0
nM flagellin
C
I
20X bound
Myc-flagellin
– – FliC FliC PIM RhFlaB RhFlaB PIM FliCΔD0 FliC PIM ΔD0 RhFlaBPIM -FliC D0
Lane 1 2 3 4 5 6 7 8 9
HA
Myc
125
90
90
70
50
Fraction-bound TLR5
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Lane: 3 4 5 6 7 8 9
D
FliC
RhFlaB
RhFlaB-FliC D0
TLR5 activity (a.u.)
2.0
1.5
1.0
0.5
0.0
nM flagellin
E
Agonist
–
TNF-α
FliC
RhFlaB
RhFlaB-FliC D0
Lane 1 2 3 4 5
P-p65
p65
P-IκBα
IκBα
90
70
90
70
50
38
50
38
P-p65 rel. to total p65
8
6
4
2
0
Lane: 1 2 3 4 5
p-IκBα rel. to total IκBα
10
8
6
4
2
0
Lane: 1 2 3 4 5
F
I
4X bound
Myc-flagellin
–
–
FliC
RhFlaB
RhFlaB-FliC D0
Lane 1 2 3 4 5
HA
*
Myc
260
125
90
90
70
50
Fraction-bound TLR5
1.25
1.00
0.75
0.50
0.25
0.00
Lane: 3 4 5
Monomer
Dimer

Figure 2.2 – Silent flagellin *Rh*FlaB lacks TLR5 binding site in D0.

(A) Flagellin binding to full-length TLR5: TLR5-HA HEK lysates were incubated with 6xHis-Myc–tagged flagellins and purified on poly-HIS binding beads. Input ("I") and bound fractions ("20X bound") were analysed by immunoblot. Left: representative blots from one of two independent experiments; right: densitometry quantification of HA signal in bound lanes relative to Myc signal, normalized to lane 3. (B) *Rh*FlaB-dependent TLR5 activity: TLR5 HEK-Blue cells were incubated with purified recombinant flagellins, and NF-κB–dependent AP levels were quantified. Error bars are SEM for n = 3. (C) Mapping TLR5 binding sites in flagellin, as in (A). (D) *Rh*FlaB chimera-dependent activation of TLR5, as in (B). (E) Activation of NF-κB pathway: TLR5 HEK-Blue cells were incubated with flagellins or TNF-α in the presence of MG132. Phosphorylation of p65 and IkBα in cell lysates was analysed by immunoblot. Left: representative blots; right: quantification of phosphorylated signal relative to total protein signal, normalized to lane 2. (F) Flagellin binding to preformed TLR5 complexes: BS3-crosslinked cells processed as in (A). Asterisk indicates the TLR5 dimer band. Quantification normalized to lane 3 monomer. (A-F) FliC = *St*FliC.

We hypothesized that *St*FliC-PIM binds the C-terminal LRRs of the TLR5 ectodomain at a location allosteric to the site of pattern recognition. Although the structure of this region of TLR5 is unsolved, the C-terminal LRRs are predicted to interact with the conserved D0 domain of flagellin (Fig. 2.1A,B) [22]. The D0 domain was not required for binding $TLR5^{N14}$ (Fig. 2.1C) and is also absent in the *St*FliC-TLR5 crystal structure [9]. Several studies previously reported the necessity of the *St*FliC D0 for TLR5 activation [9,19]. However, the mechanism is unclear and the authors unequivocally conclude that the D0 domain does not directly bind the receptor.

We tested for a TLR5 binding site in the *St*FliC D0 by assessing the binding of *St*FliC-PIM-ΔD0 to full-length TLR5. Because *St*FliC-PIM did not bind $TLR5^{N14}$, if the D0 bound TLR5 LRRs 15-22, then *St*FliC-PIM-ΔD0 should have been unable to bind full-length TLR5. Consistent with an additional binding site in the D0 of *St*FliC, *St*FliC-PIM-ΔD0 showed a substantial loss of binding to TLR5 compared with *St*FliC-PIM and *St*FliC-ΔD0 (Fig. 2.2C, lane 8 vs lanes 4,7). Given our observation that the TLR5 epitope was necessary for *Rh*FlaB to bind the receptor, such that *Rh*FlaB-PIM could not bind full-length TLR5, we predicted that the *St*FliC D0 would restore TLR5 binding to *Rh*FlaB-PIM. As expected, swapping *St*FliC D0 for the native *Rh*FlaB D0 rescued *Rh*FlaB-PIM binding (Fig. 2.2C, lanes 6, 9). Together, these results suggested that

*St*FliC D0 allosterically bound TLR5, in direct contradiction to previous findings [9,19]. The additional binding site also explained why *St*FliC pulled down more full-length TLR5 than *Rh*FlaB (Fig. 2.2A).

The discovery of an allosteric TLR5 binding site in *St*FliC prompted us to test its impact on TLR5 activation. We purified recombinant *Rh*FlaB chimera expressing the *St*FliC D0 (*Rh*FlaB-*St*FliC D0) and assayed TLR5 signalling. As expected from its greater ability to bind full-length TLR5, the chimeric flagellin was 100-fold more stimulatory than *Rh*FlaB with its native D0 (Fig. 2.2D). We confirmed the NF-κB reporter response by directly measuring phosphorylation of the NF-κB subunit p65 and inhibitor IkBα (Fig. 2.2E). After incubation with *St*FliC or *Rh*FlaB-*St*FliC D0, cells showed increased phosphorylation of p65 and IkBα, accompanied by reduced total IkBα levels, relative to untreated cells (Fig. 2.2E, lanes 3,5 versus lane 1). This indicates activation of the NF-κB pathway [23]. Although cells treated with the positive control tumour necrosis factor–α (TNF-α) (lane 2) showed a modest increase in IkBα phosphorylation and degradation, incubation with *Rh*FlaB did not induce phosphorylation or degradation of IkBα (lane 4). Combined, these results showed that the *St*FliC D0 activated *Rh*FlaB-dependent signalling of TLR5.

2.2.3. D0 binding site targeted flagellin to preformed TLR5 dimers

We hypothesized that the additional TLR5 binding site in the *St*FliC D0 increased activity in part by enabling *Rh*FlaB-*St*FliC D0 to interact with more TLR5 receptors than *Rh*FlaB. TLR5 activation requires the formation of a symmetric 2:2 flagellin:TLR5 complex [9,24]. How this complex is assembled remains unclear, although it is widely stated that flagellin binding induces TLR5 dimerisation [19,25–27]. Early cryo–electron microscopy work revealed, however, that human TLR5 forms asymmetric homodimers in the absence of flagellin, a conformation likely associated with multiple ligand binding sites and thus a possible target of the *St*FliC D0 domain [28]. We investigated whether TLR5 formed unliganded dimers by briefly treating TLR5-HA HEK cells with the membrane-impermeable crosslinker BS3 (Fig. 2.2F). In addition to monomeric TLR5, we detected a higher molecular weight species consistent with the size of a TLR5 dimer. This result suggested that preformed TLR5 dimers were present on the cell surface in the absence of ligand-induced dimerisation that is commonly invoked for TLR5.

We tested the ability of *St*FliC and *Rh*FlaB to interact with TLR5 dimers. We observed that *St*FliC bound both monomer and dimer, whereas *Rh*FlaB only interacted with the monomer (Fig. 2.2F, lanes 3 and 4). The switching out of its native D0 for *St*FliC D0 endowed *Rh*FlaB the ability to bind the dimer (Fig. 2.2F, lane 5). This result suggested that the *St*FliC D0 directly bound the ectodomain and activated TLR5 signalling in part by mediating binding to preformed TLR5 dimers. This observation supported our hypothesis that an allosteric binding site in its D0 enabled *St*FliC to interact with more receptors than *Rh*FlaB.

2.2.4. The allosteric binding site was located exclusively in the conserved C-terminal D0 domain

Given the high conservation of the flagellin D0 domains, we next asked which domain (nD0 or cD0) harboured the allosteric binding site. We recombinantly expressed and purified *Rh*FlaB chimeras containing either the *St*FliC nD0 or cD0 and tested their ability to stimulate TLR5. The *St*FliC nD0 did not affect TLR5 signalling by *Rh*FlaB, whereas *Rh*FlaB-*St*FliC cD0 phenocopied *Rh*FlaB-*St*FliC D0 (Fig. 2.3A). The *St*FliC cD0, therefore, mediated activation of TLR5. We further confirmed that the *St*FliC cD0, but not the nD0, rescued *Rh*FlaB-PIM binding to TLR5 (Fig. 2.3B, lane 6).

The *Rh*FlaB and *St*FliC cD0s were 46% identical across 39 amino acids (Fig. 2.3C). Residues in the *St*FliC cD0 are important for TLR5 activation [19], several of which were not conserved in *Rh*FlaB. We hypothesized that these amino acids mediate binding to the dimer and tested whether *Rh*FlaB mutants that express the equivalent residues in *St*FliC activated TLR5 (Fig. 2.3D). However, of the three *Rh*FlaB mutants that we tested, only *Rh*FlaB H478R showed increased stimulation of TLR5. We concluded that the allosteric binding site likely encompasses multiple residues distributed across the cD0 domain.

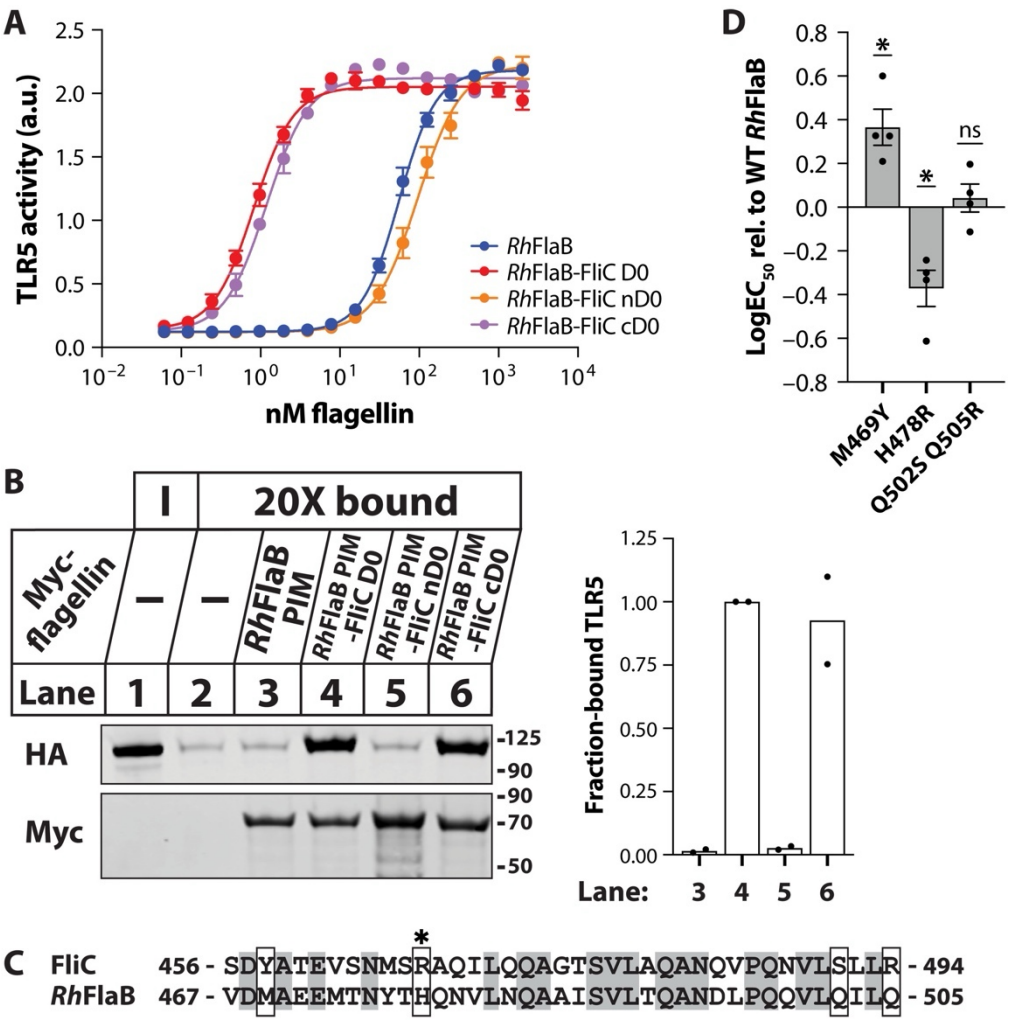


Figure 2.3 – FliC cD0 domain mediated binding to and activation of TLR5.
(A) Effect of *St*FliC nD0 and cD0 domains on *Rh*FlaB-dependent TLR5 activity, as in Fig. 2.2B. Data are means ± SEM for n = 3. (B) Effect of *St*FliC nD0 and cD0 domains on *Rh*FlaB-PIM binding to TLR5, as in Fig. 2.2A. Left: representative blots from one of two independent experiments; right: quantification of HA signal relative to Myc signal, normalized to lane 4. (C) Alignment of *St*FliC and *Rh*FlaB cD0 domain amino acid sequences. Identical residues are shaded grey, and boxes indicate mutants tested in (D). Asterisk denotes residue that increased *Rh*FlaB-dependent TLR5 activation. (D) TLR5 activation by *Rh*FlaB cD0 mutants, as in (A). Data shown are differences in $\log EC_{50}$ of mutants relative to WT *Rh*FlaB, with EC_{50} determined by weighted, nonlinear regression analysis with Hill slope constrained to 1. Significance in mean differences between *Rh*FlaB WT and mutants was calculated by one-sample *t* test against theoretical mean equal to 0 (*P < 0.05; ns, not significant). (A-C) FliC = *St*FliC.

2.2.5. Silent flagellins were prevalent in the human gut and enriched in non-industrialized populations

Next, we assessed how widespread silent flagellins were in the human microbiome. We searched for peptide sequences of silent flagellins using a published database comprising more than 33,000 flagellins (Methods; Supp. Fig. 2.1) [29]. Candidate silent flagellins were selected on the basis of their presence in human gut metagenomes and by similarity to the C-terminal region of *Rh*FlaB (Fig. 2.3; Supp. Fig. 2.1). The list was further curated to exclude flagellins containing a basic residue (R/K) at position *Rh*FlaB amino acid 478 based on our observation that *Rh*FlaB H478R showed a slight, but significant, increase in TLR5 stimulation (Fig. 2.3D). The final candidate silent flagellins were mostly, but not exclusively, from species belonging to the *Lachnospiraceae* family (74 of 76) (Supp. Fig. 2.5).

To verify whether these 76 candidate silent flagellins were silent, we expressed them recombinantly to screen for both TLR5 signalling and TLR5^{N14} binding (Fig. 2.4A). Compared with our initial screen (Fig. 2.1D), we were successful in enriching for silent flagellins: More than half (44 of 76) were weaker TLR5 agonists and stronger TLR5^{N14} binders relative to *St*FliC-PIM (Fig. 2.4A, yellow region). Given its importance in Crohn's disease, we additionally tested the flagellin CBir1, a known weak TLR5 agonist [13,30]. CBir1 bound TLR5^{N14}, categorizing it as a silent flagellin. The remaining 34 candidates were equally distributed among stimulators (red region) and evaders (blue region).

To assess whether the mechanism was the same for *Rh*FlaB as for the above set of silent flagellins, we examined the impact of swapping in the *St*FliC D0 domain on a subset representing a broad range of TLR5^{N14} binding strengths. Although the magnitude differed among candidates, the *St*FliC D0 universally increased TLR5 signalling for all silent flagellins tested, including CBir1 (Fig. 2.4B). Moreover, these silent flagellins belonged to common taxa of the human gut microbiome, including multiple species of Roseburia [17]. The *St*FliC D0 did not affect TLR5 evasion by *Hp*FlaA, consistent with previous observations [19].

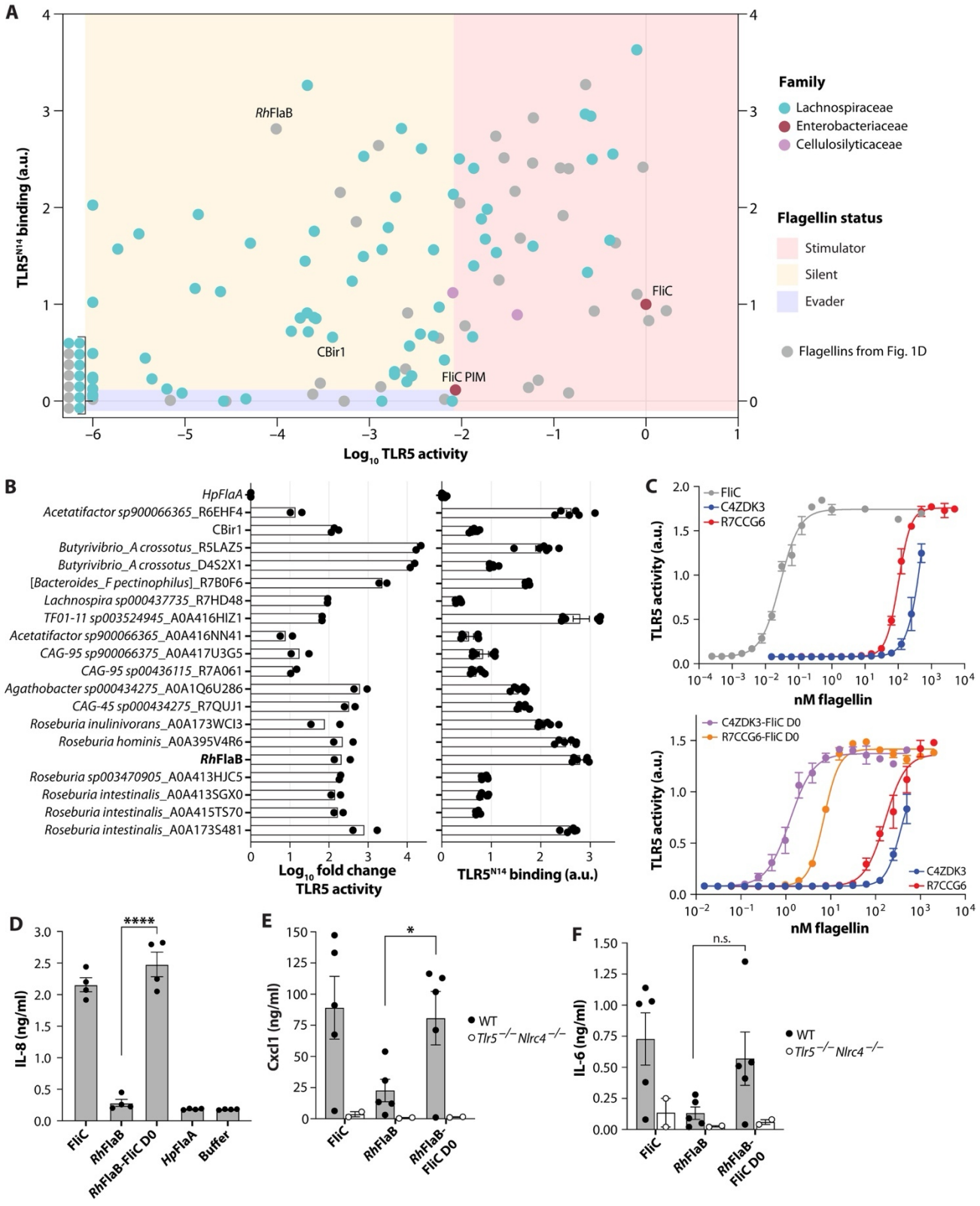

A
RhFlaB
CBir1
FliC
FliC PIM
TLR5N14 binding (a.u.)
Log10 TLR5 activity
Family
Lachnospiraceae
Enterobacteriaceae
Cellulosilyticaceae
Flagellin status
Stimulator
Silent
Evader
Flagellins from Fig. 1D
B
HpFlaA
Acetatifactor sp900066365_R6EHF4
CBir1
Butyrivibrio_A crossotus_R5LAZ5
Butyrivibrio_A crossotus_D4S2X1
[Bacteroides_F pectinophilus]_R7B0F6
Lachnospira sp000437735_R7HD48
TF01-11 sp003524945_A0A416HIZ1
Acetatifactor sp900066365_A0A416NN41
CAG-95 sp900066375_A0A417U3G5
CAG-95 sp00436115_R7A061
Agathobacter sp000434275_A0A1Q6U286
CAG-45 sp000434275_R7QUJ1
Roseburia inulinivorans_A0A173WCI3
Roseburia hominis_A0A395V4R6
RhFlaB
Roseburia sp003470905_A0A413HJC5
Roseburia intestinalis_A0A413SGX0
Roseburia intestinalis_A0A415TS70
Roseburia intestinalis_A0A173S481
Log10 fold change TLR5 activity
TLR5N14 binding (a.u.)
C
FliC
C4ZDK3
R7CCG6
TLR5 activity (a.u.)
nM flagellin
C4ZDK3-FliC D0
R7CCG6-FliC D0
D
IL-8 (ng/ml)
****
FliC
RhFlaB
RhFlaB-FliC D0
HpFlaA
Buffer
E
Cxcl1 (ng/ml)
*
WT
Tlr5−/− Nlrc4−/−
F
IL-6 (ng/ml)
n.s.

Figure 2.4 – Silent flagellins were widespread among *Lachnospiraceae* that colonize the human gut.

(A) TLR5 activity versus $TLR5^{N14}$ binding for *Rh*FlaB-like flagellins, as in Fig. 2.1D. Data are mean for n ≥ 2; values relative to *St*FliC. See Supp. Fig. 2.5. (B) Native D0s swapped for *St*FliC D0 in subset of *Lachnospiraceae* silent flagellins: TLR5 activity measured as in (A). Bar graph shows mean difference between WT and chimeric flagellins from at least two independent experiments. $TLR5^{N14}$ binding on right. (C) Top: TLR5 activation by flagellins from human stool as in Fig. 2.2B. Bottom: Effect of *St*FliC D0 on TLR5 activity. Data are means ± SEM for n = 3. See tables S1 and S2 and data file S1. (D) Flagellin-dependent responses in colonoids: IL-8 levels quantified by ELISA. Data represent means ± SEM. Significance between *Rh*FlaB and *Rh*FlaB-*St*FliC D0 means was determined by unpaired, two-tailed *t* test (****P < 0.0001). (E and F) TLR5-dependent responses in mice: Serum Cxcl1 (E) and IL-6 (F) levels were measured by ELISA after intraperitoneal injection of flagellins. Significance between *Rh*FlaB and RhFlaB-*St*FliC D0 means was determined by unpaired, two-tailed *t* test (*P < 0.05). (A-F) FliC = *St*FliC.

Given that flagellin is facultatively expressed and that expression in the gut can vary depending on external factors [12], we assessed the presence of silent flagellins directly from healthy human stool. Endogenous flagellins were isolated using TLR5 as bait and identified by mass spectrometry (MS). Peptides were searched against a custom flagellin database built from metagenome sequences generated from the same stool sample. Of the 12 flagellins identified, 10 were ascribed to *Lachnospiraceae* (Supp. Table 2.2; Supp. Data 2.1). This was consistent with the taxonomic affiliation of the abundantly expressed flagellins in humans without IBD (Supp. Fig. 2.6A,B) [18]. We recombinantly expressed and purified the top two candidates to assay TLR5 signalling. Both flagellins weakly activated TLR5 with EC_{50} values greater than 100 nM (Fig. 2.4C; Supp. Table 2.1). However, similar to *Rh*FlaB, swapping in the *St*FliC D0 for the native D0 profoundly increased their ability to stimulate TLR5.

We further verified that silent recognition occurred when TLR5 was endogenously expressed in human organoids and in mice. In three-dimensionally cultured human colon organoids, *St*FliC stimulated the secretion of interleukin-8 (IL-8) (Fig. 2.4D), a proinflammatory cytokine produced downstream from TLR5 activation [31]. IL-8 levels in *Rh*FlaB-treated organoids were similar to those of *Hp*FlaA- and buffer-treated controls, whereas organoids incubated with *Rh*FlaB-*St*FliC D0 had significantly higher IL-8 levels (Fig. 2.4D). We then characterized the response to silent

flagellin *Rh*FlaB by wild-type (WT) and *Tlr5−/−Nlrc4−/−* C57BL/6 mice. NLRC4 is an intracellular sensor of flagellin [32,33], and mice lacking both Tlr5 and Nlrc4 failed to respond to flagellin [34]. WT mice injected with *Rh*FlaB had lower levels of the epithelial proinflammatory cytokine Cxcl1 compared with animals injected with *St*FliC and *Rh*FlaB-*St*FliC D0 (Fig. 2.4E). We saw a similar trend for IL-6, an inflammatory cytokine produced by myeloid cells (Fig. 2.4F). Consistent with these cytokine responses being flagellin-dependent, Cxcl1 and IL-6 levels were not elevated in *Tlr5−/−Nlrc4−/−* double-knockout mice after injection. IL-18 levels were similar between WT animals injected with *St*FliC and *Rh*FlaB (Supp. Fig. 2.7). Because IL-18 production requires NLRC4, but not TLR5 [34], this suggested that silent flagellins induced an inflammatory response when detected intracellularly. Together, these results indicated that TLR5's silent recognition of flagellin was not species specific and that the *St*FliC D0 activated both human and mouse endogenously expressed TLR5.

Our identification of an allosteric activator of TLR5 in *St*FliC suggested a mechanism by which this receptor can respond to minute levels of stimulatory flagellin. Commensal members of the gut microbiome, such as members of *Lachnospiraceae*, can produce an array of flagellins that are silent, stimulatory, or evasive. *R. hominis* itself encodes several flagellins that fell into all three categories based on our binding and activation criteria (Fig. 2.5A; Supp. Tables 2.1,2.3) [20]. This within-species flagellin diversity reflected the flagellin diversity encoded broadly in human gut metagenomes, where all three types were detected (Fig. 2.5B; Supp. Fig. 2.8). We observed that metagenomes from non-industrialized populations encoded a greater proportion of all three flagellin types compared with metagenomes from industrialized populations, despite lower relative abundance of *Lachnospiraceae* (Fig. 2.5C; Supp. Fig. 2.6B,C). The decrease in flagellin abundance with industrialization was most pronounced for the silent flagellins (Fig. 2.5D). Reduced flagellin diversity may reflect shifts in host-microbiome interactions associated with industrialization.

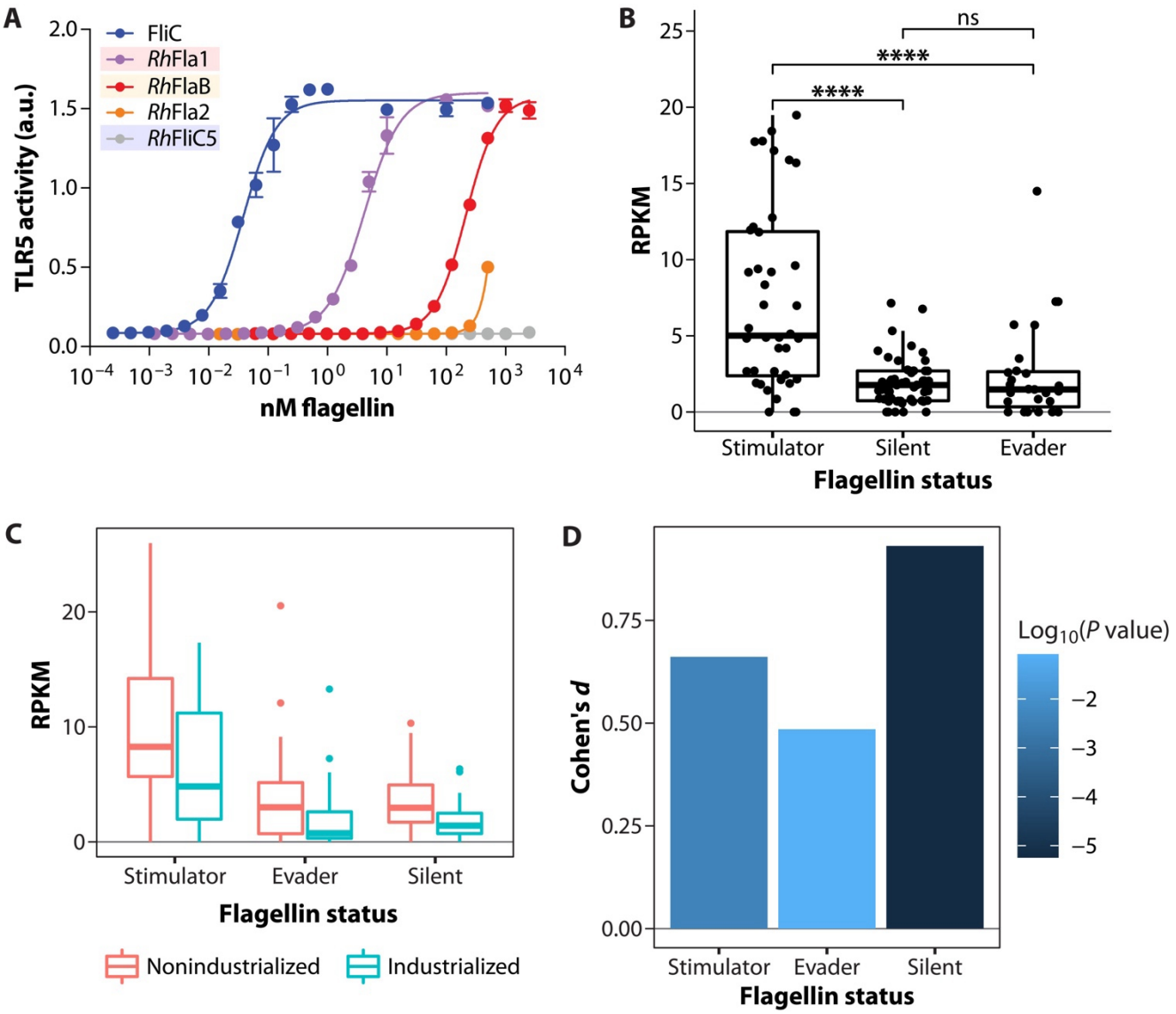


Figure 2.5 – Silent flagellins were enriched in non-industrialized populations.
(A) Within-species flagellin diversity: TLR5 activity of flagellins encoded by *R. hominis* (Supp. Table 2.3) as in Fig. 2.2B. Data are means ± SEM for n = 3. Flagellins are shaded to reflect status (see Supp. Fig. 2.4 for RhFliC5 binding and Supp. Table 2.1 for EC_{50} values). FliC = *St*FliC. (B) Abundance of stimulator, silent, and evader flagellins: Boxplots show median RPKMs of flagellins in human metagenomes (n = 1783). Significance was determined by post hoc pairwise *t* tests (analysis of variance (ANOVA), type-II, F = 25.3, P = 6.8 × 10−10). ****P < 0.0001. See Supp. Fig. 2.8. (C) Median RPKM of flagellins by industrialization status and flagellin status: See Supp. Fig. 2.8 and Supp. Data 2.2. (D) Effect sizes (Cohen's d) and *t* test P values corresponding to industrialization versus non-industrialization comparisons within each flagellin status.

2.3. Discussion

The understanding of how TLR5 interacts with its primary ligand, flagellin, comes mostly from the study of flagellins encoded by Pseudomonadota (formerly Proteobacteria), in particular, the pathogens *Salmonella* and *H. pylori*, and others such as *Escherichia coli*. These studies led to the discovery of the TLR5 epitope, a conserved region in the flagellin nD1, whose binding is considered to be required for TLR5 recognition and subsequent stimulation. We showed here that, in addition to the TLR5 epitope, the D0 of *St*FliC allosterically bound TLR5, akin to a homotropic ligand. Our work indicated that, in addition to these modes of interaction (i.e., recognition followed by activation versus nonrecognition), a third mode, very common in commensal bacteria prevalent in the gut, allows bacteria to express flagellins that retain the TLR5 epitope without inducing a robust TLR5 response. Although commensal bacteria also produced stimulator and evader flagellins (all three types can be encoded in a single genome), our analysis of metagenomes indicated that silent flagellins are very common in the human gut and therefore represent a substantial, previously unappreciated, yet physiologically relevant, population of TLR5 ligands.

Our current model proposes that TLR5 adopts different conformations, as evidenced by the presence of both monomeric and dimeric TLR5, and that flagellins have different binding preferences for these receptor states. Although the *St*FliC D0 conferred binding to dimeric TLR5, silent flagellins failed to bind this receptor state and were weak agonists relative to *St*FliC as a result. However, receptor binding to the TLR5 epitope enabled silent flagellins to activate TLR5 at high concentrations, in contrast to *Hp*FlaA. Our data further suggest that *St*FliC binding to TLR5 complexes induced a conformational change, rather than receptor dimerisation, similarly to what has been described for other TLRs [35]. This mechanism is compatible with previous studies that identify an additional TLR5 binding interface in the cD1 domain of flagellin [7, 9]. Although our work showed that *St*FliC-PIM-ΔD0 failed to bind the receptor, this does not exclude a role for the cD1 domain in TLR5 activation: TLR5 conformational changes are likely accompanied by alterations in flagellin-TLR5 binding. Mutating the *St*FliC cD1 is known to affect TLR5 activation, consistent with a direct interaction between this region and the receptor in the crystal structure of the complex.

Together, our work highlights how pattern recognition by TLR5 can occur without downstream signalling. By probing into the weak agonism of flagellins produced by commensal gut bacteria, we discovered a third class of flagellins that contain the epitope recognised by TLR5 yet poorly activate the receptor. Allosteric activation of TLR5 allows the host to tolerate silent flagellins from commensal bacteria while remaining responsive to faint levels of stimulatory flagellin.

2.4. Methods

2.4.1. Study design

The objective of this study was to identify flagellins encoded by common human gut bacteria and characterize their interaction with the host immune receptor TLR5. Publicly available gut metagenomes from non-IBD individuals were used to identify candidate flagellins. We recombinantly expressed these proteins and used biochemical assays to test their ability to bind and activate TLR5. We identified a class of flagellin, termed “silent flagellin,” characterized by strong binding and weak activity. Mutational analysis revealed that this “silent recognition” by TLR5 resulted from the inability of silent flagellins to bind TLR5 dimers through a previously unidentified cD0 binding site. Last, we accessed the prevalence of silent flagellins in human gut metagenomes globally.

2.4.2. Reagents

Chemicals were obtained from Sigma-Aldrich (2-mercaptoethanol, bromophenol blue, bovine serum albumin (BSA), LB with agar, EDTA, imidazole, guanidine HCl, Igepal CA-630, N-acetylcysteine, nicotinamide, SB202190, SDS, and Tween 20), Carl Roth (bis-tris, bicine, Hepes, NaH_2PO_4, Na_2HPO_4, urea, tris, NaCl, KCl, methanol, milk powder, MES, and $MgCl_2$), or Thermo Fisher Scientific (glycerol) unless stated otherwise. Primary antibodies included mouse anti-Myc monoclonal 4A6 (Merck 05-724), rat anti-HA monoclonal 3F10 (Roche 11867423001), rabbit anti–phospho–NF-κB p65 (Ser^{536}) monoclonal 93H1 (Cell Signaling Technology (CST) #3033), rabbit anti–phospho-IkBα (Ser^{32}) monoclonal 14D4 (CST #2859), mouse anti–NF-κB p65 monoclonal L8F6 (CST #6956), and mouse anti-IkBα monoclonal L35A5 (CST #4814). Secondary antibodies were from LI-COR (IRDye 800CW goat anti-rat IgG and IRDye 680RD/800CW goat anti-mouse IgG).

2.4.3. $TLR5^{N14}$ binding to AP-tagged flagellins

Flagellin (prey) and $TLR5^{N14}$ (bait) constructs were cloned into modified pLIB vector containing N-terminal BiP or GP64 signal peptide, respectively, and C-terminal Strep-II affinity tag as described previously [36,37]. Flagellin constructs were further modified to include C-terminal AP tag. $TLR5^{N14}$ bait construct was generated by inserting first 14 LRRs of human TLR5 and C-

terminal variable leucine repeat adaptor sequence upstream of human IgG-Fc and 6XHis tag. Constructs were codon-optimized for *Drosophila melanogaster* expression and assembled using Gibson assembly (New England Biolabs (NEB) #E2611). Plasmids were transformed into NEB 5-alpha *E. coli* (NEB #C2987), cultured in 5 ml of LB supplemented with carbenicillin (100 μg/ml; Invitrogen), and purified using ZymoPURE plasmid miniprep kit (Zymo Research #D4211).

Sequence-verified constructs were transformed into DH10EMBaCY cells to generate recombinant baculovirus genomes (bacmids) as described previously [36,38]. Bacmids were transfected into SF9 insect cells using FuGENE 6 transfection reagent (Promega #E2311) and expanded to generate initial (V0) viral stock. Two successive rounds of viral amplification were performed through 1:100 (v/v) inoculations of 25 ml of SF9 cell cultures with baculovirus-containing supernatant, each lasting 72 hours. The supernatant from the second culture (V2) was used to induce protein production. Protein production was optimized for each construct through expression trials in Hi5 cells using inoculation ratios from 1:10 to 1:1000. Proteins were harvested after 48 hours except for *St*FliC-PIM and *Hp*FlaA, which were harvested after 72 hours.

Secreted prey proteins in supernatant were collected after centrifugation (800 relative centrifugal force, 5 min) and filter-purified. Bait protein was obtained from Hi5 cell pellets by ultrasonication in lysis buffer (50 mM Hepes, 300 mM NaCl, 5% glycerol, 0.1% Tween 20, 5 mM 2-mercaptoethanol, 1× protease inhibitor (Serva #39107)) and then filter-purified. Protein expression was confirmed by immunoblotting with Strep-Tactin conjugated to horseradish peroxidase (HRP) (IBA #2-1502-001).

Flagellin candidates from human commensals were polymerase chain reaction (PCR)–amplified from pETM11 vectors using Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific #F530S). Inserts were cloned into pECIA14 vector using Gibson assembly (NEBuilder HiFi DNA Assembly Master Mix #E2621). Insect cell protein expression was performed as previously described [39,40] with minor modifications. AP-tagged flagellins were expressed via transient transfection in D*. melanogaster* Schneider 2 (S2) cells using Expres2 TR (Expres2ion Biotechnologies). During transfection, cells were shifted from 27° to 25°C. Protein expression was induced with 1 mM $CuSO_4$ 24 hours after transfection. Supernatant media were collected 72 hours after induction. Protease inhibitors (cOmplete, EDTA-free, Roche #5056489001) and

0.02% NaN_3 were added to the media. Expressed proteins were confirmed by immunoblotting using anti–Flag-HRP (Sigma-Aldrich #A8592) antibody. Nine candidates failed to express, resulting in 40 abundant and 76 *Rh*FlaB-like candidates.

Each assay contained *Rh*FlaB, *St*FliC, and *Hp*FlaA as controls. AP-tagged flagellins were quantified by incubating 25 µL of each prey sample in 100 µL of BluePhos microwell substrate (SeraCare #5120-0059) for 30 min at room temperature. Prey concentrations were normalized to *Rh*FlaB based on their relative response to the BluePhos substrate by dilution in PBS-T (phosphate-buffered saline + 0.1% (v/v) Tween 20). $TLR5^{N14}$-flagellin binding was assayed following a previously described method [39], with some modifications. Flagellin proteins and $TLR5^{N14}$ were prediluted 1:10 in PBS-T and then mixed by rocking for 2 hours at 4°C. Ninety-six–well Pierce protein A–coated plates (Thermo Fisher Scientific #15132) were activated by two washes with 200 µL of PBS-T. $TLR5^{N14}$-flagellin mixtures (200 µL) were added to each well and incubated overnight at 4°C. Each sample had three replicates per plate, alongside corresponding flagellin-only wells as controls. AP activity was quantified using BluePhos microwell substrate, with the absorbance measured at 650 nm after 3 hours. Relative $TLR5^{N14}$-flagellin binding strength was quantified by subtracting the background absorbance from flagellin-only wells from mixed wells and then normalized to *St*FliC. Negative values were set to zero.

2.4.4. Myc-tagged flagellins

Flagellin cDNAs were commercially synthesized (Integrated DNA Technologies (IDT)) with N-terminal Myc tags and cloned into pETM11 vector downstream from 6x-histidine tag (NEBuilder HiFi DNA assembly) before transformation in TOP10 cells (Invitrogen). Mutated flagellins and flagellins containing the *Salmonella* FliC (UniProt ID #P06179) D0 (amino acids 1 to 41 and amino acids 456 to 494) were generated using Gibson assembly. Plasmid inserts were Sanger-sequenced and transformed into ClearColi BL21(DE3) cells (Lucigen).

Unless stated otherwise, ClearColi-transformed cells were grown at 37°C in LB supplemented with kanamycin (50 µg/ml; Gibco) to OD600 (optical density at 600 nm) of 0.4 to 0.6/ml. Protein expression was induced with 0.1 M isopropyl-β-d-thiogalactopyranoside (MP Biomedicals), and cells were incubated for 2.5 hours at 37°C. Cells were harvested by centrifugation (3400g for 15

min), and pellets were resuspended in 1:100 volume lysis buffer (10 mM tris-HCl (pH 8), 8 M urea, and 100 mM NaH_2PO_4). Cells expressing *H. pylori* FlaA (or *Hp*FlaA-*St*FliC D0) were grown at 25°C for 4 hours after induction and then lysed in 10 mM tris-HCl (pH 8), 6 M guanidine HCl, and 100 mM Na_2HPO_4. Cells were lysed for 2 hours at room temperature, and the supernatant was collected after centrifugation (16,000 rcf for 20 min). These lysates were used to screen for TLR5 activity in Fig. 2.1D and Fig. 2.4A,B.

For flagellin purification, cleared lysates were incubated with Ni-NTA (nitrilotriacetic acid) agarose (1:4 volume; QIAGEN) for 2 hours at room temperature. Agarose was then washed once with two column volumes (CV) of lysis buffer and then washed twice with two CVs of lysis buffer (pH 6.3) and twice with two CVs of lysis buffer (pH 5.9). Ni-NTA–bound proteins were eluted with 0.5 CV of lysis buffer (pH 4.5). Eluates were dialyzed at 4°C in 20 mM tris-HCl (pH 8), 300 mM NaCl, and 5 mM $MgCl_2$ with the exception of *H. pylori* FlaA, CAZDK3, CAZDK3-*St*FliC D0, and R7CCG6-*St*FliC D0, which were dialyzed in 50 mM Hepes (pH 7.4), 114 mM NaCl, and 1.5 mM Na_2HPO_4. Purified proteins were quantified by bicinchoninic acid assay (BCA; Pierce), and aliquots were stored at −80°C. For mouse experiments, non–Myc-tagged flagellins were purified as described above and passed through polymyxin B–agarose (Sigma-Aldrich) before BCA quantification.

2.4.5. TLR5 HEK-Blue activity assay

HEK-Blue hTLR5 cells (InvivoGen) were grown in 5% CO2 at 37°C in medium (GlutaMAX Dulbecco's modified Eagle's medium (DMEM) and 10% foetal bovine serum (FBS; Gibco)) containing selection antibiotics (Zeocin (100 µg/ml; InvivoGen) and blasticidin (30 µg/ml; InvivoGen)) to 90% confluence. Cells were detached using prewarmed PBS (pH 7.4; Gibco), resuspended in medium (1 × 105 per ml), and distributed in 96-well plates (180 µL per well). Purified flagellins were serially diluted in PBS, and 20 µL were added in triplicate to HEK cells for a final volume of 200 µL per well. Plates were incubated for 18 hours (5% CO2, 37°C), and then 20 µL of medium from each well were added to 180 µL of QUANTI-Blue solution (InvivoGen). Absorbance at 635 nm was measured after 30 min. EC_{50} values for purified flagellins were calculated by plotting absorbance values against flagellin concentrations in Prism 9 (GraphPad) and performing weighted, nonlinear regression analysis. Unless stated otherwise, no constraints were set. $LogEC_{50}$ values were averaged.

Flagellin candidates were screened for TLR5 activity (Fig. 2.1D and Fig. 2.4A,B) as described above with the following modifications. HEK cells were plated at 190 µL per well, and bacterial lysates expressing flagellins were serially diluted in PBS such that final volume in the assay ranged from 1 to 1 × 10−6 µL. Lysates were added in duplicate to HEK cells to a final volume of 200 µL per well. The relative amount of flagellin expression was determined by quantifying Myc signal in 0.1 µL of lysate (diluted 10-fold in PBS) by immunoblotting. TLR5 activity was determined by plotting absorbance values against bacterial lysate volume in Prism 9 and calculating EC_{50} values from weighted, nonlinear regression analysis with the following constraints: Hill slope was set to 1 and top less than or equal to shared value. Activity was normalized to protein expression by calculating EC_{50} multiplied by Myc signal relative to *St*FliC. Values were negative $\log_{10}$-transformed. Candidates with activity less than empty vector control were assigned $\log_{10}$ TLR5 activity values of −6.

2.4.6. Full-length TLR5-flagellin pull-down assays

hTLR5-HA HEK cells (InvivoGen) were incubated in 5% CO2 at 37°C in medium (GlutaMAX DMEM and 10% FBS; Gibco) supplemented with selection antibiotic blasticidin (10 µg/ml). Cells were grown to 90% confluency in T75 flasks (Grenier), detached with prewarmed PBS (pH 7.4), and harvested by centrifugation (3400 rcf for 3 min). Pellets were resuspended in 1 ml of prechilled lysis buffer (25 mM Tris-HCl (pH 7.5), 1% (v/v) Igepal CA-630, 100 mM NaCl, 5 mM imidazole, and 10% v/v glycerol) supplemented with 1× Halt protease inhibitors (Thermo Fisher Scientific).

Cells were incubated on ice for 15 min, followed by centrifugation at 4°C (855g for 20 min). Supernatant was collected, and protein levels were quantified by BCA. For each reaction, 500 µg of cell lysate were diluted in 300 µL of lysis buffer and incubated with 150 pmol of Myc-tagged flagellin or buffer control for 2 hours at 4°C. TALON resin (Takara) was washed twice with PBS (pH 7) and twice with lysis buffer before resuspension in PBS; 20 µL of washed resin were added to each reaction. After 2 hours of incubation at 4°C, resin was pelleted at 4°C (855g for 1 min) and washed 3× with 750 µL of lysis buffer supplemented with protease inhibitors (cOmplete, mini, EDTA-free). The bound fraction was eluted with 20 µL of lysis buffer containing 200 mM imidazole and transferred to fresh tubes. Samples (25 µg of input and eluted fractions) were analysed by gel electrophoresis followed by immunoblotting.

For pull-downs with cross-linked lysates, hTLR5-HA HEK cells were plated in Nunc EasYdishes (Thermo Fisher Scientific #150460) to 70% confluence. Immediately before BS3 treatment, cells were washed with 1 ml of PBS and then incubated with 2.33 mM BS3 cross-linker (Pierce 21580) in 1.5 ml of PBS for 30 s on ice. Reaction was quenched with 1 M Tris-HCl (pH 8) to a final concentration of 50 mM. Crosslinker solution was removed, and cells were scraped into prechilled lysis buffer (750 µL) and then processed as described above with the following modification: The input sample was incubated with anti-HA resin (Roche #11815016001) for 2 hours at 4°C, and bound proteins were eluted with 20 µL of HA peptide (1 mg/ml in Tris-buffered saline (TBS); Thermo Fisher Scientific 26184) for 15 min at 37°C.

2.4.7. Flagellin diversity in human microbiome

Flagellins in genomes of isolates and metagenome-assembled genomes from microbial taxa present in the human gut microbiome were assessed using predicted proteome from Unified Human Gastrointestinal Genome catalog v.2.0 (available at www.ebi.ac.uk/metagenomics/genome-catalogues/human-gut-v2-0). We considered a predicted protein a flagellin if eggNOG ID was “COG1344” and annotated as flagellin (as opposed to other accessory proteins). On the basis of this, we identified 5404 (0.053%) flagellins spanning eight phyla.

2.4.8. Flagellin selection

The selection process of candidate flagellins is summarized in Supp. Fig. 2.1. Using the flagellin database from [29], we annotated database sequences with InterProScan [41] and only retained sequences having both Pfam domains PF00669 (flagellin N-terminal domain, “ND”) and PF00700 (flagellin C-terminal domain, “CD”). Our filtered database comprised 33,051 flagellin protein sequences. To identify flagellins abundant in the human gut, we filtered the hits from metagenomes of healthy individuals from the IBD multi-omics database of Lloyd-Price et al. [18] using the median of read counts as a cutoff. The taxonomy of those accessions was assigned using the taxonomizr R package [42] to obtain their taxids and a further search of their tax IDs within Genome Taxonomy Database (GTDB) release 95 (https://data.ace.uq.edu.au/public/gtdb/data/releases/latest/). We searched the remaining

accessions using Entrez Direct [43] in the National Center for Biotechnology Information (NCBI) Identical Protein Groups database [44] to obtain their assembly accessions, which were further searched within the GTDB taxonomy. We ranked the resulting accessions by their read counts and randomly selected across the list to get a broad taxonomic representation of flagellins from the human gut, reducing the dataset to 44 candidates. We tested the random community assembly of the final candidates using different metrics: (i) Cmean and Pagel's Lambda, implemented in the phylosignal R package [45], and the standardized effect size (SES) of (ii) mean pairwise distance (MPD) and mean nearest taxon distance (MNTD), implemented in the picante R package [46]. We retrieved the coding sequence (CDS) of each protein accession from the NCBI Identical Protein Groups database [44].

To assess the expression profile of flagellins in the human gut, we retrieved publicly available gut metatranscriptome data (available at https://ibdmdb.org/tunnel/public/HMP2/MTX/1750/products) [18]. We restricted our assessment to samples from the 26 healthy controls with associated metatranscriptomes in [18], selecting one sample (corresponding to their earliest visit) per individual. We filtered the downloaded HUMAnN2 tables [47] by retaining features annotated as flagellin and whose abundance could be attributed to a specific taxon by HUMAnN2's tiered search. We further validated that the retained flagellins contained the N-terminal and C-terminal domains. HUMAnN2's tiered search provides NCBI taxonomy annotation; we reannotated the contributing species by matching the NCBI taxonomy annotation to that of GTDB release 95 [48].

*Rh*FlaB-like candidates were identified using the flagellin database previously described. A truncated CD protein sequence of *Rh*FlaB (accession: WP_014081191.1) was mapped against the database using DIAMOND blastp [49] with parameters "--max-target-seqs 0 --evalue 10-3 --very-sensitive." The resulting 10,121 hits were filtered by (i) the median of the length of the alignment (100 amino acids), (ii) the sequence identity (39%), (iii) and the number of mismatches (n = 60), which resulted in 979 protein sequences. We mapped the ND flagellin from *Salmonella Typhimurium* (accession: AHA06007.1) against the selected hits with DIAMOND blastp using parameters "--max-target-seqs 0 --evalue 10-3 --very-sensitive," which reduced the list to 919 protein sequences. The resulting list was manually curated by selecting those sequences that did not have either amino acids arginine or lysine in position 478 of the alignment, leaving a total of 195 sequences.

To reduce the *Rh*FlaB-like flagellins to only those occurring in the human gut, we mapped gut metagenomes from the IBD multi-omics database [18] and the Franzosa et al. [50] dataset to our flagellin database using DIAMOND blastx with parameters "--max-target-seqs 1 --evalue 1e-3 --very-sensitive." The resulting 5131 protein accessions were intersected with the 195 flagellins from the former step, resulting in 145 flagellins that meet the sequence composition criteria and are also present in the human gut. From these, we dereplicated at 99% sequence identity with CD-HIT [51,52] using the parameters "-c 0.99 -M 16000 -n 5," which reduced the dataset to 85 protein accessions. We retrieved their CDSs from the NCBI Identical Protein Groups database [44]. To avoid nucleotide sequence redundancy with the previously identified abundant flagellins, we merged the lists of CDSs, and the final nucleotide sequence list was dereplicated using CD-HIT with parameters "-c 0.99 -M 16000 -n 5."

2.4.9. Estimating abundance of silent, stimulator, and evader flagellins in human gut metagenomes

We used metagenomes from the curatedMetagenomicData dataset [53]. Samples were selected on the basis of the following criteria: (i) shotgun metagenomes sequenced using the Illumina HiSeq platform with a median read length of >95 base pairs; (ii) available Sequence Read Archive accessions; (iii) labeled as adults or seniors or with a reported age of ≥18 years; (iv) without report of antibiotic consumption; (v) without report of pregnancy; (vi) nonlactating women; and (vii) labeled as "healthy" (based on random forest algorithm described in [53]). In cases of multiple samples per individual, one sample was randomly selected. The final dataset consisted of 1783 samples from 21 studies (see Supp. Data 2.2). Only the forward reads were used, and all metagenomes were subsampled to 1 million reads.

We mapped the metagenome reads to all stimulator, silent, and evader flagellins (n = 126) via DIAMOND blastx with the following parameters: --sensitive --iterate --evalue 1e-5 --query-cover 90. The results were used to calculate reads per kilobase of target per million reads (RPKM) for each flagellin in each sample. Cohen's *d* was calculated with the effsize R package [54].

2.4.10. NF-kB activation assay

HEK-Blue hTLR5 cells were plated in six-well plates and grown overnight to 90% confluency in medium (GlutaMAX DMEM and 10% FBS; Gibco). Cells were washed with 1 ml of Hanks' balanced salt solution (HBSS; Gibco #14025092) and then incubated with flagellin (10 nM), TNF-α (10 ng/μL; CST #16769), or buffer diluted in 1 ml of HBSS containing 50 μM (R)-MG132 (Sigma-Aldrich #M8699) for 1 hour at 37°C. Cells were harvested and resuspended in 200 μL of lysis buffer (25 mM Tris-HCl (pH 7.5), 1% (v/v) Igepal CA-630, 100 mM NaCl, and 10% (v/v) glycerol) supplemented with 50 μM MG132 and protease inhibitors. After incubating on ice for 15 min, cells were centrifuged at 4°C (855g) for 20 min. Supernatants were collected, protein levels were quantified by BCA, and 40 μg of lysates were diluted in Laemmli loading buffer for analysis by immunoblotting.

2.4.11. Immunoblotting

Protein samples were diluted in 1× Laemmli loading buffer (50 mM Tris-HCl (pH 6.8), 2% (w/v) SDS, 2% (v/v) glycerol, 0.05% (w/v) bromophenol blue, and 2.5% (v/v) 2-mercaptoethanol) and incubated at 95°C for 5 min. Samples and Chameleon Duo Pre-stained protein ladder (LI-COR) were loaded in 4 to 12% NuPAGE Bis-Tris protein gels (Invitrogen) and run in Mini Gel Tank (Invitrogen) with MES buffer (50 mM MES, 50 mM Tris (pH 7.3), 0.1% (w/v) SDS, and 1 mM EDTA) at 125 V. After electrophoresis, gels were transferred to nitrocellulose membranes (Pierce) at 10 V for 70 min at 4°C in transfer buffer (25 mM bicine, 25 mM Bis-Tris (pH 7.2), 1 mM EDTA, and 10% (v/v) methanol) using the Mini Blot Module (Invitrogen). Membranes were blocked for 30 min with 5% (w/v) milk in TBS-T (19 mM tris (pH 7.6), 137 mM NaCl, 2.7 mM KCl, and 0.1% (v/v) Tween 20) and then incubated in primary antibody for 1 hour at room temperature or overnight at 4°C. Membranes were washed 3× for 5 min with TBS-T and then incubated with secondary antibody for 30 min at room temperature. Membranes were washed 3× before imaging on the Odyssey CLx (LI-COR). Unless stated otherwise, primary and secondary antibodies were diluted in TBS-T with 5% (w/v) milk to 0.1 μg/ml (anti-HA), 0.5 μg/ml (anti-Myc), and 0.05 μg/ml (IRDyes). For NF-κB activation immunoblots, primary antibodies were diluted 1:1000 in TBS-T with 5% (w/v) BSA.
Band intensities from immunoblots were quantified in Image Studio (LI-COR). For pull-downs, the fraction of TLR5 bound to Myc-tagged bait was calculated by first subtracting the background HA signal and then dividing by the Myc signal.

2.4.12. Stool sample preparation for proteomics

A stool sample was previously obtained from an adult female without known disease (Cornell University Institutional Review Board, protocol number 1106002281). Stool (166 mg) was resuspended in PBS (pH 7.4) supplemented with protease inhibitors to 0.03% (w/v) and vortexed at maximum speed for 20 min. Lysates were cleared after centrifugation at 13,700g for 10 min at 4°C. Protein levels in the supernatant were quantified by BCA, aliquoted, and stored at −80°C. TLR5-HA HEK lysates were prepared from ~6.5 × 106 cells as described above in lysis buffer lacking imidazole and incubated with 90 µL of anti-HA resin for 2 hours at 4°C. Beads were washed 3× with lysis buffer before incubation with 1.8 mg of stool proteins for 2 hours at 4°C. Before incubation, stool proteins were boiled for 5 min at 98°C and diluted in lysis buffer to a final volume of 300 µL. After incubation, beads were washed 3× with lysis buffer and bound proteins were analysed by MS.

2.4.13. Flagellin database for proteomics

We extracted total DNA from the above stool sample using the QIAGEN PowerSoil kit following the manufacturer's instructions. Library preparation and sequencing were performed as described above. After quality control, the sequencing depth of the sample was 210,366,585 paired reads. We mapped the sequencing reads to our curated flagellin database using DIAMOND v.2.0.9 [55] using parameters `--evalue 1e-3 --ultra-sensitive`. The abundance of mapped reads was transformed to reads per kilobase (RPK). We assigned taxonomy to the flagellin hits by matching the NCBI protein ID and its associated assembly to GTDB r95. The final database contained 1028 flagellin sequences.

2.4.14. NanoLC-MS/MS analysis and MS data processing

Proteins were eluted from the washed beads and purified with a 12% NuPAGE Novex bis-tris gel (Invitrogen). Tryptic in-gel digestion of proteins was performed as described previously [56], and extracted peptides were desalted using C18 StageTips [57]. Eluted peptides were subjected to liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis.

Peptide analysis was performed on an Easy-nLC 1200 system coupled to an Exploris 480 mass spectrometer (Thermo Fisher Scientific) as described elsewhere [58] with slight modifications: Peptides were injected onto the column in high-performance liquid chromatography (HPLC) solvent A (0.1% formic acid) at a flow rate of 500 nl/min and subsequently eluted with a 227-min segmented gradient of 10-33-50-90% HPLC solvent B (80% acetonitrile in 0.1% formic acid). During peptide elution, the flow rate was kept constant at 200 nl/min.

In each scan cycle, the 20 most intense precursor ions were sequentially fragmented using higher-energy collisional dissociation (HCD) fragmentation. For both precursors and fragment ions, the maximal injection time mode was set to auto and the AGC target was set to standard. Precursor masses with charge states between 2 and 6 were selected with a resolution of 60,000 at a minimum intensity of 400,000 at a scan range of 300 to 1750 mass/charge ratio (m/z). They were excluded from further selection for 30 s. HCD collision energy for peptide fragmentation was set to 28%, and resolution for MS2 scans was set to 30,000.

MS data were processed with MaxQuant software suite version 1.6.7.0 [59], and database search was performed using the Andromeda search engine [60], a module of the MaxQuant. MS/MS spectra were searched against a Homo sapiens database obtained from UniProt (released 10 July 2020, 97,795 entries), a database containing 1028 flagellin sequences from various organisms (European Nucleotide Archive, study accession number: PRJEB47632), and a database consisting of 246 commonly observed contaminants. In database search, full tryptic specificity was required and up to two missed cleavages were allowed. Carbamidomethylation of cysteine was set as fixed modification, whereas oxidation of methionine and acetylation of protein N terminus were set as variable modifications. Mass tolerance for precursor ions was set to 4.5 parts per million (ppm) and to 20 ppm for fragment ions. Peptide, protein, and modification site identifications were reported at a false discovery rate of 0.01, estimated by the target/decoy approach [61]. For protein group quantitation, a minimum of two quantified peptides were required. All search parameters were kept to default values except for the following: Minimal peptide length of 5 amino acids was required, and the iBAQ algorithm was used to estimate quantitative values by dividing the sum of peptide intensities of all detected peptides by the number of theoretically observable peptides of the matched protein [62]. See Supp. Data 2.1.

2.4.15. Organoid experiments

Organoids derived from human colon were cultured in basement membrane extract (Cultrex PathClear Reduced Growth Factor, Type 2) and growth medium (Advanced DMEM/F12 supplemented with 50% (v/v) L-WRN conditioned medium described in [63], 1 mM Hepes (Gibco), 1× GlutaMAX (Gibco), 1× B27 (Gibco), 1 mM N-acetylcysteine, 10 nM gastrin (Sigma-Aldrich), epidermal growth factor (50 ng/ml; PeproTech), 10 mM nicotinamide, 500 nM A83-01 (Tocris), 10 µM SB202190,10 µM Y27632 (Tocris), and 250 nM CHIR99021 (Tocris)) for 4 days in 96-well plates. Organoids were incubated for 18 hours in the presence of flagellins (10 nM) or buffer control diluted in growth medium at 37°C. IL-8 cytokine levels were quantified from 20 µL of medium according to manufacturer's instructions (Human IL8 DuoSet enzyme-linked immunosorbent assay (ELISA); R&D Systems #DY208).

2.4.16. Mouse experiments

All animal studies were performed at Georgia State University under an approved animal protocol (Institutional Animal Care and Use Committee, no. A17047). Eight-week-old female or male C57BL/6 WT and *Tlr5−/−/Nlrc4−/−* mice, bred at Georgia State University, were treated with 10 or 20 µg of *St*FliC, *Rh*FlaB, or *Rh*FlaB-*St*FliC D0 by intraperitoneal injection. Two hours later, blood was collected from these mice via retrobulbar intraorbital capillary plexus, and serum was isolated by centrifugation at 4°C using MiniCollect serum separator tubes (Greiner Bio-One). Cxcl1, IL-6, and IL-18 in serum were quantitated using ELISA kits from R&D Systems (#DY453, #DY406, and #7625, respectively) according to the manufacturer's instructions. Absorbance values lower than blank controls were set to zero.

2.4.17. General data analysis and visualization

Data were processed, analysed, and visualised with snakemake [64], conda [65], and R [66] and the following R packages: dplyr [67], tidyr [68], ggplot2 [69], and ggpubr [70]. Protein Data Bank (PDB) structure was color-coded using PyMOL [71].

2.4.18. Statistics

Statistical analyses on mouse, organoid, and *Rh*FlaB mutant data were performed in Prism 9 (GraphPad). Differences between groups were assessed by unpaired, two-tailed *t* test. Unless stated otherwise, all other analyses were performed in R. We defined $P < 0.05$ as significant.

2.5. References

1. C. A. Janeway Jr, Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harb. Symp. Quant. Biol. 54 Pt 1, 1–13 (1989).

2. I. Botos, D. M. Segal, D. R. Davies, The structural biology of Toll-like receptors. Structure 19, 447–459 (2011).

3. R. Medzhitov, Toll-like receptors and innate immunity. Nat. Rev. Immunol. 1, 135–145 (2001).

4. B. Chassaing, R. E. Ley, A. T. Gewirtz, Intestinal epithelial cell toll-like receptor 5 regulates the intestinal microbiota to prevent low-grade inflammation and metabolic syndrome in mice. Gastroenterology 147, 1363–77.e17 (2014).

5. F. Hayashi, K. D. Smith, A. Ozinsky, T. R. Hawn, E. C. Yi, D. R. Goodlett, J. K. Eng, S. Akira, D. M. Underhill, A. Aderem, The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature 410, 1099–1103 (2001).

6. S. Maki-Yonekura, K. Yonekura, K. Namba, Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. Nat. Struct. Mol. Biol. 17, 417–422 (2010).

7. K. D. Smith, E. Andersen-Nissen, F. Hayashi, K. Strobe, M. A. Bergman, S. L. R. Barrett, B. T. Cookson, A. Aderem, Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. Nat. Immunol. 4, 1247–1253 (2003).

8. M. A. B. Kreutzberger, C. Ewing, F. Poly, F. Wang, E. H. Egelman, Atomic structure of the Campylobacter jejuni flagellar filament reveals how ε Proteobacteria escaped Toll-like receptor 5 surveillance. Proc. Natl. Acad. Sci. U. S. A. 117, 16985–16991 (2020).

9. S.-I. Yoon, O. Kurnasov, V. Natarajan, M. Hong, A. V. Gudkov, A. L. Osterman, I. A. Wilson, Structural basis of TLR5-flagellin recognition and signaling. Science 335, 859–864 (2012).

10. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

11. E. Andersen-Nissen, K. D. Smith, K. L. Strobe, S. L. R. Barrett, B. T. Cookson, S. M. Logan, A. Aderem, Evasion of Toll-like receptor 5 by flagellated bacteria. Proc. Natl. Acad. Sci. U. S. A. 102, 9247–9252 (2005).

12. T. C. Cullender, B. Chassaing, A. Janzon, K. Kumar, C. E. Muller, J. J. Werner, L. T. Angenent, M. E. Bell, A. G. Hay, D. A. Peterson, J. Walter, M. Vijay-Kumar, A. T. Gewirtz, R. E. Ley, Innate and adaptive immunity interact to quench microbiome flagellar motility in the gut. Cell Host Microbe 14, 571–581 (2013).

13. T. Steiner, S. Ivison, C. Wang, C. Elson, The A4-Fla2 flagellin, a dominant antigen in Crohn's Disease, is a poor TLR5 agonist: P-0163. Inflamm. Bowel Dis. 15, S55 (2009).

14. B. A. Neville, P. O. Sheridan, H. M. B. Harris, S. Coughlan, H. J. Flint, S. H. Duncan, I. B. Jeffery, M. J. Claesson, R. P. Ross, K. P. Scott, P. W. O'Toole, Pro-inflammatory flagellin proteins of prevalent motile commensal bacteria are variably abundant in the intestinal microbiome of elderly humans. PLoS One 8, e68919 (2013).

15. A. Almeida, S. Nayfach, M. Boland, F. Strozzi, M. Beracochea, Z. J. Shi, K. S. Pollard, E. Sakharova, D. H. Parks, P. Hugenholtz, N. Segata, N. C. Kyrpides, R. D. Finn, A unified catalog of 204,938 reference genomes from the human gut microbiome. Nat. Biotechnol. 39, 105–114 (2021).

16. K. L. Alexander, Q. Zhao, M. Reif, A. F. Rosenberg, P. J. Mannon, L. W. Duck, C. O. Elson, Human Microbiota Flagellins Drive Adaptive Immune Responses in Crohn's Disease. Gastroenterology (2021), doi:10.1053/j.gastro.2021.03.064.

17. P. Louis, H. J. Flint, Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol. Lett. 294, 1–8 (2009).

18. J. Lloyd-Price, C. Arze, A. N. Ananthakrishnan, M. Schirmer, J. Avila-Pacheco, T. W. Poon, E. Andrews, N. J. Ajami, K. S. Bonham, C. J. Brislawn, D. Casero, H. Courtney, A. Gonzalez, T. G. Graeber, A. B. Hall, K. Lake, C. J. Landers, H. Mallick, D. R. Plichta, M. Prasad, G. Rahnavard, J. Sauk, D. Shungin, Y. Vázquez-Baeza, R. A. White 3rd, IBDMDB Investigators, J. Braun, L. A. Denson, J. K. Jansson, R. Knight, S. Kugathasan, D. P. B. McGovern, J. F. Petrosino, T. S. Stappenbeck, H. S. Winter, C. B. Clish, E. A. Franzosa, H. Vlamakis, R. J. Xavier, C. Huttenhower, Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569, 655–662 (2019).

19. V. Forstnerič, K. Ivičak-Kocjan, T. Plaper, R. Jerala, M. Benčina, The role of the C-terminal D0 domain of flagellin in activation of Toll like receptor 5. PLoS Pathog. 13, e1006574 (2017).

20. A. M. Patterson, I. E. Mulder, A. J. Travis, A. Lan, N. Cerf-Bensussan, V. Gaboriau-Routhiau, K. Garden, E. Logan, M. I. Delday, A. G. P. Coutts, E. Monnais, V. C. Ferraria, R. Inoue, G. Grant, R. I. Aminov, Human Gut Symbiont Roseburia hominis Promotes and Regulates Innate Immunity. Front. Immunol. 8, 1166 (2017).

21. S. H. Duncan, R. I. Aminov, K. P. Scott, P. Louis, T. B. Stanton, H. J. Flint, Proposal of Roseburia faecis sp. nov., Roseburia hominis sp. nov. and Roseburia inulinivorans sp. nov., based on isolates from human faeces. Int. J. Syst. Evol. Microbiol. 56, 2437–2441 (2006).

22. S. N. Klimosch, A. Försti, J. Eckert, J. Knezevic, M. Bevier, W. von Schönfels, N. Heits, J. Walter, S. Hinz, J. Lascorz, J. Hampe, D. Hartl, J.-S. Frick, K. Hemminki, C. Schafmayer, A. N. R. Weber, Functional TLR5 genetic variants affect human colorectal cancer survival. Cancer Res. 73, 7232–7242 (2013).

23. A. T. Gewirtz, T. A. Navas, S. Lyons, P. J. Godowski, J. L. Madara, Cutting edge: bacterial flagellin activates basolaterally expressed TLR5 to induce epithelial proinflammatory gene expression. J. Immunol. 167, 1882–1885 (2001).

24. K. Ivičak-Kocjan, G. Panter, M. Benčina, R. Jerala, Determination of the physiological 2:2 TLR5:flagellin activation stoichiometry revealed by the activity of a fusion receptor. Biochem. Biophys. Res. Commun. 435, 40–45 (2013).

25. M.-H. Khani, M. Bagheri, A. Dehghanian, A. Zahmatkesh, S. Moradi Bidhendi, Z. Salehi Najafabadi, R. Banihashemi, Effect of C-Terminus Modification in Salmonella typhimurium FliC on Protein Purification Efficacy and Bioactivity. Mol. Biotechnol. 61, 12–19 (2019).

26. I. A. Hajam, P. A. Dar, I. Shahnawaz, J. C. Jaume, J. H. Lee, Bacterial flagellin-a potent immunomodulatory agent. Exp. Mol. Med. 49, e373 (2017).

27. A. H. López-Yglesias, X. Zhao, E. K. Quarles, M. A. Lai, T. VandenBos, R. K. Strong, K. D. Smith, Flagellin induces antibody responses through a TLR5- and inflammasome-independent pathway. J. Immunol. 192, 1587–1596 (2014).

28. K. Zhou, R. Kanai, P. Lee, H.-W. Wang, Y. Modis, Toll-like receptor 5 forms asymmetric dimers in the absence of flagellin. J. Struct. Biol. 177, 402–409 (2012).

29. D. Hu, P. R. Reeves, The Remarkable Dual-Level Diversity of Prokaryotic Flagellins. mSystems 5 (2020), doi:10.1128/mSystems.00705-19.

30. M. J. Lodes, Y. Cong, C. O. Elson, R. Mohamath, C. J. Landers, S. R. Targan, M. Fort, R. M. Hershberg, Bacterial flagellin is a dominant antigen in Crohn disease. J. Clin. Invest. 113, 1296–1306 (2004).

31. A. T. Gewirtz, P. O. Simon Jr, C. K. Schmitt, L. J. Taylor, C. H. Hagedorn, A. D. O'Brien, A. S. Neish, J. L. Madara, Salmonella typhimurium translocates flagellin across intestinal epithelia, inducing a proinflammatory response. J. Clin. Invest. 107, 99–109 (2001).

32. L. Franchi, A. Amer, M. Body-Malapel, T.-D. Kanneganti, N. Ozören, R. Jagirdar, N. Inohara, P. Vandenabeele, J. Bertin, A. Coyle, E. P. Grant, G. Núñez, Cytosolic flagellin requires Ipaf for activation of caspase-1 and interleukin 1beta in salmonella-infected macrophages. Nat. Immunol. 7, 576–582 (2006).

33. E. A. Miao, C. M. Alpuche-Aranda, M. Dors, A. E. Clark, M. W. Bader, S. I. Miller, A. Aderem, Cytoplasmic flagellin activates caspase-1 and secretion of interleukin 1beta via Ipaf. Nat. Immunol. 7, 569–575 (2006).

34. M. Vijay-Kumar, F. A. Carvalho, J. D. Aitken, N. H. Fifadara, A. T. Gewirtz, TLR5 or NLRC4 is necessary and sufficient for promotion of humoral immunity by flagellin. Eur. J. Immunol. 40, 3528–3534 (2010).

35. E. Latz, A. Verma, A. Visintin, M. Gong, C. M. Sirois, D. C. G. Klein, B. G. Monks, C. J. McKnight, M. S. Lamphier, W. P. Duprex, T. Espevik, D. T. Golenbock, Ligand-induced conformational changes allosterically activate Toll-like receptor 9. Nat. Immunol. 8, 772–779 (2007).

36. F. Weissmann, G. Petzold, R. VanderLinden, P. J. Huis In 't Veld, N. G. Brown, F. Lampert, S. Westermann, H. Stark, B. A. Schulman, J.-M. Peters, biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. Proc. Natl. Acad. Sci. U. S. A. 113, E2564–9 (2016).

37. V. Altmannova, A. Blaha, S. Astrinidis, H. Reichle, J. R. Weir, InteBac: An integrated bacterial and baculovirus expression vector suite. Protein Sci. 30, 108–114 (2021).

38. C. Bieniossek, T. Imasaki, Y. Takagi, I. Berger, MultiBac: expanding the research toolbox for multiprotein complexes. Trends Biochem. Sci. 37, 49–57 (2012).

39. E. Smakowska-Luzan, G. A. Mott, K. Parys, M. Stegmann, T. C. Howton, M. Layeghifard, J. Neuhold, A. Lehner, J. Kong, K. Grünwald, N. Weinberger, S. B. Satbhai, D. Mayer, W. Busch, M. Madalinski, P. Stolt-Bergner, N. J. Provart, M. S. Mukhtar, C. Zipfel, D. Desveaux, D. S. Guttman, Y. Belkhadir, An extracellular network of Arabidopsis leucine-rich repeat receptor kinases. Nature 553, 342–346 (2018).

40. K. Parys, N. R. Colaianni, H.-S. Lee, U. Hohmann, N. Edelbacher, A. Trgovcevic, Z. Blahovska, D. Lee, A. Mechtler, Z. Muhari-Portik, M. Madalinski, N. Schandry, I. Rodríguez-Arévalo, C. Becker, E. Sonnleitner, A. Korte, U. Bläsi, N. Geldner, M. Hothorn, C. D. Jones, J. L. Dangl, Y. Belkhadir, Signatures of antagonistic pleiotropy in a bacterial flagellin epitope. Cell Host Microbe 29, 620–634.e9 (2021).

41. P. Jones, D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, S. Hunter, InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240 (2014).

42. S. Sherrill-Mix, Functions to Work with NCBI Accessions and Taxonomy [R package taxonomizr version 0.8.0]. (2021) (available at https://CRAN.R-project.org/package=taxonomizr).

43. J. Kans, Entrez Direct: E-utilities on the Unix Command Line (National Center for Biotechnology Information (US), 2022).

44. D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, J. Ostell, K. D. Pruitt, E. W. Sayers, GenBank. Nucleic Acids Res. 46, D41–D47 (2018).

45. F. Keck, F. Rimet, A. Bouchez, A. Franc, phylosignal: an R package to measure, test, and explore the phylogenetic signal. Ecol. Evol. 6, 2774–2780 (2016).

46. S. W. Kembel, P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, C. O. Webb, Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464 (2010).

47. E. A. Franzosa, L. J. McIver, G. Rahnavard, L. R. Thompson, M. Schirmer, G. Weingart, K. S. Lipson, R. Knight, J. G. Caporaso, N. Segata, C. Huttenhower, Species-level functional profiling of metagenomes and metatranscriptomes. Nat. Methods 15, 962–968 (2018).

48. D. H. Parks, M. Chuvochina, D. W. Waite, C. Rinke, A. Skarshewski, P.-A. Chaumeil, P. Hugenholtz, A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36, 996–1004 (2018).

49. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60 (2015).

50. E. A. Franzosa, A. Sirota-Madi, J. Avila-Pacheco, N. Fornelos, H. J. Haiser, S. Reinker, T. Vatanen, A. B. Hall, H. Mallick, L. J. McIver, J. S. Sauk, R. G. Wilson, B. W. Stevens, J. M.

Scott, K. Pierce, A. A. Deik, K. Bullock, F. Imhann, J. A. Porter, A. Zhernakova, J. Fu, R. K. Weersma, C. Wijmenga, C. B. Clish, H. Vlamakis, C. Huttenhower, R. J. Xavier, Gut microbiome structure and metabolic activity in inflammatory bowel disease. Nat Microbiol 4, 293–305 (2019).

51. W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22, 1658–1659 (2006).

52. L. Fu, B. Niu, Z. Zhu, S. Wu, W. Li, CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152 (2012).

53. E. Pasolli, L. Schiffer, P. Manghi, A. Renson, V. Obenchain, D. T. Truong, F. Beghini, F. Malik, M. Ramos, J. B. Dowd, C. Huttenhower, M. Morgan, N. Segata, L. Waldron, Accessible, curated metagenomic data through ExperimentHub. Nat. Methods 14, 1023–1024 (2017).

54. M. Torchiano, Efficient Effect Size Computation [R package effsize version 0.8.1]. (2020) (available at https://CRAN.R-project.org/package=effsize).

55. B. Buchfink, K. Reuter, H.-G. Drost, Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods 18, 366–368 (2021).

56. N. Borchert, C. Dieterich, K. Krug, W. Schütz, S. Jung, A. Nordheim, R. J. Sommer, B. Macek, Proteogenomics of Pristionchus pacificus reveals distinct proteome structure of nematode models. Genome Res. 20, 837–846 (2010).

57. J. Rappsilber, M. Mann, Y. Ishihama, Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. 2, 1896–1906 (2007).

58. M. Schmitt, T. Sinnberg, K. Bratl, K. Zittlau, C. Garbe, B. Macek, N. C. Nalpas, Proteogenomics Reveals Perturbed Signaling Networks in Malignant Melanoma Cells Resistant to BRAF Inhibition. Mol. Cell. Proteomics 20, 100163 (2021).

59. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367–1372 (2008).

60. J. Cox, N. Neuhauser, A. Michalski, R. A. Scheltema, J. V. Olsen, M. Mann, Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. 10, 1794–1805 (2011).

61. J. E. Elias, S. P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat. Methods 4, 207–214 (2007).

62. B. Schwanhäusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, M. Selbach, Global quantification of mammalian gene expression control. Nature 473, 337–342 (2011).

63. H. Miyoshi, T. S. Stappenbeck, In vitro expansion and genetic modification of gastrointestinal stem cells in spheroid culture. Nat. Protoc. 8, 2471–2482 (2013).

64. F. Mölder, K. P. Jablonski, B. Letcher, M. B. Hall, C. H. Tomkins-Tinch, V. Sochat, J. Forster, S. Lee, S. O. Twardziok, A. Kanitz, A. Wilm, M. Holtgrewe, S. Rahmann, S. Nahnsen, J. Köster, Sustainable data analysis with Snakemake. F1000Res. 10, 33 (2021).

65. Anaconda Software Distribution Anaconda Documentation (2020) (available at https://docs.anaconda.com/).

66. R Core Team, R: A Language and Environment for Statistical Computing (2018) (available at https://www.R-project.org/).

67. H. Wickham, R. François, L. Henry, K. Müller, dplyr: A Grammar of Data Manipulation (2021) (available at https://CRAN.R-project.org/package=dplyr).

68. H. Wickham, tidyr: Tidy Messy Data (2021) (available at https://CRAN.R-project.org/package=tidyr).

69. H. Wickham, ggplot2: Elegant Graphics for Data Analysis (2016) (available at https://ggplot2.tidyverse.org).

70. A. Kassambara, ggpubr: “ggplot2” Based Publication Ready Plots (2020) (available at https://CRAN.R-project.org/package=ggpubr).

71. L. L. C. Schrödinger, W. DeLano, PyMOL (2020; http://www.pymol.org/pymol).

72. Y. Perez-Riverol, J. Bai, C. Bandla, D. García-Seisdedos, S. Hewapathirana, S. Kamatchinathan, D. J. Kundu, A. Prakash, A. Frericks-Zipper, M. Eisenacher, M. Walzer, S. Wang, A. Brazma, J. A. Vizcaíno, The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 50, D543–D552 (2022)

2.6. Further statements

2.6.1. Acknowledgments:

Human colon organoids from the Stanford Tissue Bank were provided by J. Y. Co from the laboratories of M. R. Amieva and D. Monack. We thank I. Droste-Borel and B. Macek from the Proteome Center Tübingen, C. Lanz and O. Weichenrieder from the Genome Center at the MPI for Biology, and I. Hang and J. Neuhold at the Vienna BioCenter Core Facilities (VBCF ProTech).

2.6.2. Funding statement

This work was supported by the Max Planck Society and grants from the Austrian Academy of Sciences through the Gregor Mendel Institute and the Vienna Science and Technology Fund Project (LS17-047) (Y.B.).

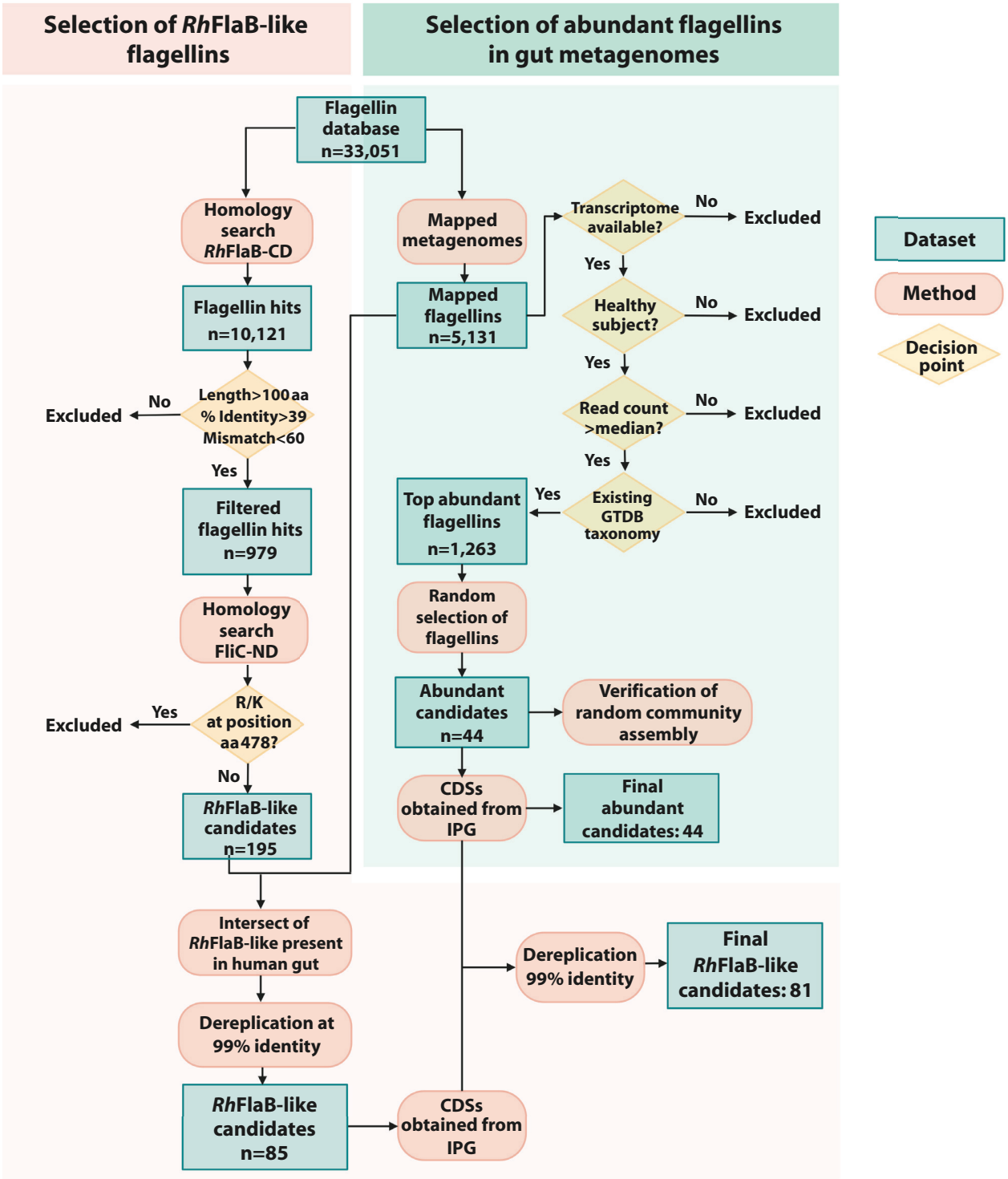
2.6.3. Competing interests

Authors declare that they have no competing interests.

2.6.4. Data and materials availability

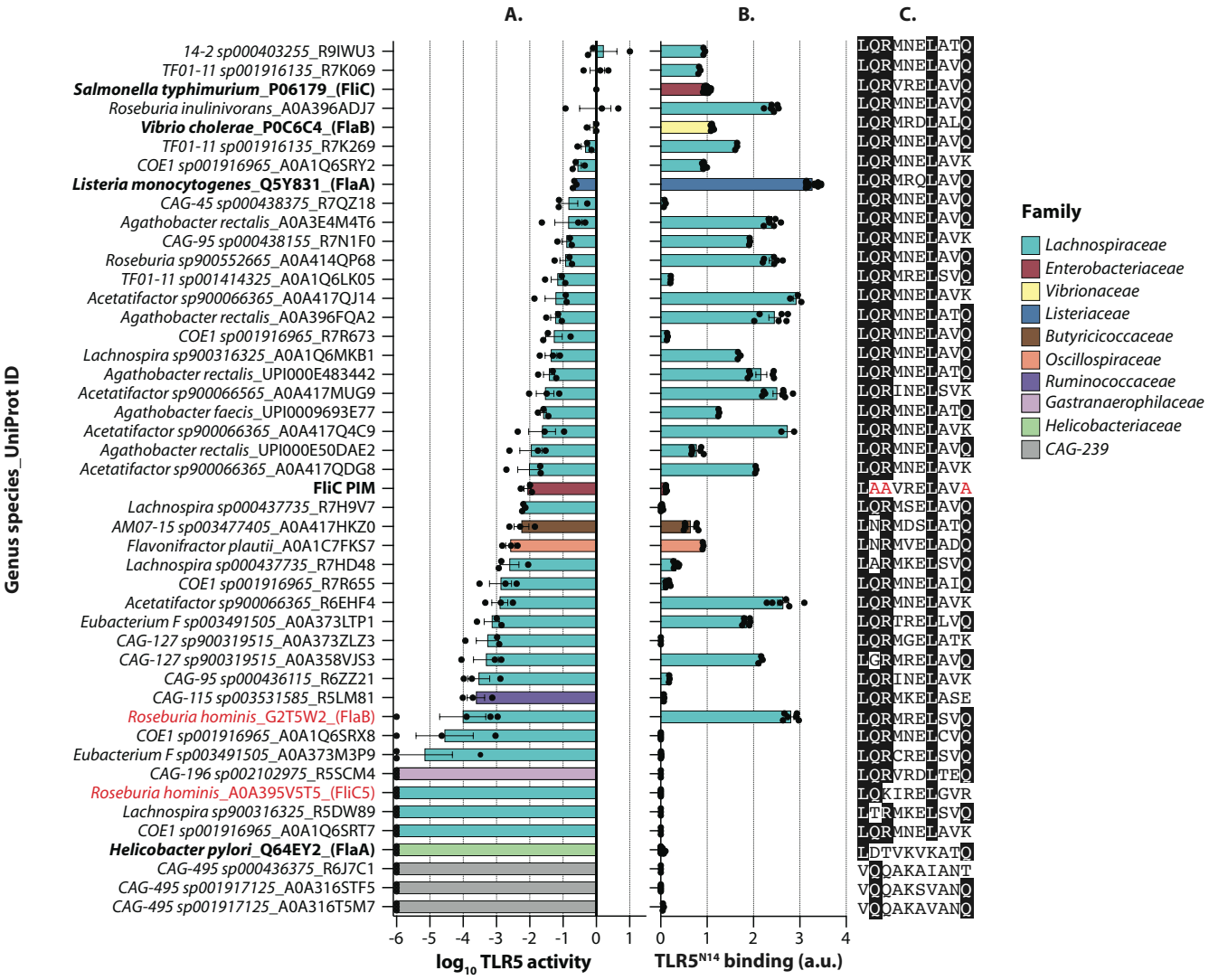
Raw sequence data of samples used for proteomics are available from European Nucleotide Archive (study accession number PRJEB47632). MS proteomics data are deposited to the ProteomeXchange Consortium via the PRIDE partner repository with dataset identifier PXD033249 (72). All other data needed to evaluate the conclusions in the paper are available in the paper or the Supplementary Materials.

2.7. Supplementary figures



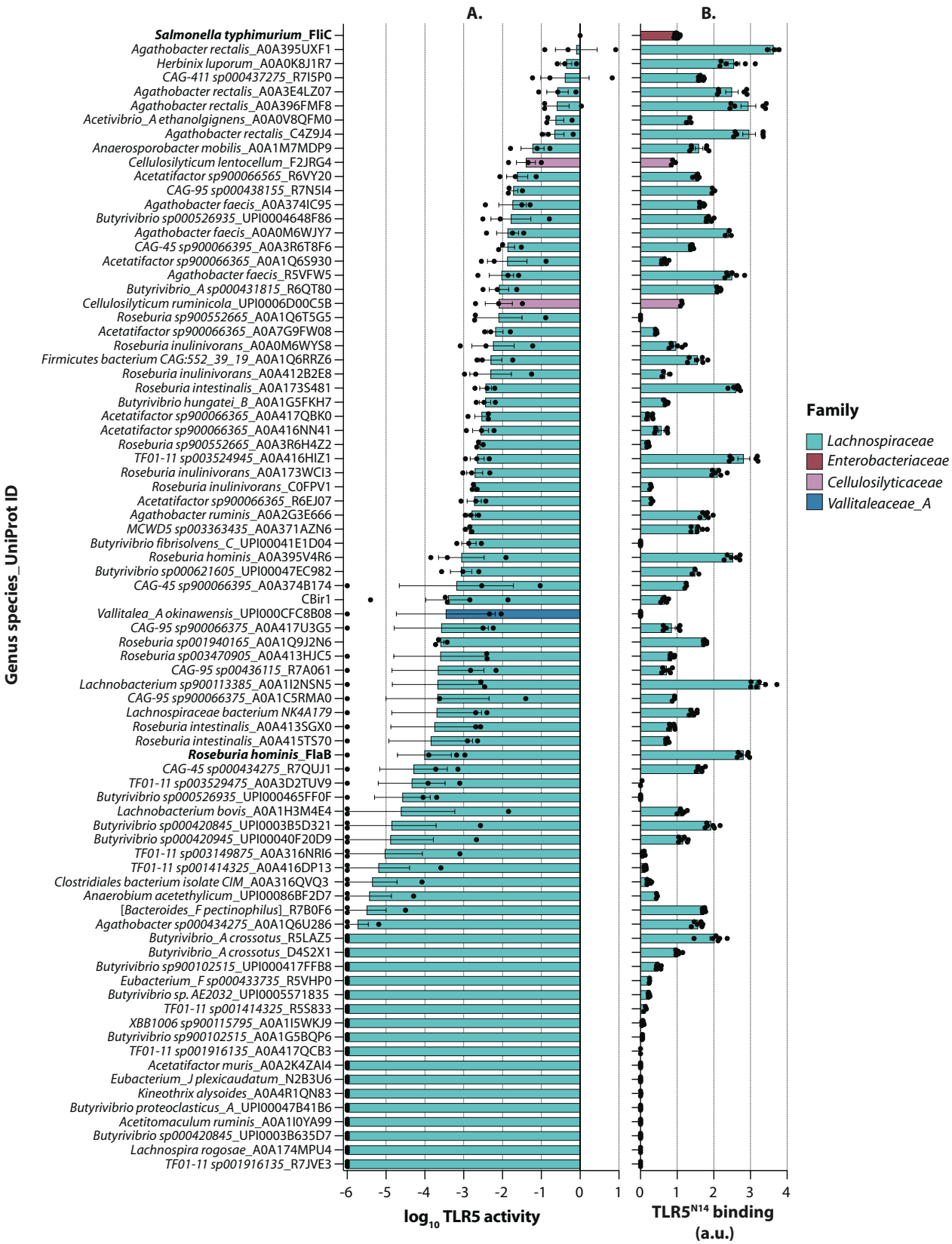
Supplementary Figure 2.1 - Selection of flagellin candidates.

The flowchart illustrates the strategy used to identify abundant flagellins (n = 44) and *Rh*FlaB-like flagellins occurring in the human gut (n = 81). FliC = *St*FliC. Left: *Rh*FlaB-like flagellins were identified by homology search of truncated C-terminal domain (CD) and N-terminal domain (ND) of FlaB of *Roseburia hominis* and FliC of *Salmonella Typhimurium*, respectively. Only *Rh*FlaB-like flagellins occurring in human gut metagenomes were kept. Right: Abundant flagellins from gut metagenomes of healthy subjects were randomly selected among flagellins with a read count over the median and an existing taxonomy on GTDB v95. The coding sequences (CDSs) of *Rh*FlaB-like candidates were de-replicated with the CDSs of previously identified abundant flagellins to avoid redundancy. Six candidates were excluded due to poor expression in insect cells resulting in 41 highly abundant and 78 *Rh*FlaB-like candidates. See methods for additional details.



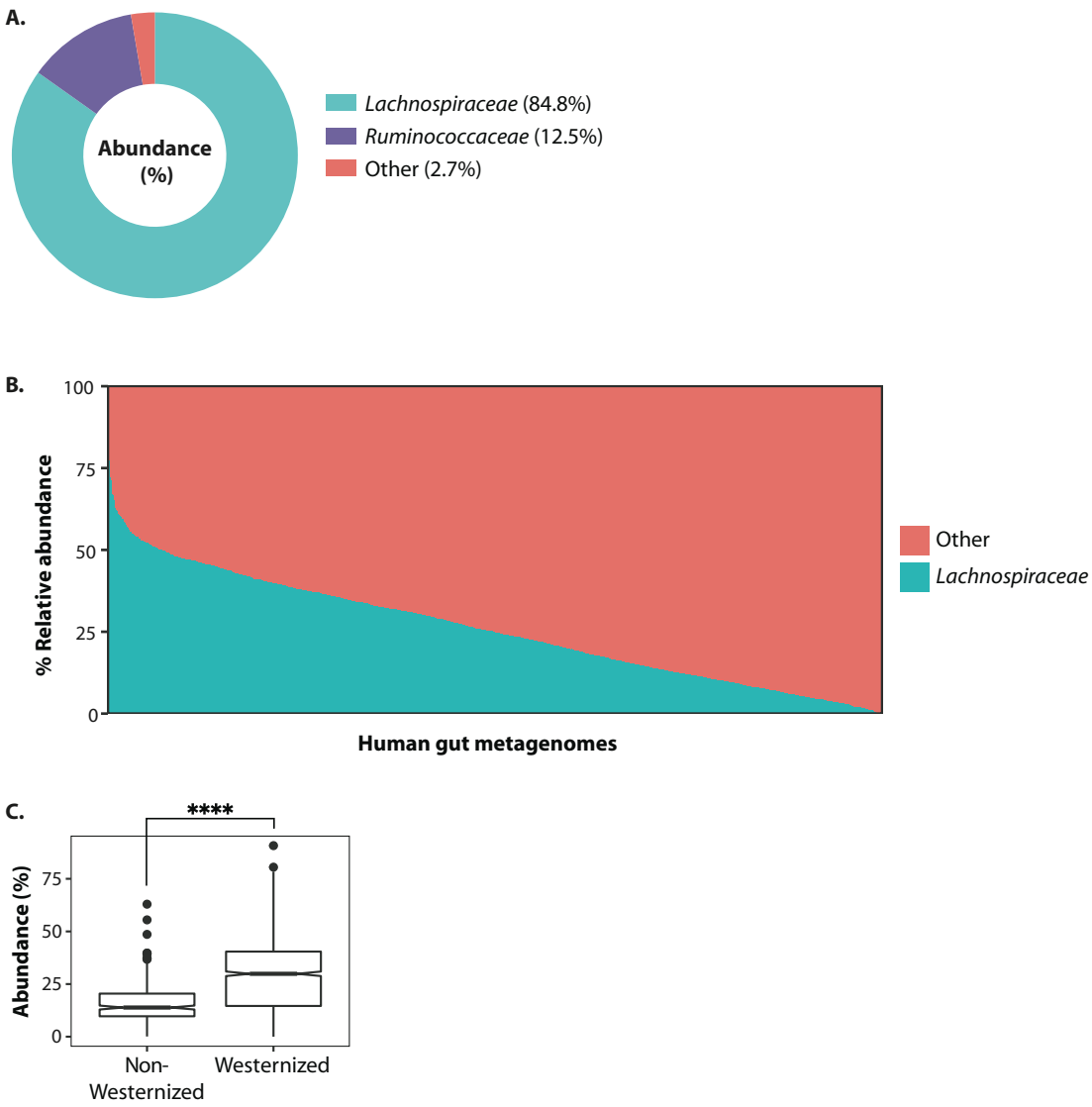
Supplementary Figure 2.2 - TLR5 recognition and activation by commensal-derived flagellins abundant in the human gut.

(A) Activation of TLR5: Myc-tagged flagellins were expressed in bacterial lysates and incubated with TLR5 HEK-Blue cells. TLR5 activity represents negative EC_{50} normalized to Myc expression. Data are mean ± SEM from at least three independent experiments. Values normalized to *St*FliC. FliC = *St*FliC. (B) Flagellin binding to $TLR5^{N14}$. AP-tagged flagellins were incubated with $TLR5^{N14}$ bait and AP activity was quantified. Data are mean ± SEM for n≥2 and normalized to *St*FliC. (C) TLR5 epitope sequences. Conserved residues required for TLR5 recognition are shaded black; residues mutated in *St*FliC-PIM are coloured red. Flagellins produced by *R. hominis* (*Rh*FlaB and *Rh*FliC5) are red; flagellins from pathogens are shown in bold. Bar colour indicates family-level taxonomy (GTDB). See Fig. 2.1D.


A.
B.
Genus species_UniProt ID
Salmonella typhimurium_FliC
Agathobacter rectalis_A0A395UXF1
Herbinix luporum_A0A0K8J1R7
CAG-411 sp000437275_R7I5P0
Agathobacter rectalis_A0A3E4LZ07
Agathobacter rectalis_A0A396FMF8
Acetivibrio_A ethanolgignens_A0A0V8QFM0
Agathobacter rectalis_C4Z9J4
Anaerosporobacter mobilis_A0A1M7MDP9
Cellulosilyticum lentocellum_F2JRG4
Acetatifactor sp900066565_R6VY20
CAG-95 sp000438155_R7N5I4
Agathobacter faecis_A0A374IC95
Butyrivibrio sp000526935_UPI0004648F86
Agathobacter faecis_A0A0M6WJY7
CAG-45 sp900066395_A0A3R6T8F6
Acetatifactor sp900066365_A0A1Q6S930
Agathobacter faecis_R5VFW5
Butyrivibrio_A sp000431815_R6QT80
Cellulosilyticum ruminicola_UPI0006D00C5B
Roseburia sp900552665_A0A1Q6T5G5
Acetatifactor sp900066365_A0A7G9FW08
Roseburia inulinivorans_A0A0M6WYS8
Firmicutes bacterium CAG:552_39_19_A0A1Q6RRZ6
Roseburia inulinivorans_A0A412B2E8
Roseburia intestinalis_A0A173S481
Butyrivibrio hungatei_B_A0A1G5FKH7
Acetatifactor sp900066365_A0A417QBK0
Acetatifactor sp900066365_A0A416NN41
Roseburia sp900552665_A0A3R6H4Z2
TF01-11 sp003524945_A0A416HIZ1
Roseburia inulinivorans_A0A173WCI3
Roseburia inulinivorans_C0FPV1
Acetatifactor sp900066365_R6EJ07
Agathobacter ruminis_A0A2G3E666
MCWD5 sp003363435_A0A371AZN6
Butyrivibrio fibrisolvens_C_UPI00041E1D04
Roseburia hominis_A0A395V4R6
Butyrivibrio sp000621605_UPI00047EC982
CAG-45 sp900066395_A0A374B174
CBir1
Vallitalea_A okinawensis_UPI000CFC8B08
CAG-95 sp900066375_A0A417U3G5
Roseburia sp001940165_A0A1Q9J2N6
Roseburia sp003470905_A0A413HJC5
CAG-95 sp00436115_R7A061
Lachnobacterium sp900113385_A0A1I2NSN5
CAG-95 sp900066375_A0A1C5RMA0
Lachnospiraceae bacterium NK4A179
Roseburia intestinalis_A0A413SGX0
Roseburia intestinalis_A0A415TS70
Roseburia hominis_FlaB
CAG-45 sp000434275_R7QUJ1
TF01-11 sp003529475_A0A3D2TUV9
Butyrivibrio sp000526935_UPI000465FF0F
Lachnobacterium bovis_A0A1H3M4E4
Butyrivibrio sp000420845_UPI0003B5D321
Butyrivibrio sp000420945_UPI00040F20D9
TF01-11 sp003149875_A0A316NRI6
TF01-11 sp001414325_A0A416DP13
Clostridiales bacterium isolate CIM_A0A316QVQ3
Anaerobium acetethylicum_UPI00086BF2D7
[Bacteroides_F pectinophilus]_R7B0F6
Agathobacter sp000434275_A0A1Q6U286
Butyrivibrio_A crossotus_R5LAZ5
Butyrivibrio_A crossotus_D4S2X1
Butyrivibrio sp900102515_UPI00041 7FFB8
Eubacterium_F sp000433735_R5VHP0
Butyrivibrio sp. AE2032_UPI0005571835
TF01-11 sp001414325_R5S833
XBB1006 sp900115795_A0A1I5WKJ9
Butyrivibrio sp900102515_A0A1G5BQP6
TF01-11 sp001916135_A0A417QCB3
Acetatifactor muris_A0A2K4ZAI4
Eubacterium_J plexicaudatum_N2B3U6
Kineothrix alysoides_A0A4R1QN83
Butyrivibrio proteoclasticus_A_UPI00047B41B6
Acetitomaculum ruminis_A0A1I0YA99
Butyrivibrio sp000420845_UPI0003B635D7
Lachnospira rogosae_A0A174MPU4
TF01-11 sp001916135_R7JVE3
-6 -5 -4 -3 -2 -1 0 1
log10 TLR5 activity
0 1 2 3 4
TLR5N14 binding (a.u.)
Family
Lachnospiraceae
Enterobacteriaceae
Cellulosilyticaceae
Vallitaleaceae_A

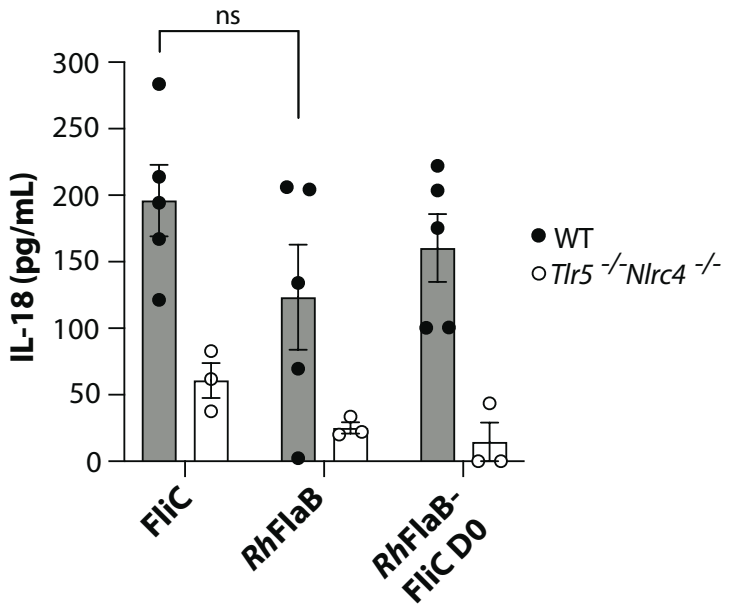
Supplementary Figure 2.3 - TLR5 recognition and activation by *Rh*FlaB-like flagellins.

(A) Activation of TLR5 by *Rh*FlaB-like flagellins. TLR5 activity measured as in Supp. Fig. 2.2A. Data are mean ± SEM from at least three independent experiments. Values normalized to *St*FliC. (B) Flagellin binding to $TLR5^{N14}$. AP-tagged flagellins were incubated with $TLR5^{N14}$ bait and AP activity was quantified. Data are mean ± SEM for n≥2 and normalized to *St*FliC. Bar colour indicates family-level taxonomy (GTDB). *St*FliC and *Rh*FlaB are shown in bold. See Fig. 2.4A.



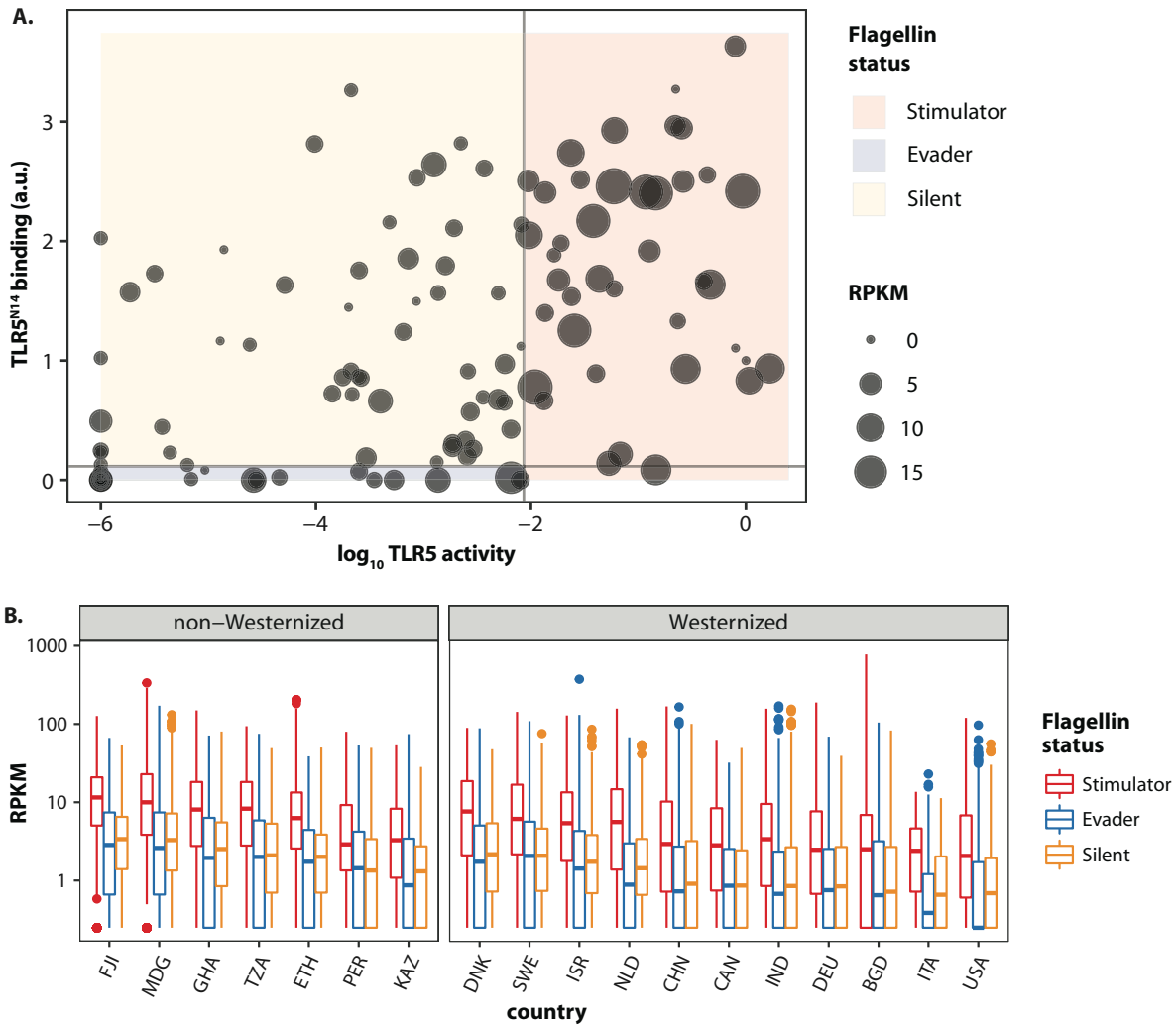
Supplementary Figure 2.4 - *Lachnospiraceae* are dominant producers of flagellin in the human gut.

(A) Relative abundance of flagellin gene transcripts from metatranscriptomes of 26 healthy controls from ref 18, as shown by family-level taxonomy (GTDB). (B) Relative abundance of *Lachnospiraceae* in human gut metagenomes from the curated Metagenome Data collection (n = 1783). Only samples labelled as "healthy" were included. The dataset comprised 21 studies and individuals from 18 countries. Metagenome profiling was conducted via Kraken2 and Bracken with a custom reference database generated from the Genome Taxonomy Database (GTDB) Release 95 [35]. (C) Median relative abundance of Lachnospiraceae in Westernized vs non-Westernized populations from (B). Significance between medians determined by Wilcoxon (****P<2.2e-16).



Supplementary Figure 2.5 - Silent flagellin *Rh*FlaB activates the NLRC4 inflammasome.

Wild-type and *Tlr5-/-Nlrc4-/-* mice were injected with indicated flagellins and IL-18 levels in serum were measured by ELISA. Significance between WT *St*FliC and *Rh*FlaB means was determined by unpaired, two-tailed *t* test (ns, not significant). FliC = *St*FliC.



Supplementary Figure 2.6 - Stimulator flagellins are more abundant than evader and silent flagellins among healthy human populations.

(A) The activity, binding, and median reads per kilobase per million reads (RPKM) for each flagellin. RPKM values were calculated by mapping metagenome Illumina reads from "healthy" individuals in the curatedMetagenomics dataset (n = 1783). (B). Median RPKM for each flagellin, grouped by country (ISO 3166-1 codes), Westernization status, and flagellin status.

2.8. Supplementary Tables

Taxonomy	**Flagellin**	**EC_{50} (nM)**	**$logEC_{50}$ ± SEM**
Salmonella typhimurium	FliC	0.0644	-1.19 ± 0.0867
	FliC-PIM (Q89A R90A Q97A)	9.44	0.975 ± 0.0936
Roseburia hominis	FlaB	138.8	2.14 ± 0.0564
	FlaB-FliC D0	2.53	0.403 ± 0.0632
	Fla1	3.37	0.528 ± 0.0926
	Fla2	>1000	3.25 ± 0.460
	FliC5	>1000	3.31 ± 0.199
Agathobacter rectalis	C4ZDK3	487.2	2.69 ± 0.0613
	C4ZDK3-FliC D0	1.01	0.004 ± 0.0915
Eubacterium_F sp003491505	R7CCG6	114.2	2.06 ± 0.0535
	R7CCG6-FliC D0	3.95	0.596 ± 0.257

Supplementary Table 2.1 - EC_{50} values for purified recombinant flagellins.

Taxonomy	Flagellin (UniProt ID)	% coverage	Maxquant score	TLR5 epitope
**Agathobacter rectalis*	WP_012744006.1 (C4ZDK3)	40.4	238	LQRMNELATQ
**Eubacterium_F sp003491505*	CDD75417.1 (R7CCG6)	29.2	131	LQRARELVVQ
**Roseburia intestinalis*	CDA56948.1 (R6AYN5)	25.2	68.0	LQRMNELATQ
**Roseburia inulinivorans*	WP_055168195.1 (A0A173SD30)	23.9	60.0	LQRMNELAVQ
**UBA9502 sp003480315*	SCG91213.1 (A0A1C5KT57)	8.4	12.4	LQRMNELAVK
**Lachnospira eligens_A*	RHL68652.1 (A0A174Z6V7)	15.1	9.66	LQRMNELATQ
**14-2 sp000403845*	EOT27140.1 (S0JEY7)	9.2	9.23	LQRMNELATQ
**Roseburia intestinalis*	RHA58738.1 (A0A3R6A0K5)	25.9	8.56	LQRMNELATQ
unclassified *Clostridium*	WP_117780849.1 (UPI000E553912)	8.5	6.99	LQRMGELATK
**TF01-11 sp001916135*	CDE69670.1 (R7K069)	8.5	6.70	LQRMNELAVQ
**Roseburia inulinivorans*	RHA89587.1 (A0A413TXC5)	25.9	5.91	LQRMNELATQ
Clostridium hydrogeniformans	WP_027633903.1 (UPI000485A193)	5.9	5.87	LQRMRELSVQ

Supplementary Table 2.2 - Endogenous flagellins identified directly from human stool by mass spectrometry using TLR5 as bait.

Asterisk denotes members of the *Lachnospiraceae* family. Conserved residues in the TLR5 epitope required for recognition are underlined.

Gene	Old gene name	Protein	TLR5 epitope
RHOM_RS00690	RHOM_00665	Fla1	LQRMNELATQ
RHOM_RS15365	RHOM_15820	FlaB	LQRMRELSVQ
RHOM_RS00845	RHOM_00820	Fla2	LDRMVELTTQ
RHOM_RS00990	RHOM_00975	FliC5	LQKIRELGVR

Supplementary Table 2.3 - Flagellins encoded by *R. hominis*.

Conserved residues in the TLR5 epitope required for recognition are underlined.

Supplementary Table 2.4 – Raw data used to generate figures.

The raw data for figures in the main text and supplement are compiled in an excel sheet, and labelled accordingly. This datasheet can be accessed online at the following address: https://www.science.org/doi/suppl/10.1126/sciimmunol.abq7001/suppl_file/sciimmunol.abq7001_table_s4.zip

2.9. Supplementary Data Files

Supplementary data files are available online at the following address: https://www.science.org/doi/suppl/10.1126/sciimmunol.abq7001/suppl_file/sciimmunol.abq7001_data_files_s1_and_s2.zip

The descriptions for supplementary data files are provided below.

Supplementary Data 2.1. – Protein Groups.

All protein groups identified in the processed raw proteomics files (spectra). Each row contains the group of proteins that could be assigned to a set of identified peptides that were processed with a 1% false discovery rate.

Supplementary Data 2.2. – curMeta.

All relevant metadata associated with the curatedMetagenomics dataset samples used in this work. The country identifiers are ISO 3166-1. All data was obtained from the full curatedMetagenomics dataset.

Chapter 3: A kinetic and structure-based comparison of silent and stimulatory flagellin interactions with TLR5

This chapter is comprised of an advanced manuscript, awaiting submission to a target journal.

Authors are listed in the intended order for publication:
Michael E. W. Bell[1,5], Miriam Haag[1,5], Marieke Böcker[1], Kateryna Maksymenko[2], Sara J. Clasen[1,5], Iris Koch[3], Veronika Altmannova[4], Katharina Hipp[3], Marcus Hartmann[2], Andrei Lupas[2], John R. Wier[4], Felipe Merino[2], Ruth. E. Ley[1,5].

[1] Department of Microbiome Science, Max Planck Institute for Biology, Tübingen 72076, Germany.
[2] Department of Protein Evolution, Max Planck Institute for Biology, Tübingen 72076, Germany.
[3] Electron Microscopy Facility, Max Planck Institute for Biology, Tübingen 72076, Germany.
[4] Friedrich Miescher Laboratory of the Max Planck Society, Max-Planck-Ring 9, Tübingen 72076, Germany.
[5] Cluster of Excellence EXC 2124 Controlling Microbes to Fight Infections, University of Tübingen, Tübingen, Germany.

Manuscript contributions: This project was conceptualised by myself, SJC, and REL. Plasmids were produced by myself and SJC. Recombinant proteins were produced by myself, MH, and MB. SPR methodology was designed by myself, SJC, and RL. SPR analyses and optimisation were performed by myself, MH, and MB, with support from KM. SPR results were analysed by myself, MH, MB, with support from KM. I designed the methodology for flagella shearing and purification. I isolated and purified the flagella with advice from VA, and JRW. I prepared samples for negative staining and Cryo-EM analyses. TEM and Cryo-EM imaging was performed by myself, IK, and KH. I resolved the density map and refined the structure with generous help from FM. I analysed the structure, with advice from FM, JRW, and MH. I wrote the complete original text for this manuscript, which was revised by REL and MH. I generated all figures other than the SPR curves in the supplement, which were made by MH.

Abstract

Toll-like receptor 5 (TLR5) detects and responds to bacterial flagellin. Some flagellins, termed "silent," bind TLR5 but elicit weak responses. This binding outcome involves interactions at both D1 and D0 interfaces, but the D1 domain binding mechanism remains unclear. We studied interactions between truncated D1 domain binding TLR5 ($TLR5^{N14}$), and the stimulatory *S. typhimurium FliC* (*St*FliC) and silent *R. hominis* FlaB (*Rh*FlaB) flagellins. Surface Plasmon Resonance revealed *St*FliC associated substantially slower, but formed longer-lasting TLR5 complexes than *Rh*FlaB. Different functions were also identified for residues in the well-established TLR5 epitope between these proteins. Cryo-EM structures identified hydrophobic pockets and a negatively charged surface unique to *Rh*FlaB, which explained the observed rapid association and shorter-lived complexes. These findings suggest silent flagellins may evade TLR5 signalling through rapid dissociation, impeding D0 binding and subsequent effector recruitment. This novel kinetic and structural insight into TLR5-flagellin interactions has potential applications in the design of interventions targeting TLR5.

3.1. Introduction

The human body is colonised by a diverse range of microbes, which are present at a 1:1 ratio with human cells [1]. The vast majority of these bacteria are located in the large intestine, and constitute the gut microbiome [2,3]. Precisely how these impact our health and well-being is of great interest, and research continues to uncover how the microbiome influences numerous aspects of human development, health, and homeostasis [4]. As a first line of defence, most complex organisms have an extensive and generalised innate immune system designed to protect us against disease-causing microbes. In humans, much of this system is focused around the human gut [5,6]. However, while the immune system must prevent pathogen invasion, it should also allow for the persistence of the numerous beneficial bacteria present in the microbiome.

The human Toll-Like Receptor (TLR) family is a group of 12 Type-1 transmembrane protein Pattern Recognition Receptors, and are key in the innate immune system. TLRs target highly conserved Microbe Associated Molecular Patterns (MAMPs) in bacteria, viruses, and other pathogens. [5–7]. Toll-Like receptor 5 (TLR5) specifically targets bacterial flagellin, the

monomeric subunit of the bacterial flagella, the organelle utilised for bacterial motility and invasion. Flagellins generally form an L-shaped structure, composed of two highly conserved domains (D0 and D1), which form the core of the flagella, and hypervariable domains, which form the outer surface. The flagellin MAMPs targeted by TLR5 are located in the D0 and D1 domains, and are integral for polymerisation when forming the flagellum (Fig. 3.1A) [8,9]. It is thought that MAMP binding facilitates the formation of 2:2 TLR5-flagellin heterodimer complexes. These recruit the signalling mediator MyD88, which forms the Myddosome complex, resulting in NF-κB pathway activation [7,10–12]. This process is considered to have a substantial temporal aspect, however this has yet to be explored [13].

Existing studies on TLR5-flagellin interactions, which focus only on common pathogens such as *Salmonella enterica* serovar *typhimurium* (*St*FliC), have shown that these MAMPs form hydrogen bonds in complex, and stabilise the TLR5-flagellin heterodimer. The most well-characterised MAMP comprises the residues *St*FliC: L88-Q97, which we refer to as the TLR5 epitope [9,14,15]. This epitope is conserved throughout most flagellated human commensal bacteria including *Lachnospiraceae*, which comprise the majority of bacteria in the gut microbiome, and contains the 'PIM' residues (*St*FliC: Q89, R90, Q97), key in TLR5-flagellin binding [14–16] (Fig. 3.1A). The lack of a TLR5-epitope is a notable feature of evasive pathogens such as *Helicobacter pylori*, which avoid TLR5 recognition through drastic modifications to their flagellin sequence and flagellum conformation [17,18].

We recently described a third class of ‘silent’ interactions between TLR5-epitope containing commensal flagellins and TLR5, characterised by strong TLR5 binding but a weak immune response. We identified the notable silent flagellin *Roseburia hominis* FlaB (*Rh*FlaB), from a common commensal, which bound TLR5 twice as strongly at the D1 domain but elicited a 4-fold weaker immune response compared to *St*FliC [14]. By comparing *St*FliC and *Rh*FlaB, we found that allosteric binding at the D0 domain was required to bind the TLR5 dimer and trigger an immune response [14,19]. We also observed that D1 binding was necessary, and identified a role for the cD1 domain in modulating the immune response. However, the mechanism behind this remains unknown.

In this study, we aim to identify this mechanism by exploring the temporal nature of interactions between TLR5^{N14}, a truncated human TLR5 construct which only binds the flagellin D1 domain, and *St*FliC or *Rh*FlaB (Fig. 3.1B). This expands on our previous comparative binding analyses by showing how the binding kinetics of this complex interaction can facilitate D0 domain binding and subsequent immune response propagation [14]. We further solved the structure of *Rh*FlaB, and identified features of the D1 domain which differed to an existing *St*FliC structure [20]. With this data, we postulate a kinetics-based mechanism by which structural features in silent flagellin D1 domains impair D0 domain binding, complex stability, and ultimately the TLR5 immune response, resolving a clearer picture of D1 and D0 binding cooperate to facilitate agonistic and silent TLR5 responses.

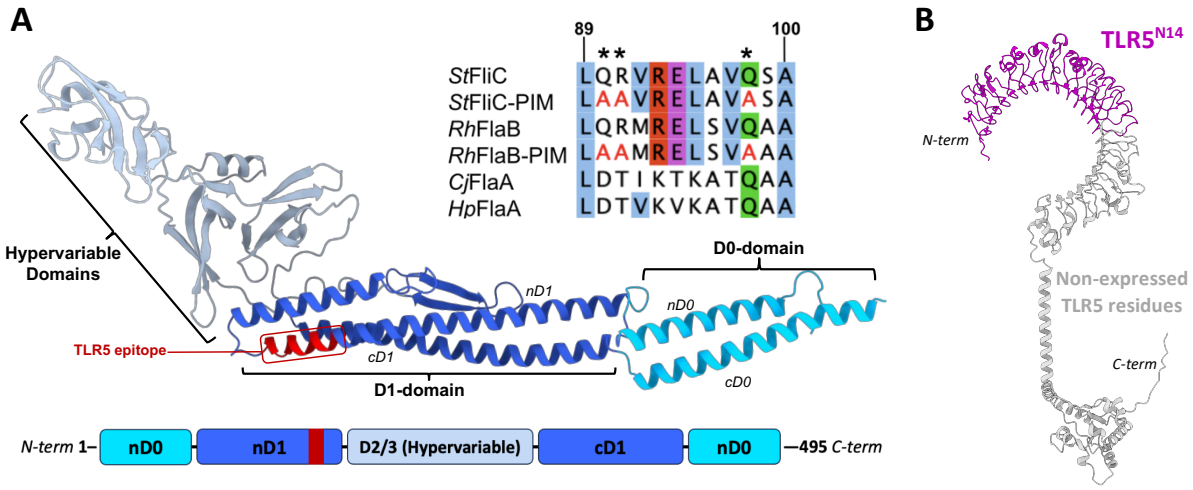


Figure 3.1 – Structural overview of flagellin and TLR5^{N14}.
(A) A *St*FliC structure showing the 'standard' flagellin domain layout (PDB ID #6jy0) and a multiple sequence alignment of the TLR5-epitope in relevant flagellins. PIM residues are indicated with '*'. Each domain is coloured separately, and the TLR5-epitope is highlighted in red. Colours correspond to the domain map. (B) The predicted structure of human TLR5, with the residues expressed in the TLR5^{N14} construct coloured in purple. Residues not expressed are shown in grey. Both structures were generated with Alphafold2 [22].

3.2. Results

3.2.1. *R. hominis* FlaB displays stronger binding to TLR5^{N14} compared to *S. typhimurium* FliC

We produced and purified a range of recombinant flagellins from *E. coli* in order to study their D1 domain interactions with insect-cell purified TLR^{N14} using Surface Plasmon Resonance (SPR) (Fig. 3.1B; Fig. 3.2A; Supp. Table 3.1). Wild-type *St*FliC, and *Rh*FlaB were the primary focus for analysis, as representative agonistic and silent flagellins (Fig. 3.1A). ΔD0 mutants (*St*FliC-ΔD0 and *Rh*FlaB-ΔD0) were studied to confirm that the flagellin D0 domain did not substantially interact with TLR5^{N14}, and PIM mutants (*St*FliC-PIM and *Rh*FlaB-PIM) allowed us to separate interactions between the TLR5 epitope and other D1 domain binding. A non-binding flagellin identified in our previous paper (Eubacterium sp. FlgL) was used as a negative control [14].

The binding strength (K_D) of the Wild-Type flagellins (WTs) showed *Rh*FlaB binds more strongly than *St*FliC (Fig. 3.2B; Table 3.1; Supp. Fig. 3.1A). While significant differences were not present in the pooled results, individual runs were significant (Supp. Table 3.2). The K_D of both flagellins PIM mutants indicated substantially weaker binding, with a more than 10-fold increase in K_D compared to the WTs (Fig. 3.2B). All flagellin mutants lacking the D0 domain (ΔD0) showed marginal but significantly lower K_D values compared to non-ΔD0 equivalents (Table 3.1; Supp. Fig. 3.1A).

Flagellin construct	K_D [nM]	k_a [1/Ms]	k_d [1/Ms]
S. typhimurium FliC	18.13 ±5.308	12072 ±2797	0.227 ±0.117
S. typhimurium FliC-ΔD0	8.772 ±1.012	15573 ±357.3	0.137 ±0.0169
S. typhimurium FliC-PIM	585.1 ±26.08	18830 ±1981	11 ±0.683
S. typhimurium FliC-PIM-ΔD0	274.3 ±13.76	42807 ±1612	11.73 ±0.274
R. hominis FlaB	12.65 ±2.56	45157 ±7459	0.555 ±0.027
R. hominis FlaB-ΔD0	7.081 ±0.087	75480 ±2424	0.534 ±0.011
R. hominis FlaB-PIM	77.99 ±21.15	9091 ±1524	0.691 ±0.129
Eubacterium sp. FlgL	87.87 ±17.34	9028 ±2803	0.755 ±0.054

Table 3.1 – Binding and Kinetics results of flagellin interactions with $TLR5^{N14}$ from Surface Plasmon Resonance analyses.

The final bindings and kinetics values from each SPR curve were pooled for analysis. Each measurement had at least 3 technical replicates. The value displayed for each binding constant is the mean ± standard deviation.

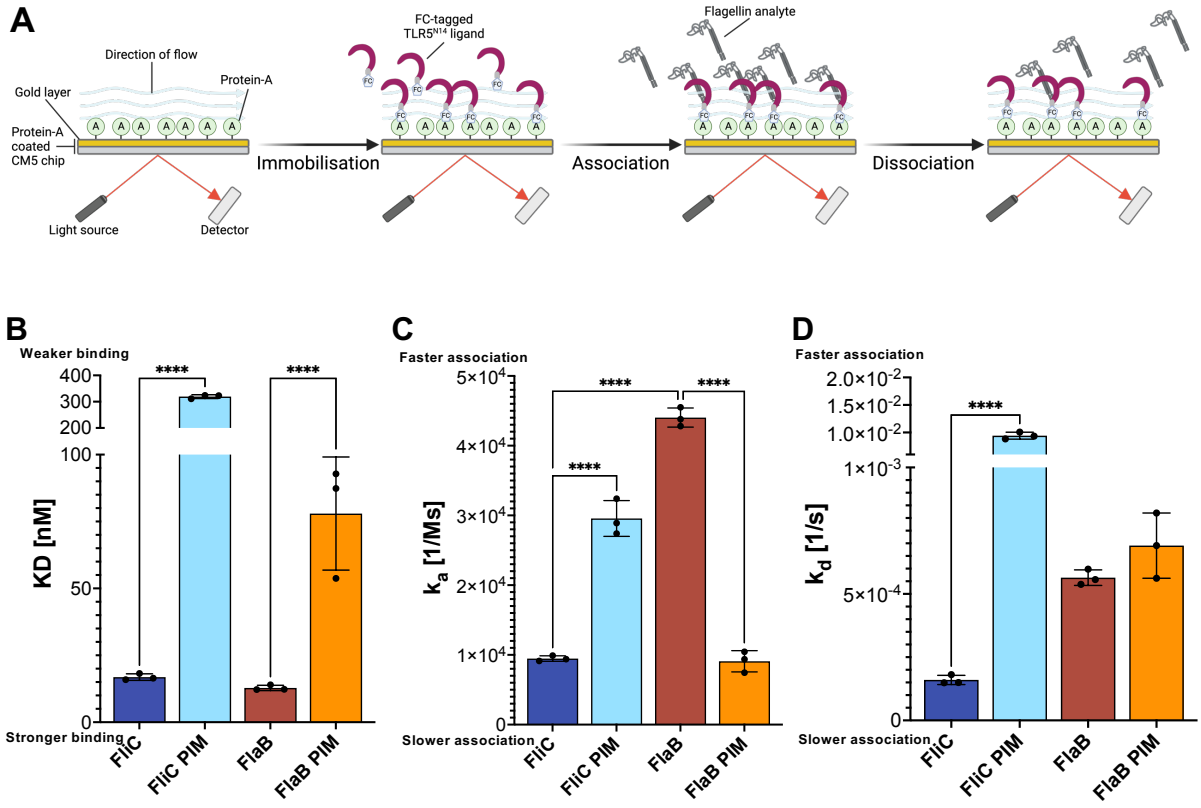


Figure 3.2 - Differing interaction kinetics are observed in *St*FliC and *Rh*FlaB binding to TLR5.

(A) Graphic overview of SPR analysis. FC-tagged TLR5^{N14} was immobilised on a protein-A coated CM5 chip. The flagellin analyte was run over the immobilised TLR5, with association (k_a) and dissociation (k_d) rates calculated. The binding strength (K_D) was calculated from the kinetics constants (k_a and k_d) with the Biacore X100 evaluation software (Cytiva). (B-D) Representative triplicate K_D (B) and kinetics (C-D) results from a single SPR run of *St*FliC and *Rh*FlaB against their respective PIM mutants. The complete dataset with all runs and ΔD0 control mutants are available (Table 3.1; Supp. Fig. 3.1), including individual curves (Supp. Fig. 3.1D-J). Error bars represent SD, significance was calculated by one-way ANOVA (****$P < 0.0001$).

3.2.2. *St*FliC and *Rh*FlaB have distinct temporal profiles when binding TLR5

By determining the association (k_a) and dissociation rates (k_d), we observed stark differences in the temporal nature of interactions between these two representative flagellins. *St*FliC both associated and dissociated substantially slower than *Rh*FlaB, though the difference was not significant for the latter (Fig. 3.2C,D; Table 3.1). This relationship was causal to the relative parity in their overall K_D, which is based on the kinetics constants. *St*FliC-ΔD0 had no significant difference in binding to *St*FliC, while *Rh*FlaB-ΔD0 showed a small but significant increase in association (Supp. Fig. 3.1B,C).

The stark differences in K_D observed between the WTs and PIM mutants were found to stem from different aspects of the binding interaction. *Rh*FlaB-PIM associated significantly slower than *Rh*FlaB, but *St*FliC-PIM unexpectedly associated with $TLR5^{N14}$ at an increased rate compared to *St*FliC (Table 3.1; Fig. 3.2C). Instead, a significant increase in the dissociation rate was observed in *St*FliC-PIM, indicating that the PIM residues have an anchoring role in this complex (Table 3.1; Fig. 3.2D). *Rh*FlaB-PIM showed no significant change in dissociation rate compared to the WT (Table 3.1; Fig. 3.2D). These characteristics were also apparent in the SPR binding curves for these constructs (Supp. Fig. 3.1D).

Eubacterium sp. FlgL (UniProt ID #R5DW89), an evasive flagellin identified in our previous study, was also analysed [14]. Interestingly, it acted almost identically to *Rh*FlaB-PIM with regards to all kinetics measurements (Supp. Fig. 3.1A-C), suggesting poor association at the primary interface, despite the presence of the TLR5 epitope.

Together, these results show that *St*FliC is bound to TLR5 for a longer duration than *Rh*FlaB. Differing roles for PIM residues in the TLR5 epitope are clear. *Rh*FlaB was almost completely reliant on these residues for association, while *St*FliC appears reliant on other D1 domain residues for association, with the PIM residues instead conferring stability.

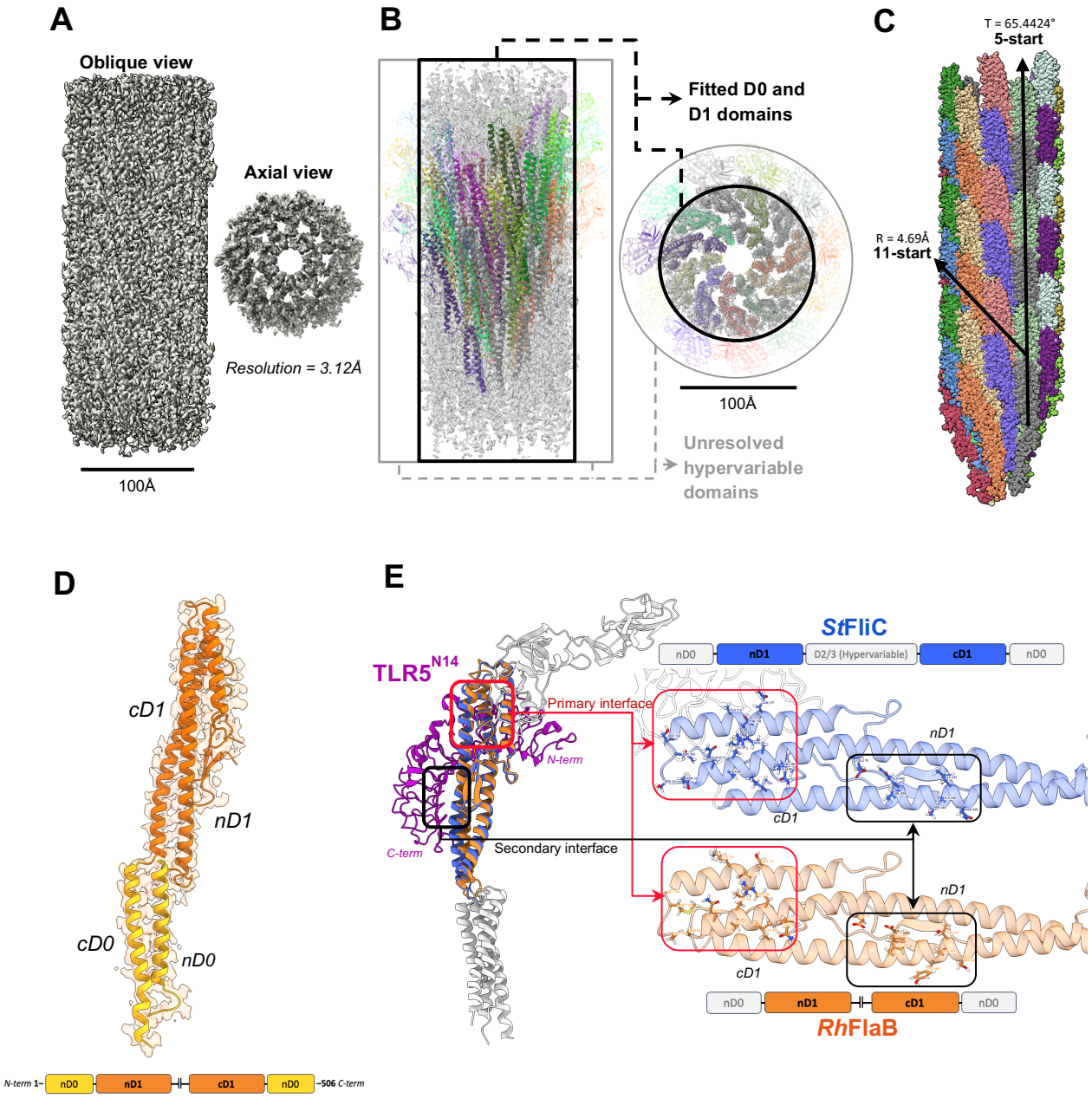

A
Oblique view
Axial view
Resolution = 3.12Å
100Å
B
Fitted D0 and
D1 domains
100Å
Unresolved
hypervariable
domains
C
T = 65.4424°
5-start
R = 4.69Å
11-start
D
cD1
nD1
cD0
nD0
N-term 1
nD0
nD1
cD1
nD0
506 C-term
E
TLR5N14
Primary interface
N-term
C-term
Secondary interface
StFliC
nD0
nD1
D2/3 (Hypervariable)
cD1
nD0
nD1
cD1
nD1
cD1
nD0
nD1
cD1
nD0
RhFlaB

Figure 3.3. Cryo-EM structure of the *R. hominis* flagella and flagellin monomer. (A) Oblique and axial view of the final flagella density map. The filament segment has a diameter and length of ca. 123 Å and 270 Å respectively. The reported resolution of the structure is 3.12Å (Supp. Fig. 3.4). (B) Fitted 22-flagellin *Rh*FlaB complex in the post-processed density map. D0 and D1 domains fit in the density map, but the hypervariable domains (transparent) were not reflected in the final structure. These are in black and grey boxes respectively. (C) 3D atomic reconstruction of the *Rh*FlaB flagella with fitted D0 and D1 domains. Repeated flagellins are shown in the same colour. The 5-/ 11-start rotational symmetry, Twist (65.4424º) and Rise (4.69Å) are indicated. (D) The D0 and D1 domains of a *Rh*FlaB monomer fitted into the electron density map. The D0 and D1 domains are coloured, and correspond to the domain map. (E) Superposition complex of *Rh*FlaB and *St*FliC (PDB ID #6jy0) bound to $TLR5^{N14}$. The $TLR5^{N14}$ structure was generated with AlphaFold2 [22]. The Primary and Secondary interfaces are highlighted. Residues implicated in *St*FliC binding which are absent in *Rh*FlaB are displayed (blue). The corresponding residues shown in *Rh*FlaB (orange) were identified by sequence and structure alignment (Supp. Fig. 3.7).

3.2.3. Isolation and purification of *Roseburia hominis* flagella

A serum containing a *Rh*FlaB-targeting antibody was produced, and checked for cross-reactivity against recombinant flagellins produced based on other flagellin identified in the *R. hominis* genome (Supp. Table 3.1). Minimal cross-reactivity was observed (Supp. Fig. 3.2). Flagella samples were mechanically sheared from *R. hominis* culture, and analysed by SDS-PAGE and western blot. Both analyses showed bands at ~60kDa and ~40 kDa in fractions 5-8, which differed from the ~54 kDa size predicted by the sequence (Supp. Fig. 3.3) [21]. As these corresponded to each-other, it confirmed the flagellum in these fractions were composed of *Rh*FlaB. The reason for the apparent difference in size was clarified later.

3.2.4. Solving the structure of the *Rh*FlaB flagellum complex and flagellin monomer

Negative staining was used to confirm sample purity and distribution (Supp. Fig. 3.4A). 2,480 TEM images of *R. hominis* flagella were taken, then manually screened (Supp. Fig. 3.4B). Auto-picking extracted 576,585 particles, which was sorted by 2D class averaging to a final dataset of 443,605 flagella particles (Supp. Fig. 3.4C). Twist (65.4424°) and Rise (4.69Å) parameters were found, facilitating 3D refinement. A final 3.12Å electron density map of the supercoiled flagella was resolved, with low variation in resolution throughout the structure (Fig. 3.3A; Supp. Fig. 3.5).

*Rh*FlaB monomer structures generated with Alphafold2 were fitted into the density map, building a 22-flagellin complex [22]. The fitting showed that the density covered the D0 and D1 domains, but lacked clear densities corresponding to the hypervariable domains (Fig. 3.3B). This explained the size discrepancy in protein bands observed during flagella shearing (Supp. Fig. 3.3), with the hypervariable domains (MW = ~24kDa) apparently degrading during expression or purification. These domains were removed from the model before refinement. Structure refinement on individual monomers produced the final map, and the complex resolving as an 11-start right-handed helix, common for similar MAMP-containing flagellins including *St*FliC (Fig. 3.3C,D) [23–25]. The final structure showed a good fit to the density map, no severe clashes, and no residues outside of energetically allowed regions (Supp. Fig. 3.6).

Superposition TLR5^{N14}:*St*FliC and TLR5^{N14}:*Rh*FlaB complexes were generated for analysis, based on a zebrafish TLR5b (zTLR5b):*St*FliC crystal structure (PDB ID #3v47 from [15]) (Fig. 3.3D). As no high-resolution human TLR5 structure exists, a TLR5^{N14} model was generated using AlphaFold2. The solved *Rh*FlaB monomer, and a comparable existing *St*FliC structure (PDB ID #6jy0 [20]) were fitted accordingly.

3.2.5. Residues involved in *St*FliC - TLR5 interactions cluster in two distinct groups in the D1 domain

Previous experimental studies were screened to identify *St*FliC D1 domain residues involved in TLR5 binding. Mapping these onto the *St*FliC monomer revealed that they clustered into two groups, hereon referred to as the Primary and Secondary interface (Fig. 3.3E). The Primary interface contains the TLR5 epitope, and is primarily located in α-nD1. The Secondary interface is centrally located in the D1 domain, primarily in α-cD1. Both interfaces have been noted in previous studies, but were named differently (Interface B and A respectively in Yoon et al and Song et al [9,15]). Due to the principal role of the TLR5 epitope in binding, we preferred to consider it here as the Primary interface. The corresponding *Rh*FlaB residues were identified by sequence and structural alignments (Fig. 3.3E; Supp. Fig. 3.7). The superposition complexes showed the Primary and Secondary interfaces are located in regions expected to interact with $TLR5^{N14}$, based on the buried solvent-accessible surfaces in the superposition complexes.

3.2.6. *Rh*FlaB and *St*FliC show distinct differences in surface hydrophobicity at both D1 binding interfaces

The Primary interface surface profile has a single hydrophobic pocket on *St*FliC, while two larger pockets are present on *Rh*FlaB (Fig. 3.4A-B). The shared hydrophobic pocket is formed by a lysine and isoleucine residue (*St*FliC: L94, I111; *Rh*FlaB: L93, I111). Based on the superposition complexes, these interact with TLR5: I257. The second hydrophobic pocket in *Rh*FlaB is primarily formed by I86 and L118, which appear to bind TLR5: F282. The two hydrophobic pockets on *Rh*FlaB bracket R89 and Q88, and neighbour Q96 (*St*FliC: R90, Q89, Q97 respectively). These are the three residues mutated in the PIM constructs, and are known to form hydrogen bonds with TLR5 (Fig. 3.4C) [9,15].

*St*FliC has two hydrophobic pockets at the Secondary interface, the largest being located between the nD1 and cD1 α-helices. A second hydrophobic pocket is located on the converse cD1 surface, and contains an isoleucine (I447) implicated in TLR5 binding (Fig. 3.4A) [9,16]. While corresponding hydrophobic pockets are present on *Rh*FlaB, they contain fewer residues and lack an isoleucine corresponding to *St*FliC: I447 (*Rh*FlaB: T448) (Fig. 3.4B). The

corresponding binding interface on $TLR5^{N14}$ is located between LRR-1 and LRR-6 (Fig. 3.4D). It contains numerous aromatic amino acids including Phenylalanine, which are predictive of protein-protein binding sites (Fig. 3.4D) [26,27].

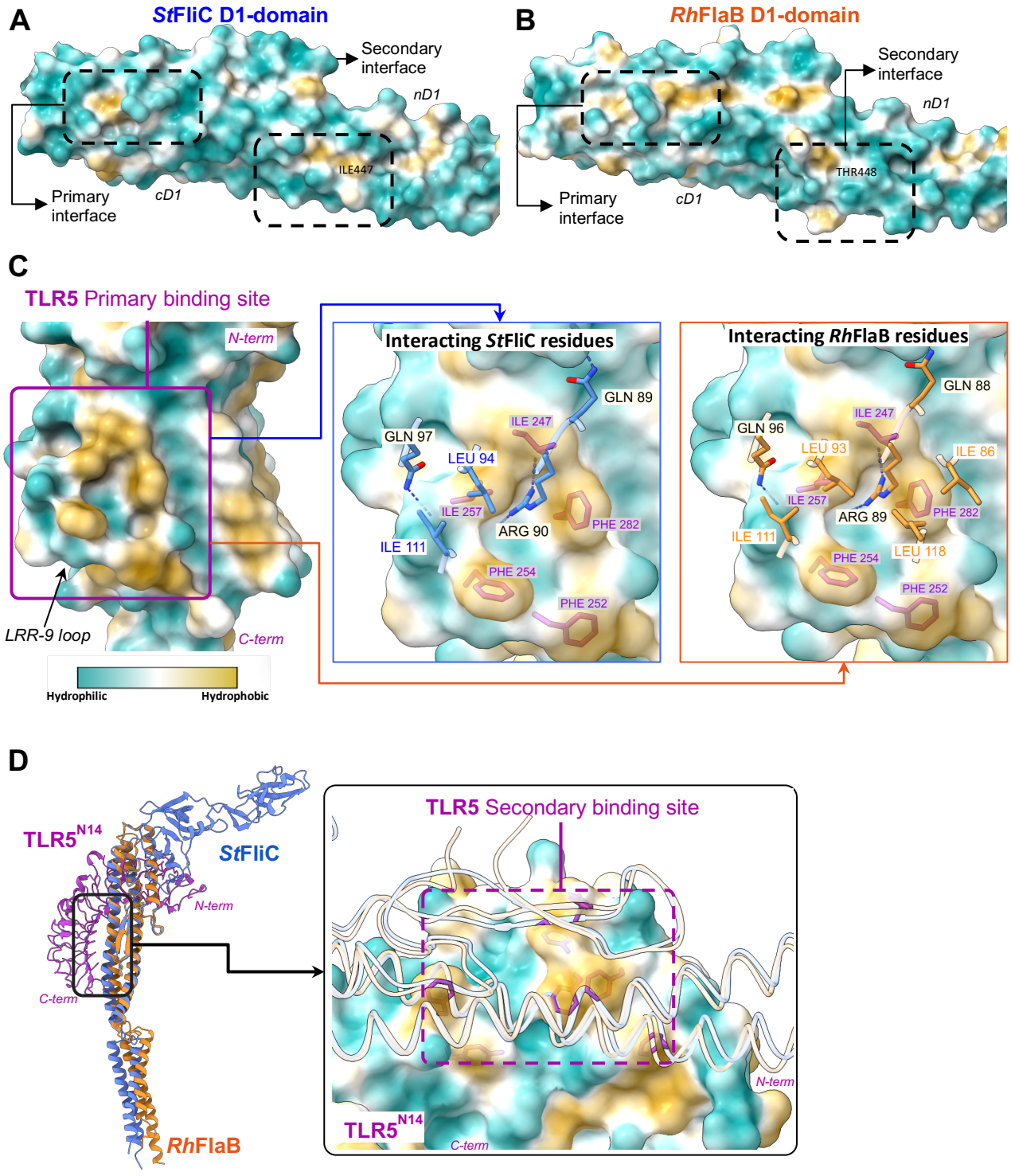

A
StFliC D1-domain
Secondary interface
nD1
ILE447
Primary interface
cD1
B
RhFlaB D1-domain
Secondary interface
nD1
THR448
Primary interface
cD1
C
TLR5 Primary binding site
N-term
LRR-9 loop
C-term
Hydrophilic
Hydrophobic
Interacting StFliC residues
GLN 89
ILE 247
GLN 97
LEU 94
ILE 257
ARG 90
PHE 282
ILE 111
PHE 254
PHE 252
Interacting RhFlaB residues
GLN 88
ILE 247
GLN 96
LEU 93
ILE 86
ILE 257
ARG 89
PHE 282
ILE 111
PHE 254
LEU 118
PHE 252
D
TLR5N14
StFliC
N-term
C-term
RhFlaB
TLR5 Secondary binding site
N-term
TLR5N14
C-term

Figure 3.4. - Comparison of Hydrophobic surface profiles at the Primary and Secondary binding interfaces.
(A-B) Hydrophobic surface profiles for *St*FliC and *Rh*FlaB D1-domain binding interfaces. The region comprising the Primary and Secondary binding interfaces are indicated. Hydrophobic regions are gold coloured. (C) Overview of the TLR5 Primary binding site (Purple), comprised of the LRR-9 loop and neighbouring residues. The *St*FliC (Blue) and *Rh*FlaB (Orange) residues expected to interact are displayed, with hydrophobic residues labelled in colour, and hydrogen-forming PIM residues labelled in black. Hydrophobic TLR5 residues expected to interface with flagellin are displayed (Purple). (D) Overview showing the TLR5 Secondary binding site expected to interact with flagellin from the superposition complex. Hydrophobic phenylalanine residues expected to interact with hydrophobic pockets in the flagellins are displayed. *St*FliC (Blue) and *Rh*FlaB (Orange) structures are overlaid for context.

3.2.7. *Rh*FlaB displays a highly negatively charged surface across D1 domain interfaces

Electrostatic interactions between individual residues were particularly difficult to analyse with superposition complexes, as direct charged interactions cannot be extrapolated. Consequently, a general view of the surface charge profiles across the binding interfaces was explored. *St*FliC displays a mix of negative, neutral, and positively charged residues at the Primary interface. The Secondary interface is primarily neutral, with few negatively charged residues (Fig. 3.5A). Interestingly, *Rh*FlaB showed a distinct strongly negatively charged surface across both TLR5 binding interfaces (Fig. 3.5B). This was most extreme at the Secondary interface, which displays a completely negatively charged surface. The corresponding TLR5 interfaces displayed primarily neutral surface charges, resulting from the aforementioned abundance of non-polar phenylalanine residues (Fig. 3.5C). Consequently, no pockets of charged residues were present on TLR5, which could be presumed to favour binding the negatively charged *Rh*FlaB.

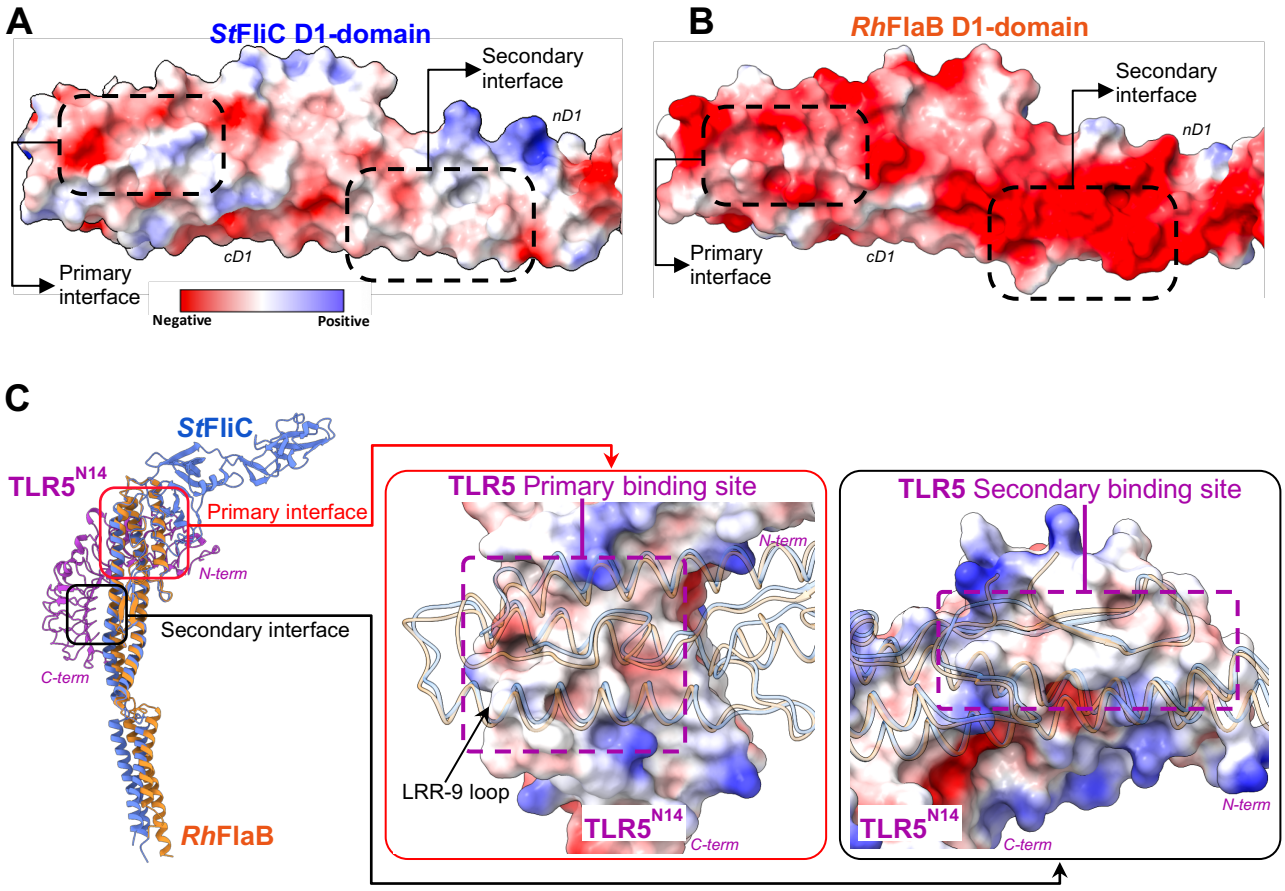


Figure 3.5 - Comparison of Electrostatic surface profiles at the Primary and Secondary binding interfaces.

(A-B) Electrostatic surface profiles for *St*FliC and *Rh*FlaB D1-domain binding interfaces. The region comprising the Primary and Secondary binding interfaces are indicated. Negative (Red), positive (Blue), and neutral (White) surface charges are represented. (C) Superposition showing the corresponding TLR5 Primary (Red) and Secondary (Black) binding interfaces expected to interface with flagellin. Interfacing regions are composed of numerous hydrophobic residues, which result in a primarily neutrally charged surface profile. *St*FliC (Blue) and *Rh*FlaB (Orange) structures are overlaid for context.

3.3. Discussion

In our previous work, we uncovered an allosteric binding site in the flagellin D0 domain which was key for TLR5 immune signalling [8,14]. However, the mechanism of action, including the role of the D1 domain in silent interactions, remained unsolved. We theorised that temporal factors, which are expected to play a role in TLR5 immune signalling, could contribute to this mechanism [13]. By using SPR to investigate the kinetics of flagellin D1 domain binding with $TLR5^{N14}$, we found that *St*FliC formed longer lasting complexes with TLR5, despite a weaker apparent K_D. *Rh*FlaB was solely reliant on the TLR5 epitope for binding, but *St*FliC association was instead facilitated by interactions at a secondary D1 domain interface, and the epitope instead integral for complex stability. Surface features of a novel *Rh*FlaB structure, absent from *St*FliC, were identified, which likely drive differences in their kinetic binding profiles.

Taken together, these findings shed light on how binding at the D1 domain can facilitate silent, stimulatory, and evasive responses when binding TLR5. These outcomes are primarily determined by binding kinetics, and not K_D. In light of our previous findings, our results show how the D1 and D0 domains may work in concert to modulate TLR5 activity, and build the foundation for further kinetic studies investigating flagellin D0 domain interactions. The identification of structural motifs driving the kinetic binding profiles can inform the study of further bacterial flagellins, and be applied in the design of ligands targeting TLR5.

3.3.1 Differences in the kinetic profiles of *Rh*FlaB and *St*FliC to TLR5 show a potential temporal mechanism of immune activity avoidance

Using Surface Plasmon Resonance (SPR), we investigated binding (K_D), association (k_a) and dissociation (k_d) rates for *St*FliC, *Rh*FlaB, and several informative mutants, against $TLR5^{N14}$. This truncated construct allowed for us to focus on the well-established D1-binding domains, without influence from potential D0 domain binding. ΔD0 flagellins were run to confirm this, and unexpectedly showed a small but significant decrease in K_D in almost all cases. However, the impact was marginal when compared to PIM mutants. The kinetic profiles from corresponding D0 and ΔD0 flagellins were similar, which indicates that this is likely an artefact of reduced protein size or a reduction in non-specific interactions.

The K_D measurements corroborate previous work, with *Rh*FlaB binding more strongly than *St*FliC, and both being substantially impaired in the PIM mutants [14]. The slower dissociation observed in *St*FliC indicated that it bound $TLR5^{N14}$ for a substantially longer duration than *Rh*FlaB, despite the slow association rate. The substantial increase in K_D seen in the PIM mutants stemmed from different aspects of the binding interaction. *Rh*FlaB-PIM showed very poor association compared to *Rh*FlaB, which was expected based on the purported role of these residues in complex formation [9,16]. Interestingly, no significant increase in dissociation rate was observed. This could be an artefact of the low association driven by indirect interactions; however, it gives context to the instability of the WT complex. Surprisingly, *St*FliC-PIM associated with $TLR5^{N14}$ at a comparable rate to the WT, but a >30-fold increase in dissociation was observed.

These findings reveal novel roles for the PIM residues in the Primary interface. They serve as the primary residues in *Rh*FlaB facilitating $TLR5^{N14}$ binding, but in *St*FliC play an anchoring role in the complex, with association primarily achieved by other residues, likely in the Secondary interface. The findings also highlight the limits of using K_D alone to study binding, as these flagellins with similar K_D values have clearly distinct interactions with TLR5.

The apparent inability of Eubacterium sp. FlgL to associate despite containing the PIM residues is surprising. As its kinetic profile mirrored *Rh*FlaB-PIM, it must utilise an unknown mechanism to evade Primary interface binding. Further study should compare its surface profile to *Rh*FlaB, to identify differences at the Primary interface, and examine if they share surface characteristics at the Secondary interface.

From these results, we theorise a new mechanism of immune response avoidance which results from 'silent' interactions. Pathogens such as *S. typhimurium* use several mechanisms to trigger an immune response and breach the gut epithelium [28,29]. The wider range of TLR5-associating residues and longer binding duration fit this profile, increasing the chances of association and complex formation. *Rh*FlaB associates quickly but for a short duration, and avoids TLR5-binding outside of the highly conserved PIM residues. We envisage that this prevents TLR5 from completing all downstream processes, forestalling an immune response.

3.3.2. Differences in the hydrophobic surface profile of the *Rh*FlaB Primary interface improve the association efficiency compared to *St*FliC

The two hydrophobic pockets found on the *Rh*FlaB primary interface surface directly bracket the PIM residues. This would improve the likelihood of binding in an orientation that facilitates the insertion of the hydrogen-bond forming PIM residues into the known binding pocket, and helps explain the high association rate [9,15]. The presence of only a single hydrophobic site in *St*FliC explains the opposite, with binding in the correct alignment less likely, limiting the chance of stable complex formation.

3.3.3. Differences in Secondary binding site surface features provide insight into why *St*FliC, but not *Rh*FlaB, can bind at this interface

The Secondary interface appears important for *St*FliC association, but has a negligible role in *Rh*FlaB. The existing structural information on interactions at this interface interaction were likely impacted by artefacts from using a truncated *St*FliC [15], limiting our ability to explore residue-residue interactions. Consequently, we have derived results from general properties of the Secondary interface.

The *St*FliC Secondary interface contains two broad hydrophobic pockets, which align over hydrophobic phenylalanine residues on the corresponding TLR5 binding site. While these were also present in *Rh*FlaB, they were smaller in scope, and occluded by protruding hydrophobic residues which clashed with TLR5 in the superposition complexes. We expect that hydrophobic aromatic residues on the TLR5 surface could more readily interact with hydrophobic pockets in *St*FliC, evidenced by the presence of *St*FliC: I447 in one of these pockets. However, nothing here definitively explained the lack of binding at the *Rh*FlaB Secondary interface.

While both Flagellins share a mostly negatively-charged Primary interface, the highly negatively charged *Rh*FlaB Secondary interface is a dramatic contrast. In the flagellum, these negatively charged surfaces are oriented towards the hollow centre, and may play a role in protein trafficking [30]. However, there is no clear advantage when binding to TLR5, and the negative charges may instead negatively impact diffusion and binding potential [31–34].

3.4. Conclusion

In this study, we provide a potential mechanism for silent-flagellin interactions, based on a binding profile with rapid association and dissociation. While displaying a high apparent K_D, *Rh*FlaB remains bound for only a short period, which may limit the ability for TLR5 to initiate a downstream immune response. The ability to bind at both D1 domain interfaces appears key for promoting a stable complex, which was supported by the evader sharing a kinetic profile with *Rh*FlaB-PIM. From structural data, we infer that differences in hydrophobic profiles likely play a substantial role in defining the association rate in *St*FliC and *Rh*FlaB. Notable differences in the electrostatic surface profiles were observed, and are likely partly responsible for preventing *Rh*FlaB Secondary interface binding. This temporal mechanism of immune-activity avoidance would serve as a reasonable compromise for bacteria aiming to avoid an immune response, without requiring the dramatic flagella restructuring found in ε-proteobacteria. Thus, it has the potential to explain the persistence of numerous high-abundance flagellated bacteria in our human gut, many of which play important roles in the microbiome.

3.5. Methods

3.5.1. Surface Plasmon Resonance analyses

Experimental design

This study aimed to investigate the dynamics of the binding interaction between flagellins and TLR5. Surface Plasmon Resonance (SPR) was chosen as binding kinetics are not measured in other established binding assays. *St*FliC and *Rh*FlaB were chosen as a representative TLR5 agonist and silent flagellin respectively, based on previous work [14]. Primary interface mutants (PIM) and ΔD0 mutants of these flagellins were analysed to study the individual contribution of the TLR5 epitope and D0 domain, respectively, to the $TLR5^{N14}$ binding. PIM mutants were made with three substitutions in the TLR5 epitope (*St*FliC: Q89A, R90A, Q97A; *Rh*FlaB: Q88A, R89A, Q96A). These residues are directly involved in TLR5 D1 domain binding, and the substitutions cause drastically reduced binding, and stimulatory activity [16,18]. ΔD0 mutants are truncated to remove the D0 domain, which is not expected to interact with $TLR5^{N14}$.

Flagellin Purification

Codon-optimised plasmids containing the recombinant flagellins were produced in our previous study, and expressed following the same protocol (See Chapter 2.4.4) [14]. For purification, cell pellets were resuspended in 1:10 w/v of resuspension buffer (20 mM Tris-HCl, 300 mM NaCl, 5 mM Imidazole, 1% TritonX-100, pH 7.5, protease inhibitor (cOmplete, EDTA-free, Roche #5056489001)), then sonicated at least 5 times with a SONOPULS mini20 ultrasonic homogenizer (BANDELIN) (1 min, 5 s pulses, 40% amplitude). Lysed cells were cleared by centrifugation (30 min, 18,000 rcf, 4°C) and filtered. Columns containing nickel-nitrilotriacetic acid (Ni-NTA) agarose (QIAGEN #30250) were equilibrated with 2 column volumes (CV) of wash buffer (20 mM Tris-HCl, 300 mM NaCl, 20 mM Imidazole, pH 7.5). Cleared lysates were incubated with Ni-NTA agarose in the column for 1 hour at 4 °C, then flowed through. The resin was washed with 2 CV wash buffer, then eluted with 0.5 CV elution buffer (20 mM Tris-HCl, 300 mM NaCl, 300 mM Imidazole, pH 7.5).

The eluent was concentrated to 1 mL with Pierce 10K MWCO protein concentrators (Thermo Fisher Scientific #8852) and further purified by size exclusion chromatography (SEC) with an Äkta Pure protein purification system (Cytiva) equipped with a Superdex® 200 Increase 10/300 GL column (Cytiva) in SEC buffer (300 mM NaCl, 50 mM HEPES, 1 mM TCEP, 5% Glycerol, pH 7.5). The eluent was collected in a 96-well plate, and verified by SDS-PAGE. Protein samples were then pooled, and the concentration measured (DS-11+ Spectrophotometer, DeNovix). All samples were stored at -20°C.

Purification of TLR5^{N14}

The production of TLR5^{N14} was performed following the same protocol as our previous study (See Chapter 2.4.3) [14]. In brief, IgG-FC-tagged TLR5^{N14} was synthesised with an N-terminal GP67 secretion sequence [35], and cloned into a modified pLIB vector containing a C-terminal 6xHis affinity tag. The vector was cloned into DH10EMBaCY cells to generate a recombinant baculovirus genome (bacmid). SF9 insect cells were inoculated with the bacmid, and the virus expanded through two rounds of amplification to a 'V2' expression culture. Protein production was performed in Hi5 cells using a 1:100 (v/v) inoculation of V2 supernatant. After 72 hours, the supernatant was harvested by centrifugation (800 rcf, 30 minutes, 4°C) and filter purified (Steritop 0.22 µm 250 mL Threaded Bottle Top Filter; Millipore # SCGPS02RE).

The supernatant was affinity purified using an Äkta Start protein purification system (Cytiva) equipped with two HisTrap HP 5 mL protein purification columns (Cytiva #175248). The supernatant was run through the HisTrap columns, which were then washed with 12 CV of wash buffer. Samples were eluted with 8 CV elution buffer in 4 mL fractions, and expression checked by SDS-PAGE. A final SEC purification step was performed with the same methodology as previously described. Samples were pooled and stored at -20°C.

Surface Plasmon Resonance

Flagellin samples were dialyzed in 1L of PBS-T (pH 7.5) at 4°C twice for ≥1 hour, and a third time overnight, in Slide-A-Lyzer 10K MWCO dialysis cassettes (Thermo Fisher Scientific #66380). Following the third dialysis, 500 mL of dialysis buffer was collected, filtered, and degassed for use as the SPR running buffer. Each SPR running buffer was used only for the samples it was used to dialyse. The protein concentration was measured, the sample diluted to the highest concentration (25 ng/µL for *Rh*FlaB constructs, 50 ng/µL for *St*FliC constructs) in SPR running buffer, then a fivefold 1:2 dilution series was made for single-cycle analysis.

SPR was performed using a Biacore X100 (Cytiva) equipped with a Sensor Chip Protein A (Cytiva #29127558). Pure $TLR5^{N14}$ was defrosted on ice, diluted to 300 ng/µL in SPR running buffer, then captured for 800 seconds. Run conditions were optimised for *St*FliC (360 s contact, 240 s dissociation, 30 µL/min flow rate) and *Rh*FlaB (180 seconds contact, 240 seconds dissociation, 30 µL/min flow rate), with mutants run at the same conditions as their respective WT. Only two samples could be measured per run, so each run paired WT *St*FliC or *Rh*FlaB against each-other or a respective mutant. SPR running buffer measured as a blank reference under the same run conditions, and subtracted from the sample curve. Three technical replicates were performed per sample per run.

Curve fitting (Supp. Fig. 3.1C-F) and analysis were performed with the Biacore X100 Evaluation Software (Cytiva). The affinity constant (k_a), dissociation constant (k_d), and binding strength (K_D) were determined for each replicate. KD values shown were calculated based on the reaction rate constants k_a and k_d, due to the relevance of the kinetics constants in this study. All curves passed the software's quality-control metrics. Closeness of the fitted curve was determined by the Chi-square and Rmax values.

Data analysis

Statistical analysis was performed with Prism 10 (GraphPad). Kinetics data are presented as arithmetic means ± standard deviations (SD). Statistical significance was determined with a one-way analysis of variance (ANOVA) with Tukey's multiple comparisons test. *P*-values < 0.05 were considered statistically significant.

3.5.2. Obtaining a density map for *R. hominis* flagella with Cryo-EM

R. hominis cell culturing

R. hominis cell culture was performed under anaerobic conditions. *R. hominis* (DSM #16839) pellets stored in liquid nitrogen were defrosted on ice, then transferred to 10mL BHIS (DSM #215c) at room temperature. The sample was grown for 24 hours at 37°C, and split to inoculate four Balch tubes containing 10mL YCFA medium (modified) (DSM #1611) at 1:1000 (v/v) concentration. The samples were cultured at 37°C for 16 hours.

R. hominis flagella shearing and purification

Shearing was performed in aerobic conditions. One tube (10 mL) of culture was transferred to a 50mL tube, passaged 15 times through a 18G syringe (Agani #123-3035309), then transferred to a 15mL tube. This was repeated for the remaining 3 cultures. Cells were cleared through two rounds of table-top centrifugation (2800 rcf, 30 min, 4°C), with the supernatant transferred to a clean tube between runs. The supernatant was transferred to 4.7 mL OptiSeal Polypropylene Tubes (Beckman Coulter #361621). The remaining material was pelleted by ultracentrifugation using a TLA-110 rotor and Optima MAX-XP Ultracentrifuge (Beckman Coulter) (100k rcf, 1 hr, 4°C). The supernatant was discarded, and the pellet resuspended in 200 µL PBS.

Gradient centrifugation was performed in 5 mL Open-Top Thinwall Polypropylene Tubes (Beckman Coulter # 326819). A continuous 20%-60% sucrose gradient in PBS was made using a Gradient Master 108 (Biocomp), and cooled to 4°C. The resuspended pellet was carefully added atop the gradient, then ultracentrifuged using a SW55-Ti rotor and Optima L-100 XP Ultracentrifuge (32400 rpm (Approximately 127000 rcf), 3 hr, 4°C). 500 µL fractions were manually aliquoted by pipetting, then checked by SDS-PAGE.

An antibody against *Rh*FlaB (Ab-*Rh*FlaB) was generated, and checked for cross-reactivity. Western blotting was used to confirm that fractions containing protein bands seen by SDS-PAGE corresponded to *Rh*FlaB. Fractions with the highest protein concentration were pooled and transferred to 4.7 mL OptiSeal Polypropylene Tubes, and ultracentrifuged (TLA110 rotor, 150k rcf, 1 hr, 4°C). The sample pellet was washed in PBS, pelleted again, then resuspended a final time in 100 µL PBS. The concentration of each sample was determined with the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific # 23225).

Negative staining

Flagella samples were imaged by Negative Stain for quality control prior to Cryo-EM. Samples were diluted to a concentration of 50 ng/µL with PBS. 4 µL were added to a carbon-coated 300 mesh copper grid, and incubated for 2 minutes. The liquid was removed with filter paper, and the grid washed thrice with water. The grid was washed twice with 1% uranyl acetate (UA), then stained for 5 minutes with 15 µL UA. The UA was carefully removed with filter paper, then the grid left to dry for at least 5 minutes. Sample grids were imaged by TEM using a Tecnai Spirit (FEI). Representative images showing flagellin concentration and distribution were taken for each sample.

Cryo-EM imaging and data processing

Purified flagella samples were directly frozen on R1.2/1.3 400 mesh copper grids (Quantifoil) using an EM GP automatic plunge freezer (Leica). Grids were imaged with a Talos Arctica Cryo-TEM (Thermo Fischer Scientific). Sample squares were mapped automatically with the manufacturer's software, and holes then manually selected for imaging. After automated image capture, poor-quality images manually removed from the dataset. MotionCorr2 and CTF-correction were performed with TranSPHIRE [36], then automated filament picking was performed with Cryolo [37]. Extracted particles imported to Relion [38]. 2D classification was used to filter out incorrect particles. An initial model was generated based on a hollow cylinder, and used for 3D refinement with helical reconstruction [39]. Exact twist and Rise parameters were determined with local searches of symmetry. Iterative refinements were performed using 3D refinement maps as the initial model. Post processing, CTF refinement and Bayesian polishing were performed to resolve the final density map.

Structure fitting & refinement

ChimeraX and ISOLDE were used for structure building and refinement [23,40]. An existing 22-flagella complex structure (PDB ID #6jy0 from [24]) was fitted into the density map. An initial *Rh*FlaB structure was generated with AlphaFold2 based on the type strain (DSM16839) [22,41,42]. *Rh*FlaB monomers were fitted based on the fitted 6jy0 complex, producing a 22-flagellin *Rh*FlaB complex. A central flagellin with the most inter-protein interactions was refined, then used to rebuild the complex. This process was repeated until no further clashes appeared when rebuilt. The structure was then refined as a monomer to resolve remaining outliers.

3.5.3. Comparison and analysis of protein structures

Identification of TLR5-binding regions of StFliC and RhFlaB

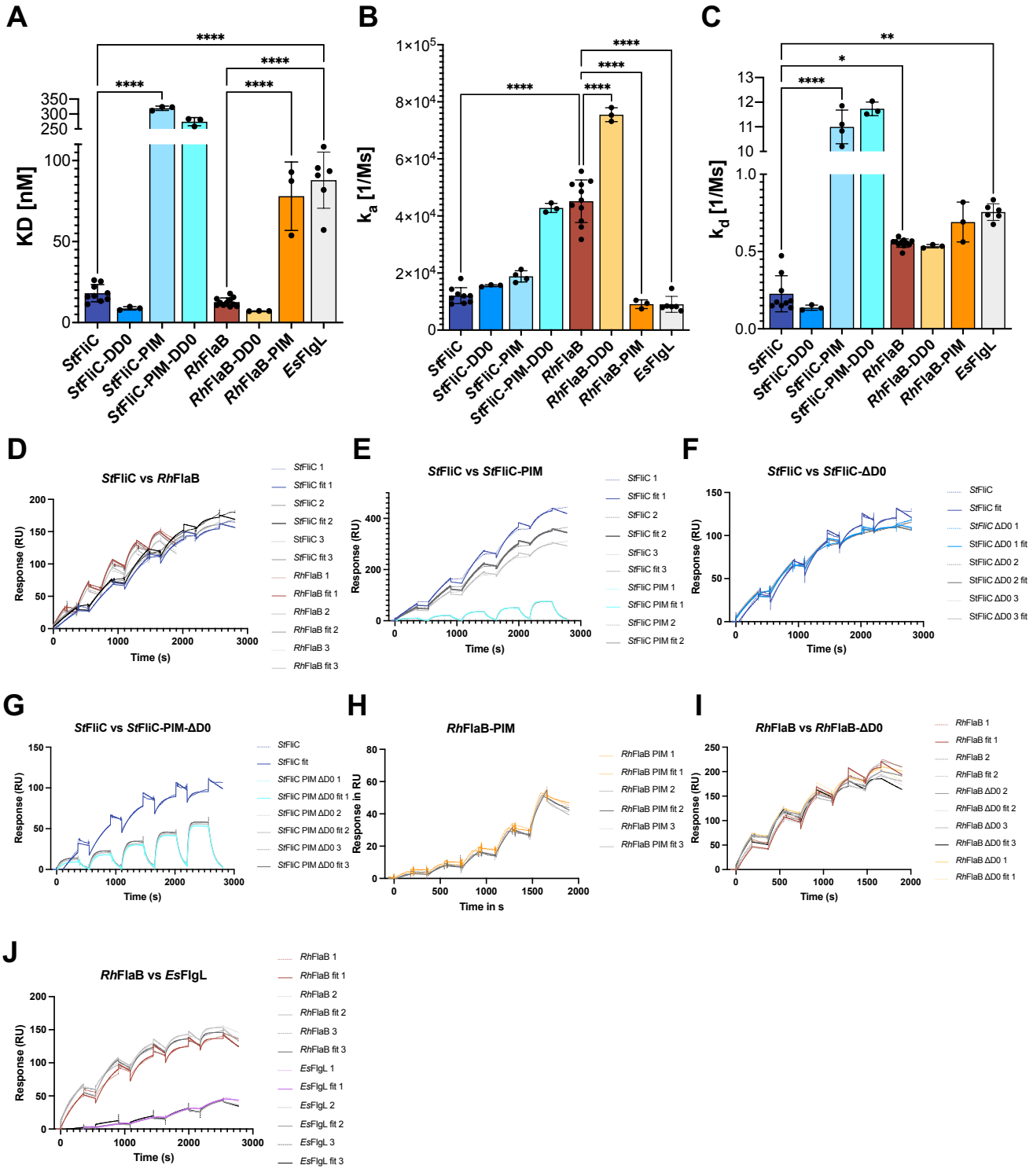
A literature search was conducted for specific residues in *St*FliC implicated in TLR5-Flagellin binding. Numerous *St*FliC residues have been experimentally shown to significantly impact the human TLR5 innate immune response [9,15,43–45]. The corresponding residues in *Rh*FlaB were identified by MSA using Clustal O (1.2.4) (Supp. Fig. 3.7). A final dataset of 21 residues which differed between *St*FliC and *Rh*FlaB were identified, and mapped onto the *St*FliC and *Rh*FlaB structures to verify their structural alignment (Fig. 3.3E).

Analysis of binding interfaces

Superposition complexes were made using ChimeraX, to examine how the experimental structures and binding interfaces aligned with $TLR5^{N14}$. An existing comparable complex, Zebrafish TLR5b (zTLR5b)–*St*FliC (PDB ID #3v47 from [15]), was used for reference. A $TLR5^{N14}$ structure generated with AlphaFold2 was fitted to zTLR5b. The *Rh*FlaB monomer structure and a comparable experimental *St*FliC monomer (PDB ID #6jy0 from [24]) were fitted based on flagellin in the zTLR5b–*St*FliC complex. The superposition showed that the alignment of the flagellins was imperfect, and the flagellin from the complex appeared to differ in confirmation. Consequently, the experimental *Rh*FlaB and *St*FliC structures were fitted either based on the TLR5 epitope, or the secondary interface, as appropriate for the analysis.

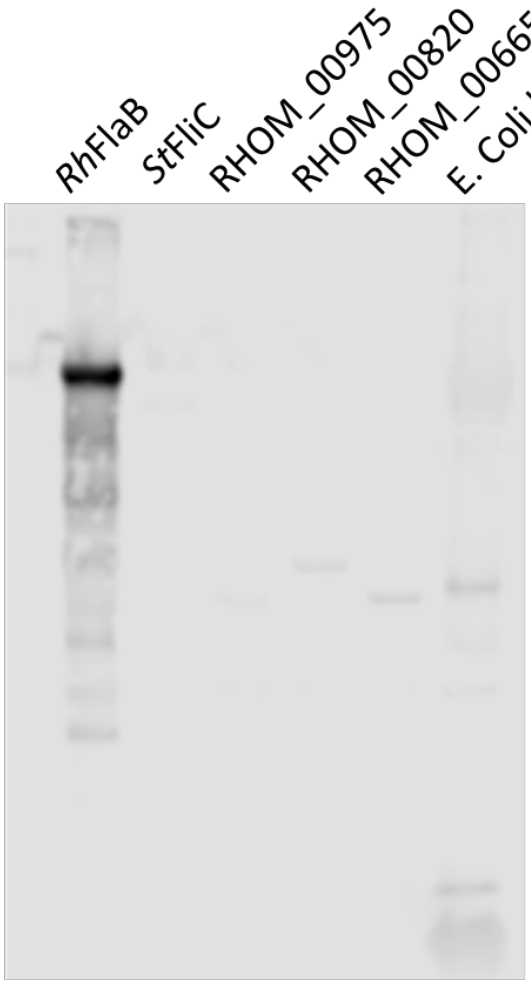
The binding interfaces were analysed in the context of their alignment in superposition complexes. The corresponding interfaces on $TLR5^{N14}$ were identified based on predictions from ChimeraX. For the Primary interface, the PIM residues were adjusted using ChimeraX, to reflect the hydrogen-bond forming rotamer orientations seen in zTLR5b [15]. H-bonds were simulated to confirm the same interactions were expected in human TLR5 (Fig. 3.4C). Hydrophobic and electrostatic surface profiles were generated using ChimeraX to analyse the binding interfaces.

3.6. Supplementary Figures



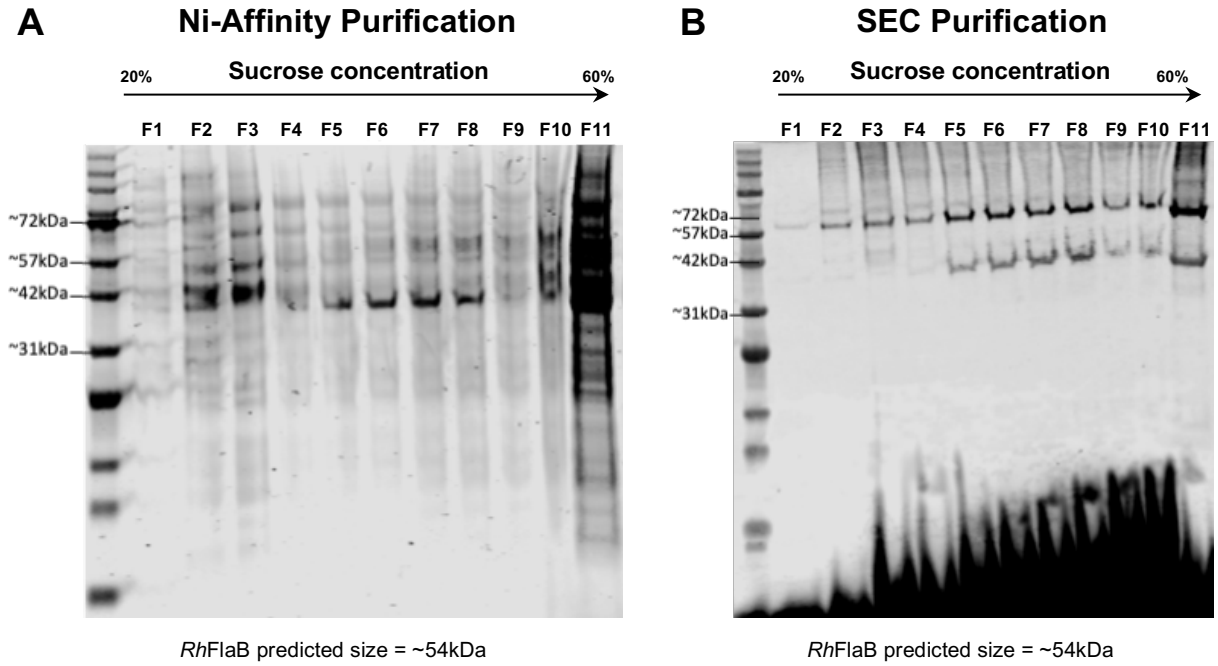
Supplementary Figure 3.1 - Combined Surface Plasmon Resonance results and per-run binding curves.

(A-C) Pooled K_D (A) and kinetics (B-C) results from all SPR runs of all studied flagellins. *St*FliC and *Rh*FlaB were always run as reference against their respective mutants and the negative control, which is reflected in the large number of datapoints for these flagellins. All other flagellin constructs had at least three-technical replicates each. Error bars represent SD. Significance was calculated by one-way ANOVA (***P* < 0.01, ****P* < 0.001, *****P* < 0.0001). (D-J) Raw (dotted lines) and fitted (solid lines) curves from paired SPR analyses. Curve data was generated with the Biacore X100 software (Cytiva) and plotted with Prism 10 (Graphpad).



Supplementary Figure 3.2 – Western blot analysis of cross-reactivity by the Anti-*Rh*FlaB antibody.

The *Rh*FlaB antibody was used to stain a SDS-PAGE gel containing *Rh*FlaB, *St*FliC, three further flagellins from the *R. hominis* genome, and *E. coli* lysate. Bands were stained with a IRDye 800CW Goat anti-Rabbit secondary antibody (LI-COR #32211), and visualised with an Odyssey CLx (LI-COR). A clear band is visible for *Rh*FlaB. Very faint bands are present in the other *R. hominis* flagellins and E. coli, but not at a comparable intensity.

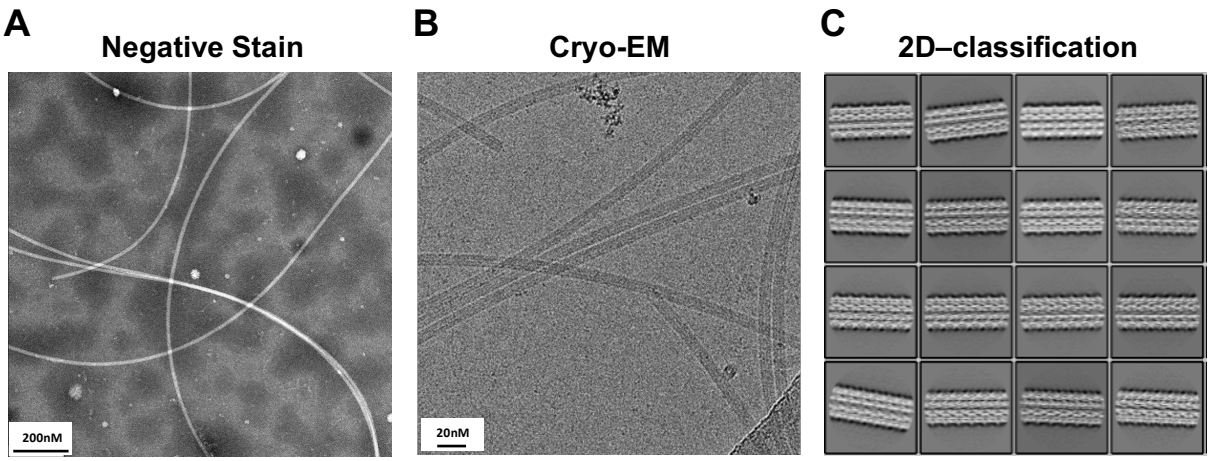


*Rh*FlaB predicted size = ~54kDa

*Rh*FlaB predicted size = ~54kDa

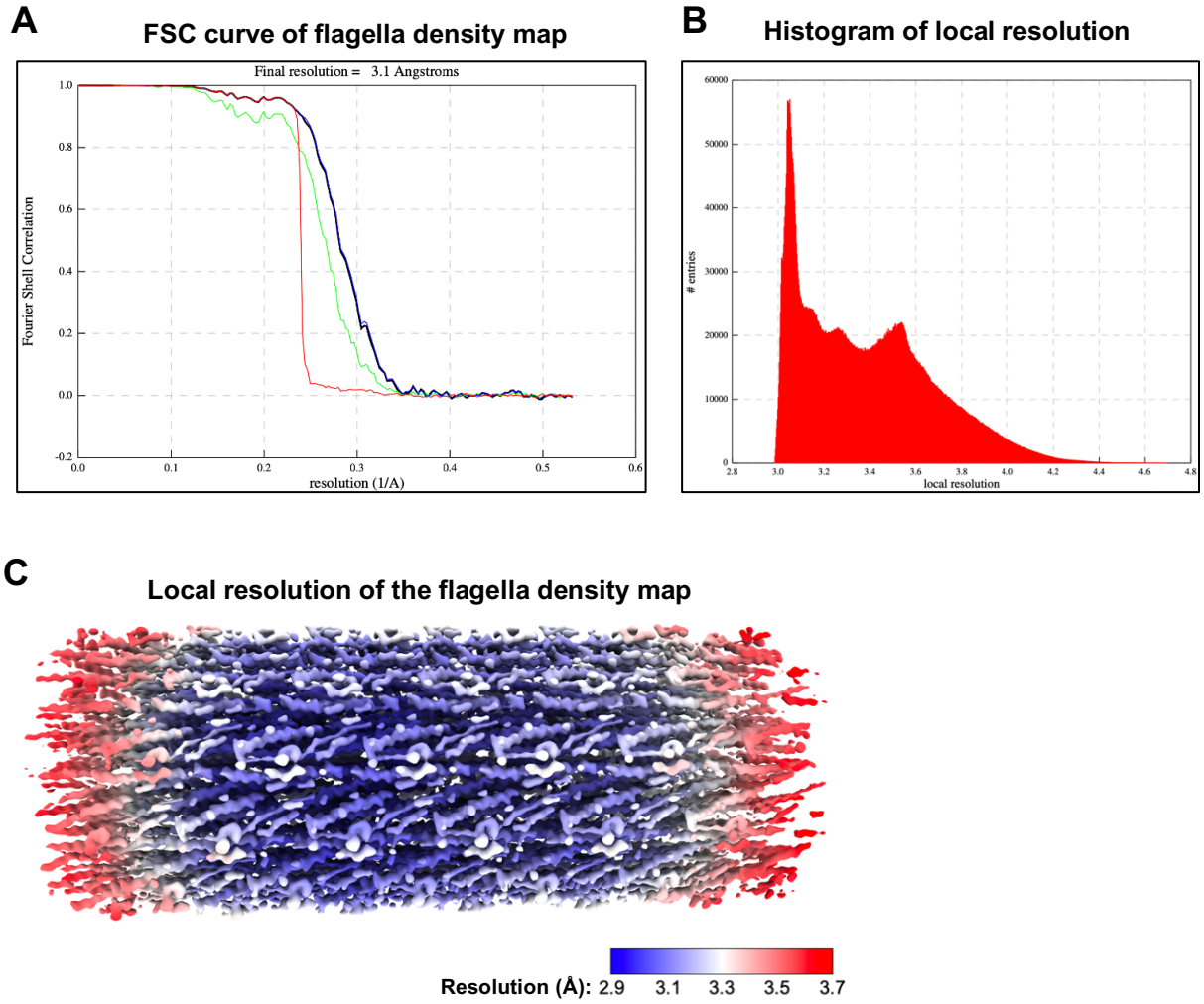
Supplementary Figure 3.3 – SDS-PAGE and western blot analysis of *R. hominis* flagella purified with a sucrose gradient.

(A) A SDS-PAGE gel of fractions taken from the gradient purification of *R. hominis* flagella. The approximate sucrose concentration of each fraction is indicated by the arrow for both gels. Protein bands were observed in all fractions, but distinct dark bands at ~42kDa were identified in fraction 5 to 8 (F5-F8). (B) A western blot of the same fractions showing two bands at ~60kDa and ~42kDa. The darkest bands were observed between F5-F8, corresponding to those identified in the SDS-PAGE gel. These fractions were identified as containing *Rh*FlaB flagella, and utilised for structural analyses.



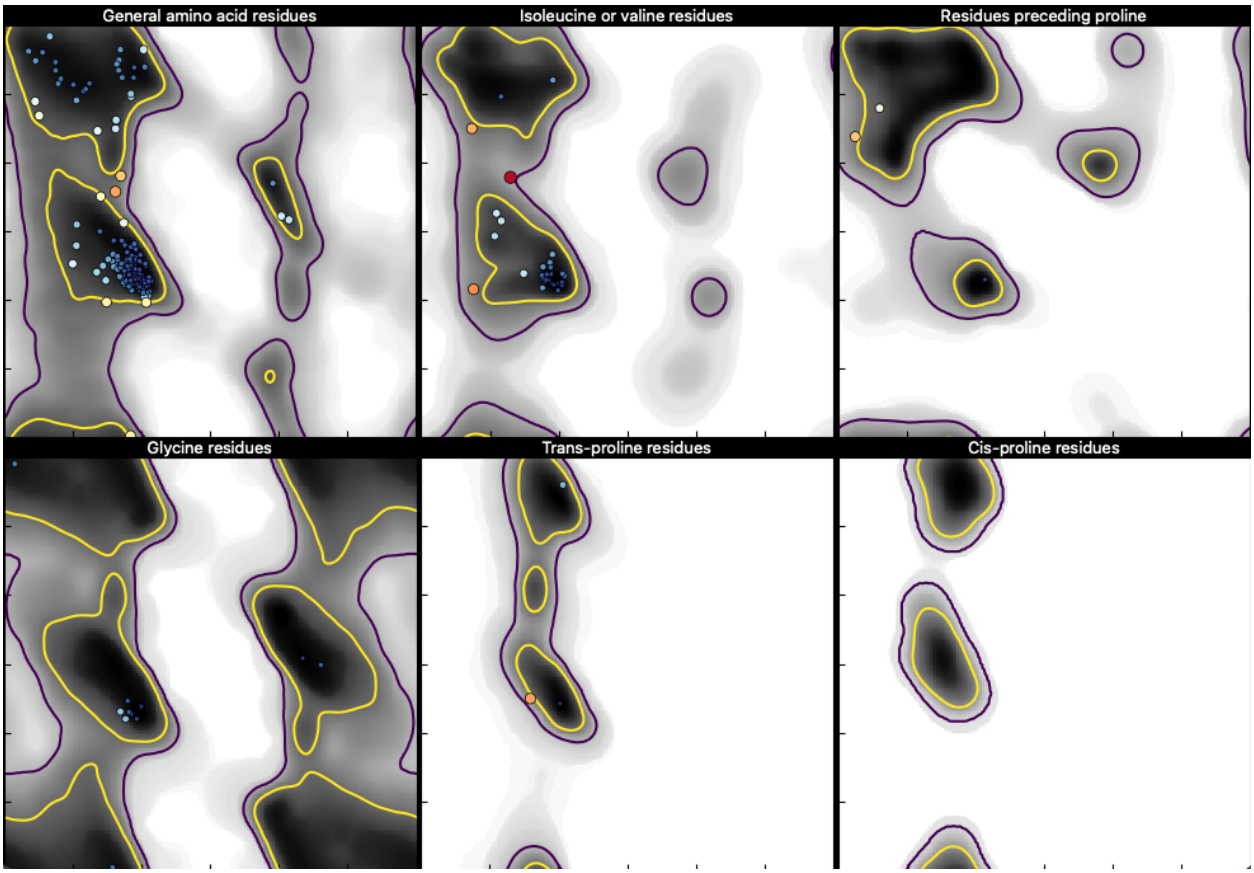
Supplementary Figure 3.4 – Representative Negative stain, Cryo-EM, and 2D-classification images of flagella sheared from *R. hominis*.

(A) Representative negative stain image showing sheared *R. hominis* flagella. Negative staining was used to check for contamination, flagellum density, and flagellum distribution. This sample showed minimal contamination, and well-distributed flagella with few flagella clumping together or overlapping. (B) Representative Cryo-EM image from the dataset. Limited contamination was visible, flagella were present in moderate density, and well distributed throughout the micrograph, facilitating particle extraction. (C) A selection of 'good' 2D classes from images of flagellum particles in Cryo-EM micrographs. These 2D classes display flagella segments with distinct surface features, no contamination, and clear contrast against the background. 2D classes were generated with Relion.



Supplementary Figure 3.5 – Analysis of the variation in resolution of the final *R. hominis* flagellum density map.

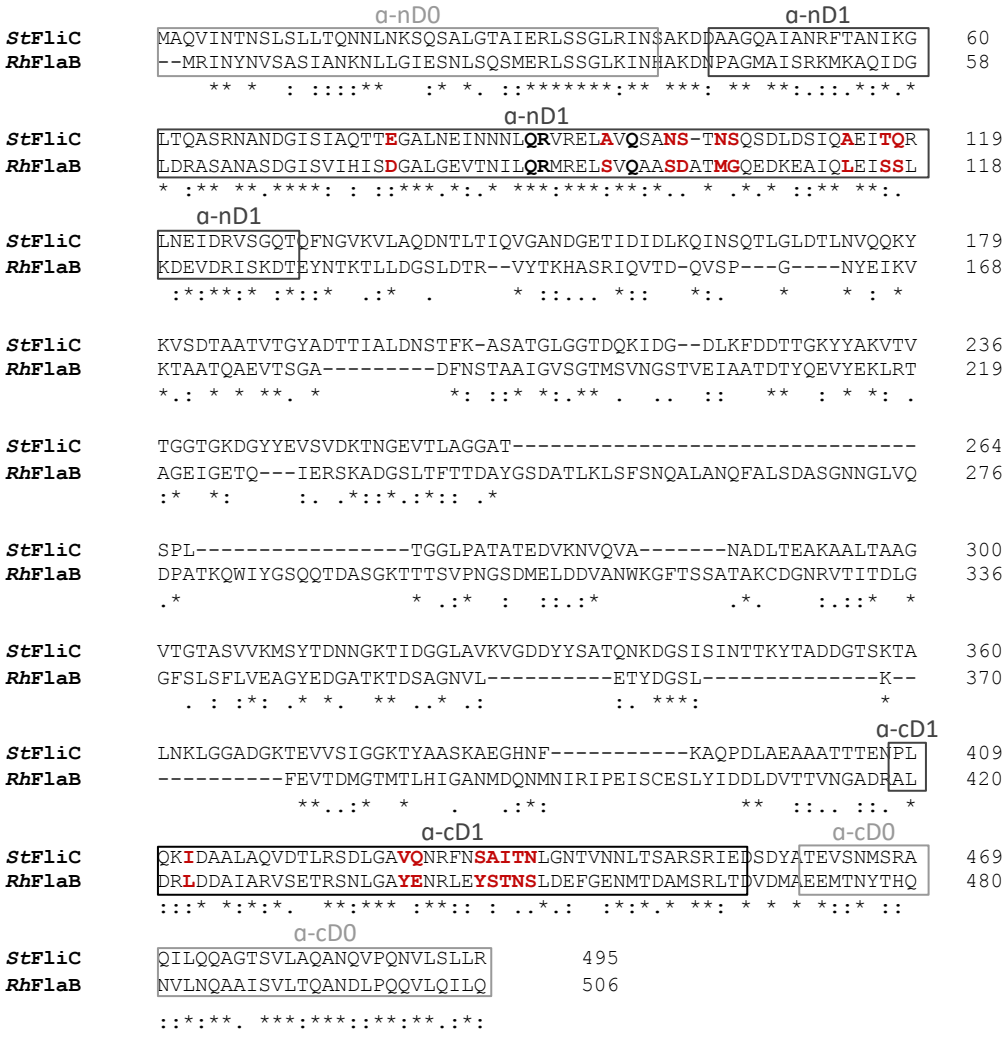
(A) Fourier Shell Correlation curve of the *R. hominis* flagellum density map generated using Relion. A narrow window of resolution is indicated by the rapid drop in the FSC curve. (B) Histogram of local resolution showing the number of points (entries) with the indicated local resolution. The highest cluster of entries were located between 3Å-3.2Å, and almost all points showed a resolution < 4Å. (C) A local resolution map applied to the final post-processed flagella segment. Approximately ~80% of the map is resolved at 2.9Å-3.2Å. Lower resolution regions (> 3.2Å) were located primarily at the ends of the flagella segment. Structure-building was performed within the higher resolution region. The local resolution map was generated with Relion, and visualised on the density map using ChimeraX.



Supplementary Figure 3.6 - Ramachandran plot of the final *Rh*FlaB monomer structure.

The Ramachandran plot displays points representing every Carbon atom in the final *Rh*FlaB structure. Most residues fell within favoured regions, identified by the yellow lines. Less than 1% of residues fell in allowed regions that were not favoured, located between the purple and yellow lines. No residues fell outside of allowed regions. The plot was generated with ISOLDE on ChimeraX [23].

CLUSTAL O(1.2.4) multiple sequence alignment



Supplementary Figure 3.7 - Sequence Alignment analysis of *St*FliC and *Rh*FlaB comparing conserved regions.

The sequence alignment of the *St*FliC and *Rh*FlaB sequences was made using Clustal O (1.2.4), with sequences obtained from UniProt (ID #FLIC_SALTY; ID #G2T5W2_ROSHA). Regions comprising the D0 and D1 domains are highlighted in grey and black respectively. Residues mutated in PIM mutants are indicated with bold black letters. *St*FliC residues known to be involved in TLR5 binding, which differ from those in *Rh*FlaB, are highlighted in red. * = conserved, : = strongly similar, . = weakly similar.

3.7. Supplementary Tables

Pairwise comparison	Mean Difference	Adjusted P Value	Summary
Binding strength (K_D) in nM			
*St*FliC vs. *St*FliC PIM	-566.9	<0,0001	****
*St*FliC vs. *St*FliC dD0	9.356	0.4116	ns
*St*FliC vs. *St*FliC PIM dD0	-256.2	<0,0001	****
*St*FliC vs. *Rh*FlaB	5.475	0.2496	ns
*St*FliC vs. *Rh*FlaB PIM	-59.86	0.1441	ns
*St*FliC vs. *Rh*FlaB dD0	11.05	0.362	ns
*St*FliC vs. *Es*FlgL	-69.75	0.0027	**
*Rh*FlaB vs. *Rh*FlaB PIM	-65.34	0.0916	ns
*Rh*FlaB vs. *Rh*FlaB dD0	5.572	0.1457	ns
*Rh*FlaB vs. *Es*FlgL	-75.22	0.0011	**
Association (k_a) in 10^3/s			
*St*FliC vs. *St*FliC PIM	-6758	0.1003	ns
*St*FliC vs. *St*FliC dD0	-3501	0.7725	ns
*St*FliC vs. *St*FliC PIM dD0	-30734	0.0367	*
*St*FliC vs. *Rh*FlaB	-33085	<0,0001	****
*St*FliC vs. *Rh*FlaB PIM	2981	0.9152	ns
*St*FliC vs. *Rh*FlaB dD0	-63408	0.0114	*
*St*FliC vs. *Es*FlgL	3044	0.6263	ns
*Rh*FlaB vs. *Rh*FlaB PIM	36066	0.0172	*
*Rh*FlaB vs. *Rh*FlaB dD0	-30323	0.0472	*
*Rh*FlaB vs. *Es*FlgL	36129	0.0017	**

Dissociation (k_d) in 10^{-4}/ms			
*St*FliC vs. *St*FliC PIM	-0.01077	<0,0001	****
*St*FliC vs. *St*FliC dD0	9E-05	0.9412	ns
*St*FliC vs. *St*FliC PIM dD0	-0.0115	<0,0001	****
*St*FliC vs. *Rh*FlaB	-0.000328	0.0011	**
*St*FliC vs. *Rh*FlaB PIM	-0.000464	0.3097	ns
*St*FliC vs. *Rh*FlaB dD0	-0.000308	0.351	ns
*St*FliC vs. *Es*FlgL	-0.000529	0.0026	**
*Rh*FlaB vs. *Rh*FlaB PIM	-0.000136	0.5849	ns
*Rh*FlaB vs. *Rh*FlaB dD0	2.03E-05	0.9629	ns
*Rh*FlaB vs. *Es*FlgL	-0.000201	0.0021	**

Supplementary Table 3.1 - Tukey's multiple comparisons test pairwise analysis of SPR data.

Binding and kinetics data were pooled for analysis, and the mean difference between samples calculated using Prism 10 (Graphpad). The significance of differences between samples was determined using an ordinary One-Way ANOVA in Prism 10 (Graphpad). Differences were considered significant if $p < 0.05$. ns = not significant, * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, **** = $p < 0.0001$.

Fitted Curve	K_D in nM	k_a in $10^3 s^{-1}$	k_d in $10^{-4} s^{-1} M^{-1}$
D			
*St*FliC fit 1	15.89	9.39	1.49
*St*FliC fit 2	16.45	9.13	1.50
*St*FliC fit 3	18.23	9.90	1.82
*Rh*FlaB fit 1	12.28	43.76	5.38
*Rh*FlaB fit 2	12.22	45.54	5.57
*Rh*FlaB fit 3	13.97	42.84	5.98
E			
*St*FliC fit 1	13.22	12.64	1.67
*St*FliC fit 2	14.46	11.98	1.73
*St*FliC fit 3	11.28	12.03	1.36
*St*FliC PIM fit 1	570.2	19.37	111.00
*St*FliC PIM fit 2	612.4	16.65	102.00
F			
*St*FliC fit	25.41	14.32	3.64
*St*FliC ΔD0 fit 1	8.80	15.17	1.34
*St*FliC ΔD0 fit 2	9.77	15.85	1.55
*St*FliC ΔD0 fit 3	7.75	15.70	1.22
G			
*St*FliC fit	26.18	18.06	4.73
*St*FliC PIM ΔD0 fit 1	281.10	41.37	116.30
*St*FliC PIM ΔD0 fit 2	283.47	42.50	120.40
*St*FliC PIM ΔD0 fit 3	258.50	44.55	115.20
H			
RhFlaB PIM fit 1	53.77	10.45	5.61
RhFlaB PIM fit 2	92.84	7.44	6.91
RhFlaB PIM fit 3	87.36	9.38	8.20
I			
RhFlaB fit 1	10.72	51.09	5.48
RhFlaB fit 2	10.09	48.54	4.90
RhFlaB ΔD0 fit 1	7.16	73.14	5.23
RhFlaB ΔD0 fit 2	7.10	75.32	5.35
J			
RhFlaB fit 1	17.86	31.78	5.68
RhFlaB fit 2	15.59	36.13	5.63
RhFlaB fit 3	14.62	38.22	5.59
EsFlgL 1	81.91	8.25	6.76
EsFlgL 2	90.85	84.52	7.68
EsFlgL 3	93.45	7.90	7.38

Supplementary Table 3.2 – Individual binding and kinetics results obtained from individual SPR curves.

Fitted curves for each flagellin binding curve were determined with the Biacore X100 software (Cytiva), and used to calculate the Binding strength (K_D), Association rate (k_a), and Dissociation rate (k_d). The results for each fitted SPR binding curve are shown. The letter corresponds to the associated raw and fitted curve displayed in Supplementary Figure 3.1D-J. This table was produced by Miriam Haag.

3.8. Bibliography

1. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol. 2016;14. doi:10.1371/journal.pbio.1002533

2. Savage DC. Microbial Ecology of the Gastrointestinal Tract. Annu Rev Microbiol. 1977;31: 107–133.

3. Kennedy MS, Chang EB. The microbiome: composition and locations. Prog Mol Biol Transl Sci. 2020;176: 1.

4. Hou K, Wu Z-X, Chen X-Y, Wang J-Q, Zhang D, Xiao C, et al. Microbiota in health and diseases. Signal Transduction and Targeted Therapy. 2022;7: 1–28.

5. Fitzgerald KA, Kagan JC. Toll-like Receptors and the Control of Immunity. Cell. 2020;180: 1044–1066.

6. Dempsey PW, Vaidya SA, Cheng G. The art of war: Innate and adaptive immune responses. Cell Mol Life Sci. 2003;60: 2604–2621.

7. Akira S, Uematsu S, Takeuchi O. Pathogen recognition and innate immunity. Cell. 2006;124: 783–801.

8. Forstnerič V, Ivičak-Kocjan K, Plaper T, Jerala R, Benčina M. The role of the C-terminal D0 domain of flagellin in activation of Toll like receptor 5. PLoS Pathog. 2017;13: e1006574.

9. Song WS, Jeon YJ, Namgung B, Hong M, Yoon S-I. A conserved TLR5 binding and activation hot spot on flagellin. Sci Rep. 2017;7: 40878.

10. Feuillet V, Medjane S, Mondor I, Demaria O, Pagni PP, Galán JE, et al. Involvement of Toll-like receptor 5 in the recognition of flagellated bacteria. Proc Natl Acad Sci U S A. 2006;103: 12487–12492.

11. Kawasaki T, Kawai T. Toll-like receptor signaling pathways. Front Immunol. 2014;5: 461.

12. Lin S-C, Lo Y-C, Wu H. Helical assembly in the MyD88–IRAK4–IRAK2 complex in TLR/IL-1R signalling. Nature. 2010;465: 885–890.

13. Oda K, Kitano H. A comprehensive map of the toll-like receptor signaling network. Mol Syst Biol. 2006;2: 2006.0015.

14. Clasen SJ, Bell MEW, Borbón A, Lee D-H, Henseler ZM, de la Cuesta-Zuluaga J, et al. Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. Sci Immunol. 2023;8: eabq7001.

15. Yoon SI, Kurnasov O, Natarajan V, Hong M, Gudkov AV, Osterman AL, et al. Structural basis of TLR5-flagellin recognition and signaling. Science. 2012;335. doi:10.1126/science.1215584

16. Smith KD, Andersen-Nissen E, Hayashi F, Strobe K, Bergman MA, Barrett SLR, et al. Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. Nat Immunol. 2003;4: 1247–1253.

17. Andersen-Nissen E, Smith KD, Strobe KL, Barrett SLR, Cookson BT, Logan SM, et al. Evasion of Toll-like receptor 5 by flagellated bacteria. Proc Natl Acad Sci U S A. 2005;102: 9247–9252.

18. Kim J-H, Namgung B, Jeon YJ, Song WS, Lee J, Kang SG, et al. Helicobacter pylori flagellin: TLR5 evasion and fusion-based conversion into a TLR5 agonist. Biochem Biophys Res Commun. 2018;505: 872–878.

19. Latz E, Verma A, Visintin A, Gong M, Sirois CM, Klein DCG, et al. Ligand-induced conformational changes allosterically activate Toll-like receptor 9. Nat Immunol. 2007;8: 772–779.

20. Yamaguchi T, Toma S, Terahara N, Miyata T, Ashihara M, Minamino T, et al. Structural and Functional Comparison of Salmonella Flagellar Filaments Composed of FljB and FliC. Biomolecules. 2020;10: 246.

21. Gasteiger E, Hoogland C, Gattiker A, Duvaud S 'everine, Wilkins MR, Appel RD, et al. Protein Identification and Analysis Tools on the ExPASy Server. The Proteomics Protocols Handbook. 2005; 571–607.

22. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.

23. Croll TI. ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. Acta Crystallographica Section D: Structural Biology. 2018;74: 519–530.

24. Mimori Y, Yamashita I, Murata K, Fujiyoshi Y, Yonekura K, Toyoshima C, et al. The structure of the R-type straight flagellar filament of Salmonella at 9 A resolution by electron cryomicroscopy. J Mol Biol. 1995;249: 69–87.

25. Blum TB, Filippidou S, Fatton M, Junier P, Abrahams JP. The wild-type flagellar filament of the Firmicute Kurthia at 2.8 Å resolution in vivo. Sci Rep. 2019;9: 14948.

26. Bordner AJ. Predicting protein-protein binding sites in membrane proteins. BMC Bioinformatics. 2009;10: 312.

27. Kotha SR, Staller MV. Clusters of acidic and hydrophobic residues can predict acidic transcriptional activation domains from protein sequence. Genetics. 2023;225. doi:10.1093/genetics/iyad131

28. Martinoli C, Chiavelli A, Rescigno M. Entry Route of Salmonella typhimurium Directs the Type of Induced Immune Response. Immunity. 2007;27: 975–984.

29. Li Q. Mechanisms for the Invasion and Dissemination of. Can J Infect Dis Med Microbiol. 2022;2022: 2655801.

30. Gatsogiannis C, Merino F, Prumbaum D, Roderer D, Leidreiter F, Meusch D, et al. Membrane insertion of a Tc toxin in near-atomic detail. Nat Struct Mol Biol. 2016;23: 884–890.

31. Xu Y, Wang H, Nussinov R, Ma B. Protein charge and mass contribute to the spatio-temporal dynamics of protein-protein interactions in a minimal proteome. Proteomics. 2013;13: 1339–1351.

32. Estrada VE, Zhang N, Wennerström H, Danielsson J, Oliveberg M. Diffusive intracellular interactions: On the role of protein net charge and functional adaptation. Curr Opin Struct Biol. 2023;81: 102625.

33. He W, Zhang J, Sachsenhauser V, Wang L, Bardwell JCA, Quan S. Increased surface charge in the protein chaperone Spy enhances its anti-aggregation activity. J Biol Chem. 2020;295: 14488–14500.

34. Farías-Rico JA, Ruud Selin F, Myronidi I, Frühauf M, von Heijne G. Effects of protein size, thermodynamic stability, and net charge on cotranslational folding on the ribosome. Proc Natl Acad Sci U S A. 2018;115: E9280–E9287.

35. Kretzschmar T, Aoustin L, Zingel O, Marangi M, Vonach B, Towbin H, et al. High-level expression in insect cells and purification of secreted monomeric single-chain Fv antibodies. J Immunol Methods. 1996;195: 93–101.

36. Stabrin M, Schoenfeld F, Wagner T, Pospich S, Gatsogiannis C, Raunser S. TranSPHIRE: automated and feedback-optimized on-the-fly processing for cryo-EM. Nat Commun. 2020;11: 1–14.

37. Wagner T, Merino F, Stabrin M, Moriya T, Antoni C, Apelbaum A, et al. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Communications Biology. 2019;2: 1–13.

38. Scheres SHW. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012;180: 519–530.

39. He S, Scheres SHW. Helical reconstruction in RELION. J Struct Biol. 2017;198: 163–176.

40. Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. 2021;30: 70–82.

41. Travis AJ, Kelly D, Flint HJ, Aminov RI. Complete Genome Sequence of the Human Gut Symbiont Roseburia hominis. Genome Announc. 2015. doi:10.1128/genomea.01286-15

42. MacDougall A, Volynkin V, Saidi R, Poggioli D, Zellner H, Hatton-Ellis E, et al. UniRule: a unified rule resource for automatic annotation in the UniProt Knowledgebase. Bioinformatics. 2020;36: 4643–4648.

43. Andersen-Nissen E, Smith KD, Bonneau R, Strong RK, Aderem A. A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. J Exp Med. 2007;204: 393–403.

44. Jacchieri SG, Torquato R, Brentani RR. Structural Study of Binding of Flagellin by Toll-Like Receptor 5. J Bacteriol. 2003;185: 4243.

45. Murthy KGK, Deb A, Goonesekera S, Szabó C, Salzman AL. Identification of Conserved Domains in Salmonella muenchen Flagellin That Are Essential for Its Ability to Activate TLR5 and to Induce an Inflammatory Response in Vitro. J Biol Chem. 2004;279: 5667–5675.

Chapter 4: Exploring the role of conformation in TLR5-TIR domain signalling specificity

4.1. Introduction

Toll-like receptors (TLRs) are a class of Type-1 transmembrane pattern recognition receptors that provide a first line of defence against pathogens, serving as a key component of the innate immune system. Among these is Toll-like Receptor 5 (TLR5), which specifically recognizes and responds to extracellular bacterial flagellin. Flagellin is the principal protein component of the bacterial flagellum, a bacterial organelle essential for motility and invasion [1–4]. Additionally, TLR5 plays a role in cell homeostasis, with disruptions to its signalling pathway linked to inflammatory conditions such as inflammatory bowel disease (IBD) [5,6].

The TLR5 signalling pathway has been well-characterised, primarily through extrapolation from other TLRs. The binding of flagellin to TLR5 is believed to induce a conformational change in the receptor, leading to the recruitment of the adaptor protein myeloid differentiation primary response 88 (MyD88) [7,8]. MyD88 then recruits several IL-1 receptor-associated kinases (IRAKs), specifically IRAK4 and IRAK1, forming a complex known as the Myddosome. This complex recruits the tumour necrosis factor receptor-associated factor 6 (TRAF6), resulting in the phosphorylation and degradation of IκB proteins, thereby activating the nuclear factor-kappa B (NF-κB) pathway. Active NF-κB translocates into the nucleus, promoting the transcription of immune factors, including pro-inflammatory cytokines, which facilitate an immune response to clear bacterial infection [9–13].

Central to this pathway is the TLR5 Toll/Interleukin-1 Receptor (TIR) domain, a conserved region located in the cytoplasmic portion of TLR5 [14–16]. In TLR5 heterodimers, the TIR domains come into close proximity and dimerise, facilitating MyD88 recruitment [1,17]. Numerous studies have characterised the TIR domains of various TLRs, their dimerisation interfaces and MyD88 binding sites. The structure of a putative TLR10-TIR dimer has also been solved [18–20]. However, only one mutagenic study has specifically investigated the TLR5-TIR,

and no experimental TLR5-TIR structures exist [15]. Understanding the structural and functional dynamics of the TLR5-TIR domain during dimerisation, ligand-induced conformational change, and MyD88 recruitment is crucial for elucidating how TLR5 mediates innate immune responses, and developing targeted interventions for the TLR5 signalling pathway.

In this chapter, I present a project aimed at generating TLR5-TIR dimers in a conformation capable of triggering the NF-κB pathway. I hypothesised that flagellin binding to TLR5 initiates a conformational change that allows the TIR domains to dimerise in an 'active' conformation. Only active dimers should be capable of binding MyD88, preventing unliganded TLR5 homodimers from provoking an immune response [7,17]. Artificial TLR5-TIR dimers were generated using GCN4 coiled-coils, which have a rotation period of 7 amino acids, allowing the attached TIR domain to be expressed at approximately 51º intervals [21,22]. These constructs were transfected into HEK cells to study their impact on TLR5 pathway activation.

Ultimately, no GCN4-TIR construct was able to stimulate the TLR5 signalling pathway at an intensity comparable to the canonical TLR5 flagellin agonist *Salmonella typhimurium* FliC (*St*FliC). While a narrow range of rotations displayed limited ability to trigger the TLR5 signalling pathway, none were sufficiently promising for further pursuit in this dissertation. However, a series of troubleshooting experiments were performed to identify avenues for future improvement. If an active dimer were found, it would support the theory of flagellin-induced conformational change in initiating the TLR5 signalling pathway, and provide insights into mechanisms by which silent flagellins could bind TLR5 without provoking such conformational changes required for signalling [7].

4.2. Results

4.2.1. The design of GCN4 coiled-coil constructs

The GCN4 leucine zipper sequence was obtained from the literature, and seven variants at each possible rotational conformation were made by the successive removal of single amino acids (Table 4.1) [21]. The TIR domain of human TLR5 was identified by analysis of the existing TLR5 consensus sequence (NCBI ID #CCDS31033.1), and validated against a structural model of TLR5 (Fig. 4.1A) [29], then added to the C-terminus of each GCN4 coiled coil sequence with a short linker. A plasma membrane (PM) targeting sequence from the *Rous sarcoma* virus was added to the N-terminus, to mimic the *in vitro* localisation of the TLR5-TIR [23]. A final C-terminal TEV-protease sequence was added prior to 6xHis and V5 affinity tags. The expected TIR dimer complex was predicted by AlphaFold2 Multimer [30] (Fig. 4.1B), and the predicted folding of the GCN4-TIR aligned with that of a TLR5-TIR model generated using AlphaFold2 (Fig. 4.1C).

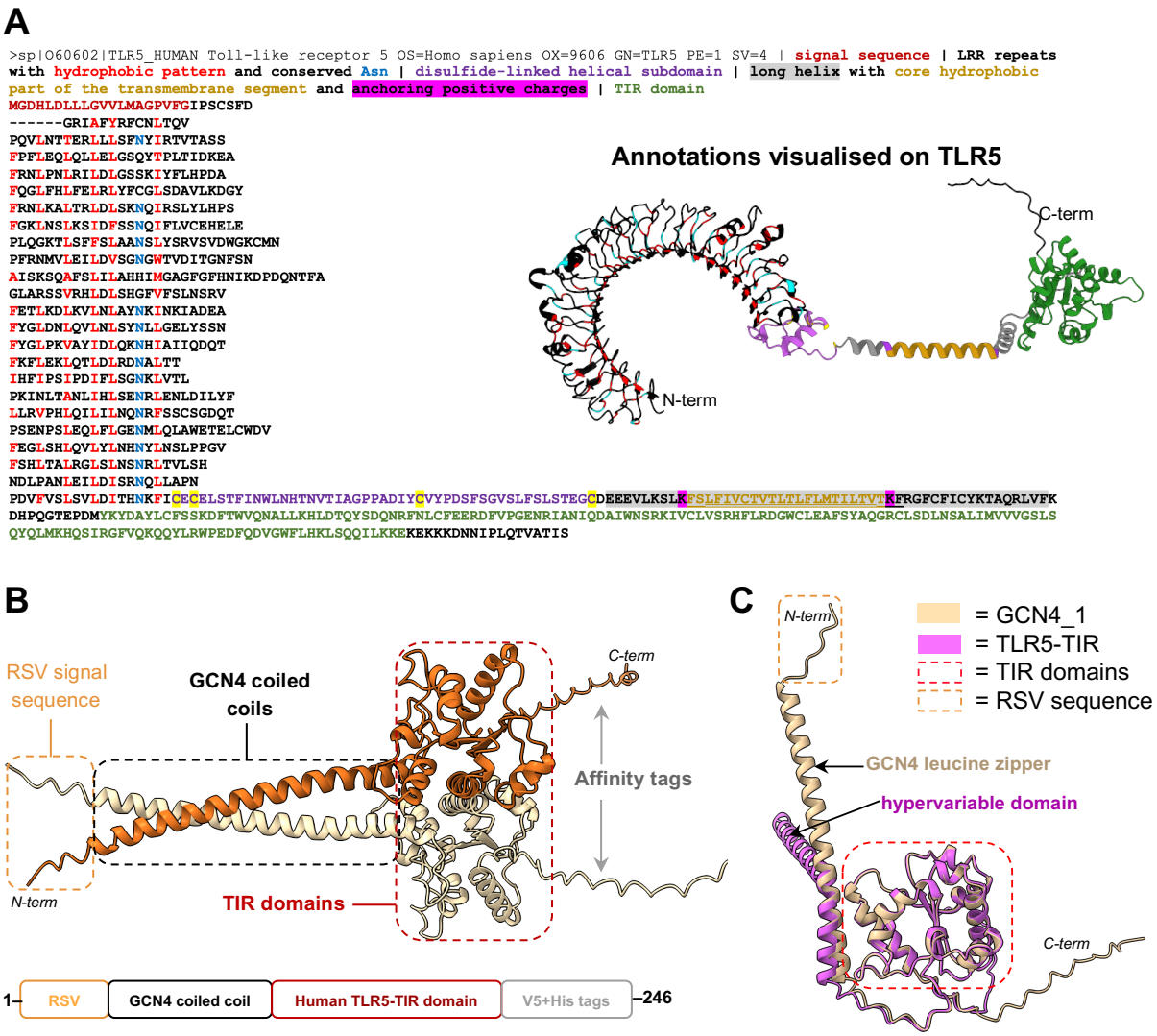


Figure 4.1 – TLR5 sequence annotation and construct validations using predicted GCN4-TIR models.

(A) A sequence analysis and annotation of human TLR5, highlighting key features including the hydrophobic regions and the TIR domain. The annotations were overlaid on a human TLR5 model generated with AlphaFold2 [29]. Colours on the model correspond to those used in the annotation. The TIR domain sequence identified was used in the GCN4-TIR construct. (B) An Alphafold2 model of the predicted GCN4-TIR dimer, generated with Alphafold2 multimer [30]. Key features of the structure are identified on the structure, and in the simplified map. (C) Superposition of GCN4_1 and the TLR5-TIR domain. Both structures were generated with Alphafold2, and display a near identical predicted secondary structure.

GCN4-TIR construct ID	Annotated sequences (RSV tag, GCN4 sequence, Linker regions, TIR sequence, and V5 tag)
GCN4_1	MGSSKSKPKDMKQLEDKVEELLSKNYHLENEVARLKKLVGERFRGFCFICYKTAQRLVFKDHPQGTEPDMYKYDAYLCFSSKDFTWVQNALLKHLDTQYSDQNRFNLCFEERDFVPGENRIANIQDAIWNSRKIVCLVSRHFLRDGWCLEAFSYAQGRCLSDLNSALIMVVVGSLSQYQLMKHQSIRGFVQKQQYLRWPEDFQDVGWFLHKLSQQILKKEKEKKKDNNIPLQTVATISGKPIPNPLLGLDST
GCN4_2	MGSSKSKPKDMKQLEDKVEELLSKNYHLENEVARLKKLVGERRGFCFICYKTAQRLVFKDHPQGTEPDMYKYDAYLCFSSKDFTWVQNALLKHLDTQYSDQNRFNLCFEERDFVPGENRIANIQDAIWNSRKIVCLVSRHFLRDGWCLEAFSYAQGRCLSDLNSALIMVVVGSLSQYQLMKHQSIRGFVQKQQYLRWPEDFQDVGWFLHKLSQQILKKEKEKKKDNNIPLQTVATISGKPIPNPLLGLDST
GCN4_3	MGSSKSKPKDMKQLEDKVEELLSKNYHLENEVARLKKLVGERGFCFICYKTAQRLVFKDHPQGTEPDMYKYDAYLCFSSKDFTWVQNALLKHLDTQYSDQNRFNLCFEERDFVPGENRIANIQDAIWNSRKIVCLVSRHFLRDGWCLEAFSYAQGRCLSDLNSALIMVVVGSLSQYQLMKHQSIRGFVQKQQYLRWPEDFQDVGWFLHKLSQQILKKEKEKKKDNNIPLQTVATISGKPIPNPLLGLDST
GCN4_4	MGSSKSKPKDMKQLEDKVEELLSKNYHLENEVARLKKLVGERFCFICYKTAQRLVFKDHPQGTEPDMYKYDAYLCFSSKDFTWVQNALLKHLDTQYSDQNRFNLCFEERDFVPGENRIANIQDAIWNSRKIVCLVSRHFLRDGWCLEAFSYAQGRCLSDLNSALIMVVVGSLSQYQLMKHQSIRGFVQKQQYLRWPEDFQDVGWFLHKLSQQILKKEKEKKKDNNIPLQTVATISGKPIPNPLLGLDST
GCN4_5	MGSSKSKPKDMKQLEDKVEELLSKNYHLENEVARLKKLVGERCFICYKTAQRLVFKDHPQGTEPDMYKYDAYLCFSSKDFTWVQNALLKHLDTQYSDQNRFNLCFEERDFVPGENRIANIQDAIWNSRKIVCLVSRHFLRDGWCLEAFSYAQGRCLSDLNSALIMVVVGSLSQYQLMKHQSIRGFVQKQQYLRWPEDFQDVGWFLHKLSQQILKKEKEKKKDNNIPLQTVATISGKPIPNPLLGLDST
GCN4_6	MGSSKSKPKDMKQLEDKVEELLSKNYHLENEVARLKKLVGERFICYKTAQRLVFKDHPQGTEPDMYKYDAYLCFSSKDFTWVQNALLKHLDTQYSDQNRFNLCFEERDFVPGENRIANIQDAIWNSRKIVCLVSRHFLRDGWCLEAFSYAQGRCLSDLNSALIMVVVGSLSQYQLMKHQSIRGFVQKQQYLRWPEDFQDVGWFLHKLSQQILKKEKEKKKDNNIPLQTVATISGKPIPNPLLGLDST
GCN4_7	MGSSKSKPKDMKQLEDKVEELLSKNYHLENEVARLKKLVGERICYKTAQRLVFKDHPQGTEPDMYKYDAYLCFSSKDFTWVQNALLKHLDTQYSDQNRFNLCFEERDFVPGENRIANIQDAIWNSRKIVCLVSRHFLRDGWCLEAFSYAQGRCLSDLNSALIMVVVGSLSQYQLMKHQSIRGFVQKQQYLRWPEDFQDVGWFLHKLSQQILKKEKEKKKDNNIPLQTVATISGKPIPNPLLGLDST

Table 4.1 – Complete annotated sequence of all GCN4-TIR constructs.

The complete amino acid sequence for each GCN4-TIR construct (GCN4_1-7) is shown, annotated with the key features described in Figure 4.1B. The segment highlighted in grey is has a single deletion in each successive construct, allowing the TIR to be expressed seven different rotational conformations.

4.2.2. Transient Transfections of HEK-Blue TLR5 reporter cells

GCN4_TIR constructs (GCN4_1-7) and a non-stimulatory di-ubiquitin negative control, were cloned into pEF6 plasmids, sequence verified, then transiently transfected into HEK-Blue reporter cells. Preliminary analysis of NF-κB pathway activation in transfected cells, using the QuantiBlue assay, displayed clear differences in activity between samples. However, numerous experiments showed highly inconsistent results (Trial 1-7 in Fig. 4.2). Substantial variation was observed between the response in GCN4-TIR transfected cells and the negative control (baseline) between experiments, for each construct (Fig. 4.2). No GCN4-TIR construct displayed a consistently higher response, when normalised to the negative control.

Qualitative observations of each experiment highlighted potential causes for the observed inconsistency. Widespread cell death was apparent in transfected samples compared to controls, though the extent both within and between experiments was highly variable (Fig. 4.3A). Cell-death was also observed in transfection controls with no plasmid, implying this was construct-independent. Base level IL-8 expression in samples with more cell death was generally lower, and generally corresponded to the intensity of protein bands from western blots of cell lysate extracts (Fig. 4.3B). However, normalising the absorbance data against band intensity did not improve consistency between runs. Immunofluorescence microscopy showed a weak signal for the anti-V5 tag, which was absent from a non-stained control (Fig. 4.3C). However, these were not consistently visible in composite images, and were not clearly localised to the outer membrane (Fig. 4.3D).

In order to improve consistency and signal, I attempted to optimise the transfection conditions by varying the plasmid and transfection reagent load, and the cell volume. However, the inconsistencies remained. Instead, I chose to generate stably transfected cells, with protein expression controlled by an inducible promoter.

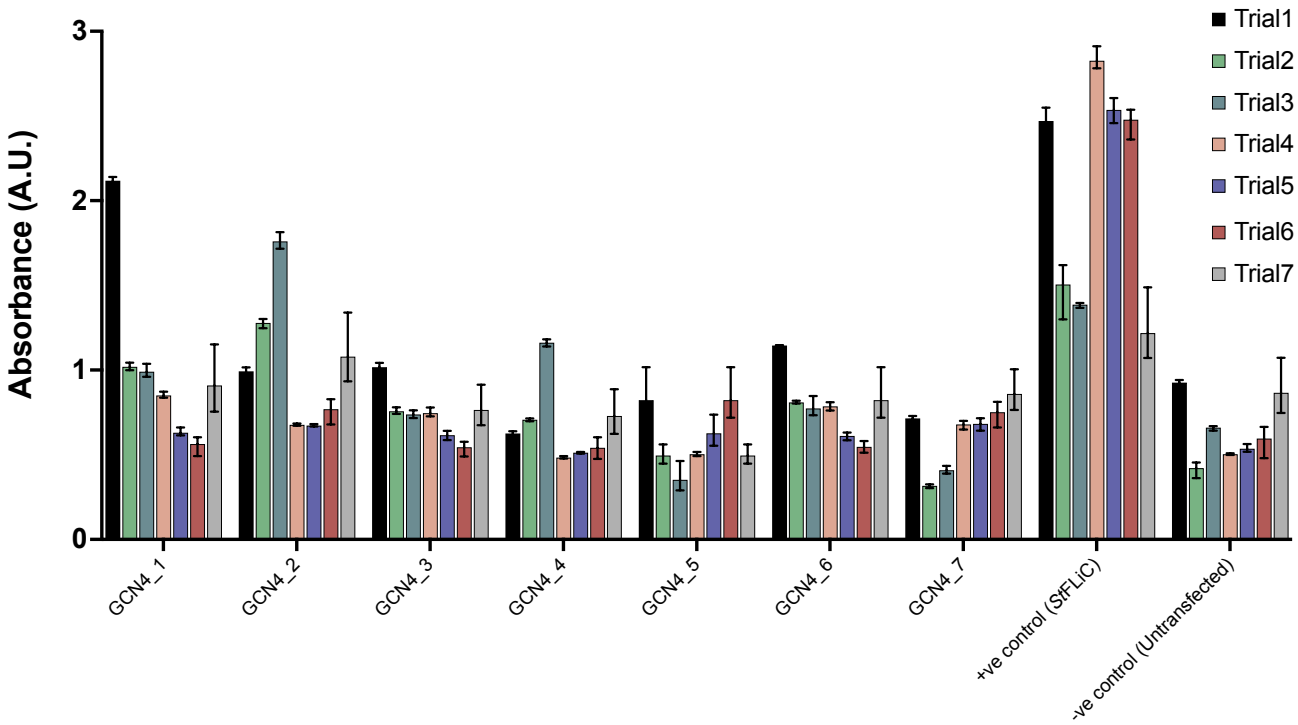


Figure 4.2 – GCN4-TIR induced TLR5 pathway activation in transiently transfected HEK-Blue cells.

Pooled results of 7 experimental replicates measuring the NF-κβ pathway activity transiently transfected HEK-Blue cells. Each bar represents the mean of 3 technical replicates per protein in each experiment. Error bars represent range. Analysis and visualisation were performed with Prism 10 (Graphpad).

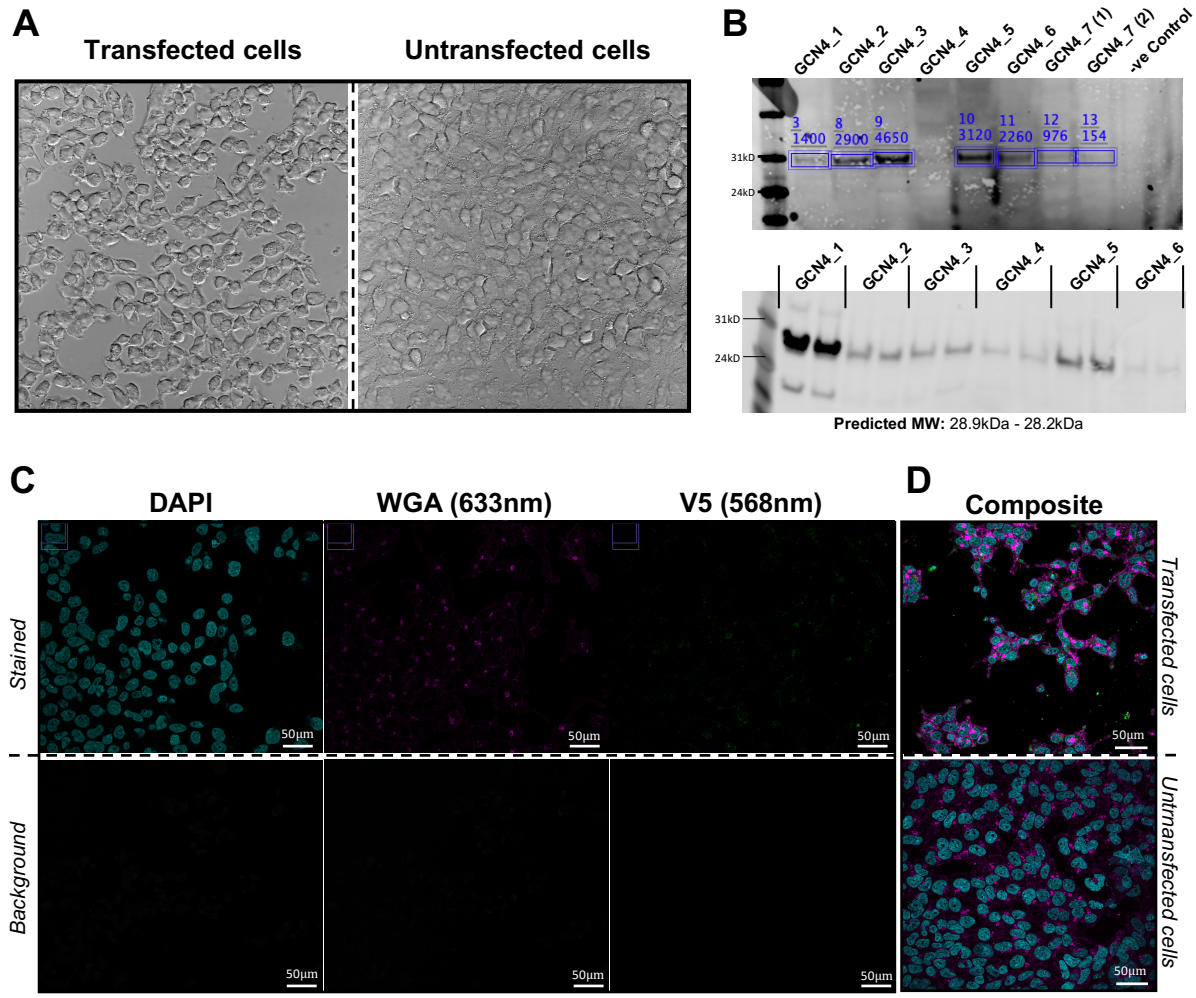


Figure 4.3 – Troubleshooting Transient Transfections of HEK-Blue cells with GCN4-TIR constructs.

(A) Representative light microscopy images comparing transfected and untransfected HEK-Blue cells, highlighting differences in cell density and activity. (B) Anti-His western blots on cell lysates from two separate experiments. The top blot displays transfections of each GCN4-TIR construct, and the bottom displays two replicates of several GCN4-TIR transfections. Samples within each blot were transfected in the same 24-well plate. Band intensity was calculated with Image studio (LI-COR). (C) Immunofluorescence microscopy image on stained and un-stained HEK-Blue cells transiently transfected with GCN4_1, used to confirm the fluorescence wasn't from the background. Stains targeted the nucleus (DAPI), plasma membrane (WGA), and GCN4_7 (V5). (D) Composite images comparing the V5 signal in transfected and untransfected HEK-Blue cells. All microscopy images were processed with FIJI (ImageJ).

4.2.3 Stable Transfections of Flp-In™ T-REx™ HEK-293 cells

GCN4-TIR constructs were cloned into pcDNA5/FRT/V5-His-TOPO™ plasmids (Invitrogen), then stably transfected into Flp-In™ T-REx™ HEK-293 cells (Invitrogen) along with a CAT+ transfection control plasmid. Tetracycline (Tet) induced expression in stable cell lines was confirmed by western blot, though double bands were observed in each GCN4-TIR construct (Fig. 4.4A). No bands were observed in the CAT+ control plasmid, as expected.

As this cell line was not compatible with the Quanti-Blue assay, NF-kB pathway activation was measured with an ELISA targeting human Interleukin 8 (IL-8). The baseline varied substantially between experiments, but was consistent between samples within each experiment (Fig. 4.4B). No substantial differences in IL-8 expression were observed following the normalisation of induced samples to paired uninduced controls (Fig. 4.4C). Both GCN4_7 and GCN4_1 showed a clear increase against baseline and the CAT+ and non-transfected controls. The largest increase of 4.11 pg/mL was observed in the latter, which was low compared to the 50.7 pg/mL increase in the positive control stimulated with *St*FliC.

4.2.4. Troubleshooting GCN4 construct production & expression

Several troubleshooting steps were taken to validate the experimental theory behind this study, and identify potential improvements to the methodology. Native-PAGE blots were run on both tris-glycine and tris-acetate gels to see if dimers could be observed, but displayed only high molecular weight smears which did not resolve into separate bands (Fig. 4.5A). Immunofluorescence microscopy was used to check if GCN4-TIR constructs localised to the PM, but I was unable to clearly distinguish a GCN4-TIR signal from the background (Fig. 4.5B). SDS-PAGE analysis of HEK-cell lysates showed no clear bands at the expected size for GCN4-TIR constructs, indicating that low protein expression may be causal to the lack of signal, despite clear western blot bands (Fig. 4.5C). Thus, several transfection, induction, and expression conditions were trialled, but did not noticeably improve apparent expression (Fig. 4.5D).

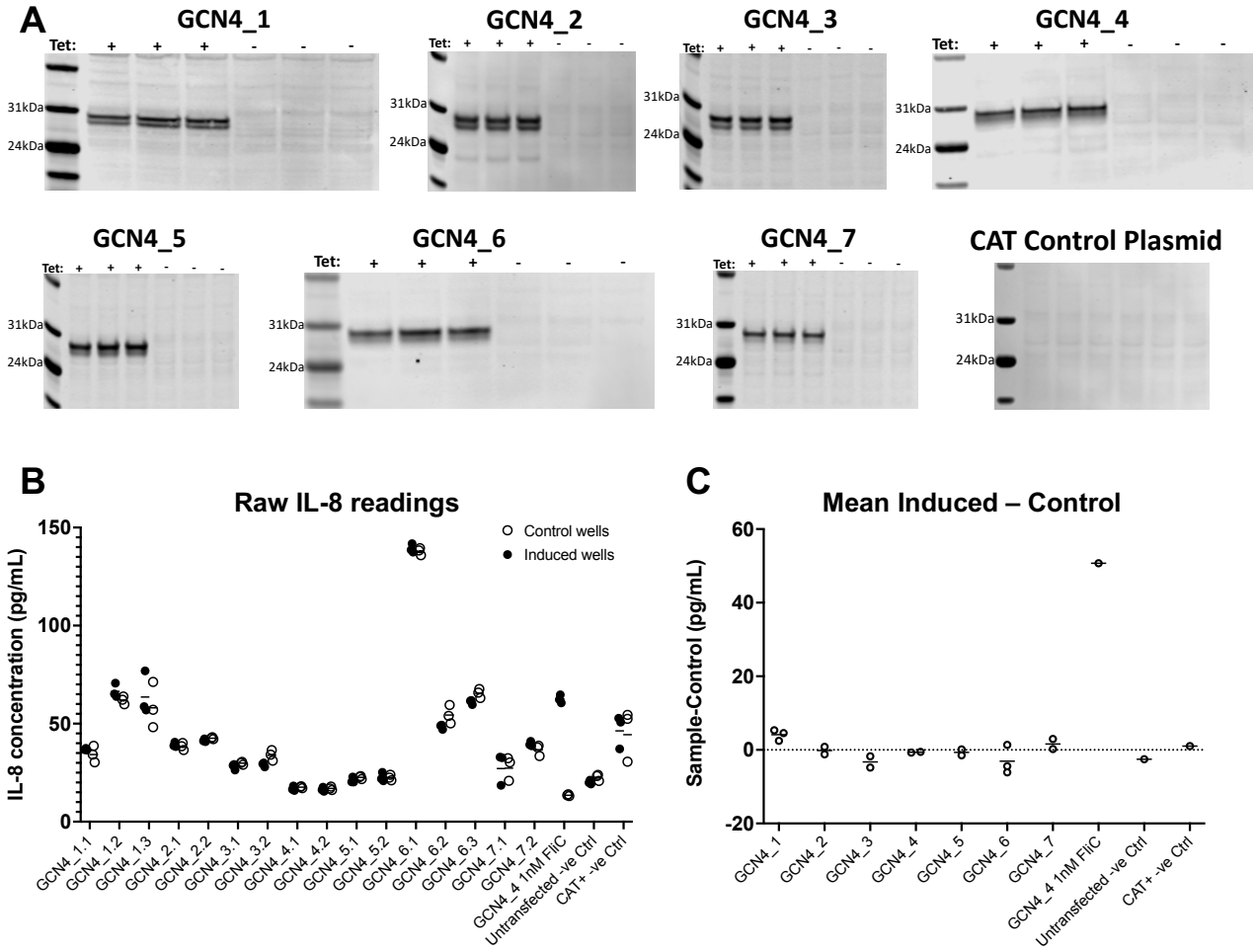


Figure 4.4 – Induction of stably transfected HEK-293 cells and IL-8 response measurements.

(A) Anti-His western blots confirmed tetracycline-induced expression of each GCN4-TIR construct. 1 µL of Tet was added to each induced well. Double bands were observed at the expected size for each construct. No bands were present in the CAT+ control transfection. (B) Raw results from IL-8 ELISA on stably transfected cells. Each dot represents a technical replicate, and the X axis identifies biological replicates. Induced wells were analysed with paired uninduced wells in the same plate. Positive (*St*FliC) and Negative (Untransfected & CAT+) controls are displayed for reference. Horizontal lines represent the mean. (C) The mean IL-8 response induced by GCN4-TIR expression was normalised to the baseline (Uninduced IL-8 signal) for each biological replicate, then plotted for each construct. The horizontal line represents the mean.

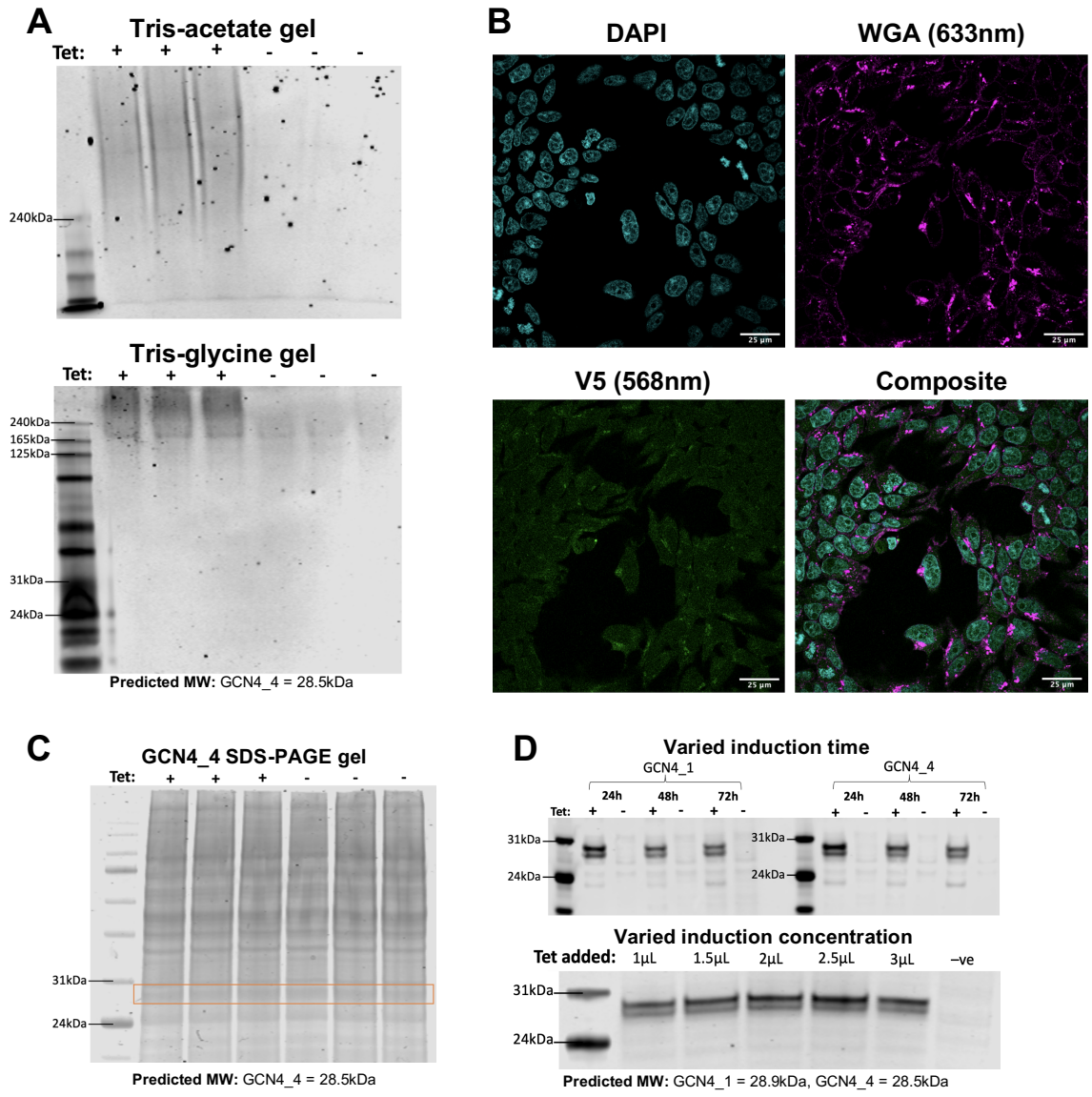


Figure 4.5 – Troubleshooting and optimisation of stably transfected HEK cells. (A) Anti-His western blots of induced and uninduced GCN4_4 cell-lysate native stains were performed with Tris-acetate and Tris-glycine gels to visualise separate bands. (B) Representative immunofluorescence microscopy images of HEK-293 cells stably transfected with GCN4_4 and induced with tetracycline. Stains targeted the nucleus (DAPI), plasma membrane (WGA), and GCN4_4 (V5). (C) SDS-PAGE gel on raw cell lysates from induced and uninduced GCN4_4 stable lines. The orange box indicates the expected size of GCN4-TIR constructs. (D) Anti-His western blots comparing expression levels at different induction times (Top) and tetracycline concentrations (Bottom), performed with GCN4_1 and 4, and GCN4_4 constructs, respectively.

4.3. Discussion

Though ultimately unsuccessful, the design of this study is sound based on existing literature. Using GCN4 coiled-coils to create artificial protein dimers is well established, and practical given the requirements for this study [24–26]. As TLR5 dimerisation is required for the initiation of an immune response, facilitated by the TLR5-TIR, artificial TLR5-TIR dimers should be capable of stimulating the TLR5 immune pathway [7,17,27]. This is expected to occur at the plasma membrane, thus the inclusion of the established PM-targeting N-terminal signalling sequence should have localised the GCN4_TIR constructs to the correct region [3,28]. Thus, the inability to stimulate a substantial TLR5 response in HEK cells is most likely caused by experimental conditions.

Transient transfections displayed clear differences between samples within each separate experiment, but the lack of consistency between experiments meant that these results were not reliable. These inconsistencies were particularly highlighted in the positive control stimulated with *St*FliC, which showed a drastically varied response level when compared to the non-transfected negative control between experiments (Fig. 4.2). The troubleshooting analyses indicated that this inconsistency was likely construct-independent, as transfections with no plasmid also showed cell death. It is surprising that consistency was not improved through adjusting the transfection conditions, and indicates that transient transfections are likely unsuitable for this strategy.

The change to stable transfections was practical, as stable cell-lines would not be susceptible to the transfection-related cell death observed in transient transfections, and this greatly improved the consistency of my results. Tet-induced expression further supported my conclusion that transient transfections were causal to the aforementioned inconsistency, as no qualitative differences were apparent between induced and uninduced cells.

Despite this, no GCN4-TIR construct was able to initiate a substantial TLR5 immune response. Stable lines expressing GCN4_1 and GCN4_7 both stimulated IL-8 activity over multiple replicates, though to a very limited extent (Fig. 4.4C). This could mean that they are not at the right orientation, but are close enough to allow for limited interactions with MyD88 and a partial response. This would indicate that TIR-dependent pathway activation is highly sensitive, with only a narrow window of orientations capable of readily binding downstream effectors. This

specificity would help to prevent stimulation by non-liganded TLR dimers, thus limiting non-targeted TLR5 activation [17]. Future experiments could examine finer rotational conformations between those of GCN4_1 and 7, however this would require the use of different coiled-coil constructs, and was not feasible within the timeframe of this study.

The numerous troubleshooting experiments have helped to identify areas for improvement. There is a limited understanding on the level of TLR5 expression which is required to initiate a response [5]. Consequently, the protein expression level may a potential factor impacting the magnitude of response, despite the strong western blot bands observed (Fig. 4.4A). I was unable to substantially improve protein expression, but this would be a clear area for optimisation in future. The double-banding observed in western blots indicates another potential issue, as the marginal size difference implies that there is N-terminal cleavage occurring in a substantial proportion of the samples (Fig. 4.4A). Such cleavage would separate the PM targeting sequence from the GCN4-TIR constructs, preventing translocation, and could explain the distributed V5 signal (Figure 4.5B). Consequently, redesigning the construct to minimise the potential for cleavage should be considered.

The addition of a longer linker to the GCN4_1 would allow for more rotation independent dimerisation, giving the TLR5-TIR freedom to associate in several orientations. This would also facilitate interactions between asymmetric TIRs, which was not possible in this study, and has some grounding in the existing literature [17]. If such a dimer is able to stimulate a TLR5 response, coiled-coil dimers with a wider range of rotamers should be designed, initially focused around the expected conformation of GCN4_1.

4.4. Conclusion

This study was ultimately unable to clarify how the TLR5-TIR interacts as a dimer to stimulate the TLR5 signalling pathway. However, my results can facilitate the development of improved methodologies that can be applied in future research. If TIR dimerisation is found to be orientation-dependent, it would support our hypothesis of flagellin-induced conformational change, and identify a mechanism by which silent flagellins may evade immune responses [7]. Determining the structure of a stimulatory TLR5-TIR domain dimer would enable detailed investigation into the interactions between TLR5-TIRs and MyD88, and subsequent Myddosome formation. These insights would enhance our wider understanding of TLR signalling initiation, and inform the development of methods to manipulate these medically important immune response pathways.

4.5. Methods

4.5.1. Design and production of GCN4-TIR constructs

A consensus TLR5 sequence (UniProt ID #O60602) was manually annotated to identify the core domains. The identified TLR5-TIR was added to the C-terminus of a GCN4 coiled coil sequence from existing literature, with a short linker, and rotational variants generated by removing single amino acids from the C-terminus of the GCN4 sequence (Table 4.1) [21]. A 10-amino acid plasma-membrane trafficking sequence from the Rous sarcoma virus was added to the N-terminus [23]. Sequences were codon optimised for *Homo sapiens*, and synthesised by IDT.

For transient transfections, GCN4-TIR sequences were cloned into pEF6/V5-His-TOPO™ plasmids (Invitrogen #K961020), which added C-terminal V5 and 6xHis tags, using NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs (NEB) #M5520). For stable transfections, the GCN4-TIR sequence with V5 and 6xHis tags in pEF6 plasmids were amplified by PCR, and cloned into pcDNA™/FRT/TO plasmids (Invitrogen # V652020) using NEBuilder HiFi DNA Assembly Master Mix (NEB #M5520). The assembly was verified by sanger sequencing, then transformed into NEB 5-alpha *E. coli* (NEB #C2987). Transformed cells were cultured in 5 ml of LB with 100 µg/ml carbenicillin (Invitrogen), and purified using a ZymoPURE plasmid miniprep kit (Zymo Research # D4212).

4.5.2. Cell culturing

HEK-Blue hTLR5 reporter cells (InvivoGen #hkb-htlr5) and HEK-293 reporter cells were cultured in 5% CO_2 at 37°C in Growth media (DMEM+GlutaMAX (Gibco #31966021) with 10% FBS (Gibco #26140079) and 100 µg/ml Zeocin (InvivoGen)). The media for HEK-Blue cells was further supplemented with 30 µg/ml Blasticidin (InvivoGen). At 90% confluence, cells were detached using PBS (37°C, pH 7.4) (Gibco), diluted to 50% confluence in growth medium, and transferred to either 24-well or 6-well cell culture plates (Cellstar).

4.5.3. Flagellin Protein Production and purification

Myc-tagged *St*FliC proteins were designed and cloned as described in Chapter 2.4.4. Flagellin expression, and purification of the flagellins was performed with BL21(DE3) *E. coli* as described in Chapter 3.5.1.

4.5.4. Transfection of HEK-Blue and HEK-293 TLR5 reporter cells

Transient and stable transfections were performed on HEK-Blue and HEK-293 TLR5 reporter cells respectively, at 70%-90% confluence. Transfections were performed in either 24-well or 6-well plates using the Lipofectamine™ 3000 Transfection kit (Thermo Fischer Scientific #L3000), following the manufacturer's protocol. Selection for stable lines was started after 2 days through the addition of 200 µg/ml Hygromycin B Gold (Invivogen #ant-hg). Colonies which persisted were expanded, and tested for sensitivity to Zeocin. Cells which were resistant to Blasticidin and Hygromycin but sensitive to Zeocin were identified as stably transfected.

4.5.5. Quanti-Blue assay

Quanti-Blue assays measured Secreted Embryonic Alkaline Phosphatase (SEAP) production in transfected HEK-Blue cells, which are a proxy for TLR5 pathway activation. Each plate contained three biological replicates per sample, including the transfection control. 2 days post-transfection, 20 µL of medium was transferred to a 96-well plate, with 3 technical replicates per sample. 180 µL of QUANTI-Blue solution (InvivoGen #rep-qbs) was added to each well, then incubated for 30 minutes at room temperature. The absorbance at 635 nm was measured after 30 min with a Synergy H1 microplate reader (BioTek). The absorbance values for the three technical replicates were averaged to find the absorbance for each biological replicate, which were plotted in Figure 4.2.

4.5.6. IL-8 Enzyme Linked Immunosorbent Assay (ELISA)

IL-8 ELISAs were used to measure TLR5 pathway activation in stably transfected HEK-293 cells. The production of GCN4-TIR constructs was induced with 1 µg/ml tetracycline (Sigma-Aldrich #T3258), and the positive control stimulated with 1 µg/ml *St*FliC, 24 hours prior to analysis. Paired uninduced wells were analysed alongside each sample, to determine baseline IL-8 expression. Analysis was performed with the Human IL-8 ELISA Kit (Invitrogen # KHC0081), following the manufacturer's protocol. Three technical replicates were measured for each sample, and the mean was plotted for each biological replicate in Figure 4.4B. To determine the impact of GCN4-TIR construct expression on activity, the expression in induced cells was normalised against the baseline expression in non-induced cells for each replicate, plotted in Figure 4.4C.

4.5.7. HEK Cell Lysis

Both Stable and Transiently transfected HEK cells were lysed in the same manner. Remaining media from sample wells was removed and discarded. Cells were washed once in 500 µL PBS (37°C, pH 7.4), and then incubated in 500 µL PBS (37°C, pH 7.4) at 37°C for 10 minutes to detach cells. Cells were resuspended by pipetting, transferred to 1.5mL Eppendorf tubes, then pelleted by tabletop centrifugation (20,000 rcf, 1 min). The PBS was removed, the pellet resuspended in 100 µL HEK lysis buffer (1% NP-40, 100 mM NaCl, 25 mM TRIS, 20 mM Imidazole, 10% Glycerol, pH 7.5), and then incubated on ice for 15 minutes. Cellular debris was pelleted by tabletop centrifugation (10000 rcf, 4°C, 20 min), and the supernatant containing the HEK cell lysate extracted. The protein concentration of HEK cell lysates was quantified with a Pierce™ BCA Protein Assay Kit (Thermo Scientific #23225), following the manufacturer's protocol.

4.5.8. Protein Blotting

20 µg of HEK cell lysate was used for SDS-PAGE and western blot analyses. Protein samples were stained with NuPAGE LDS sample buffer (Novex #NP0007) and boiled at 95°C for 3 minutes. Samples and Chameleon Duo Pre-Stained Protein Ladder (LI-COR #928-60000) were run on Bolt 4-12% NuPAGE Bis-Tris WedgeWell gels (Invitrogen #NW04120) in a Mini Gel Tank (Invitrogen #A25977) with MES buffer (50 mM MES, 50 mM Tris (pH 7.3), 0.1% (w/v) SDS, and 1 mM EDTA) at 100 V.

40 µg of HEK cell lysate was used for Native-PAGE analysis. Protein samples were stained with Tris-Glycine Native Sample Buffer (Novex #LC2673). Samples and a protein ladder were run on NativePAGE 4-16% Bis-Tris gels (Invitrogen #BN1002) or NuPAGE 3-8% Tris-Acetate gels (Invitrogen #EA0375) in a Mini Gel Tank with either NativePAGE Running Buffer (Invitrogen #BN2001) or Tris-Glycine Native Running Buffer (Invitrogen #LC2672) at 100 V.

For SDS-PAGE, gels were directly stained with ReadyBlue Protein Gel Stain (Sigma-Aldrich #RSB) for 1 hour, then washed twice for 30 minutes in water. Gels were imaged with an Odyssey CLx (LI-COR). For western Blots, the protein was transferred to a nitrocellulose membrane (Thermo Fischer Scientific #88018) in transfer buffer (NuPAGE Transfer Buffer (Invitrogen #NP00061), 10% (v/v) methanol) at 4°C for 70 min at 10 V, using the Mini Blot

Module (Invitrogen #B1000). Membranes were blocked for 1 hour at room temperature with blocking buffer (TBS, 0.1% (v/v) Tween-20, 5% (w/v) milk), then stained with a primary antibody for 1 hour at room temperature. Membranes were washed thrice for 5 min with TBS-T and incubated with a secondary antibody for 30 min at room temperature. Membranes were washed twice in TBS-T, then imaged with an Odyssey CLx (LI-COR).

Primary antibodies targeting His tags (TetraHis (QIAGEN #34670)) and V5 tags (Anti-V5 Antibody (Novex #R96025)) were diluted 1:1000 (v/v) and 1:2000 (v/v) respectively in blocking buffer. Secondary antibodies (IRDye 660RD/IRDye 880CW (LI-COR #926-32210/#926-68070)) were diluted 1:10000 (v/v) in TBS-T. Western blot band intensities were quantified with Image Studio (LI-COR).

4.5.9. Immunofluorescence Microscopy

HEK-cells used in immunofluorescence microscopy were grown on Poly-D-Lysine Cellware 12mm round Coverslips (Corning #354086) in 6-well plates. A non-transfected/uninduced control was imaged alongside each sample to check for non-target signal. Non-stained cells were also imaged to check for potential background at the measured wavelengths. Transient transfection or Tet induction were performed at 70% confluence as previously described. 24 hours post-transfection or post-induction, cells were washed twice with PBS (37°C, pH 7.4), then fixed in 1.5 mL PBS (37°C, pH 7.4) with 2% Paraformaldehyde (Sigma-Aldrich #16005) for 30 minutes at room temperature. Cells were washed twice in PBS (37°C, pH 7.4), then fixation was quenched using 1.5 mL PBS (37°C, pH 7.4) with 100 mM glycine in PBS for 1 hour at room temperature. Cells were washed twice with PBS (37°C, pH 7.4), then stained.

Staining was performed in darkness. Cells on the coverslip were incubated in Wheat germ agglutinin (WGA) Alexa Fluor 633 conjugate (Invitrogen #W21404), pre-diluted to 5 µg/mL in PBS (pH 7.4), for 2 minutes at room temperature, then washed twice in PBS (37°C, pH 7.4). Cells were permeabilized using Permeabilization buffer (PBS (pH 7.5), 1% (w/v) BSA (Sigma-Alrich #A2153), 0.1% (w/v) Saponin (Sigma-Aldrich #47036)) for 1 hour at room temperature. The buffer was removed, and the coverslip incubated in Primary antibody mix (PBS (pH 7.4), 1% (w/v) BSA, 0.1% (w/v) Saponin, 1:100 Anti-V5 Antibody) for 1 hour at room temperature, or overnight at 4°C. The coverslip was washed for 5 minutes in PBS (37°C, pH 7.4) thrice, then incubated in Secondary antibody mix (PBS (pH 7.4), 1% (w/v) BSA, 0.1% (w/v) Saponin, 1:5000

IRDye 880CW) for 1 hour at room temperature. The coverslip was washed for 5 minutes in PBS (37°C, pH 7.4) thrice, then stained using 1 µg/mL DAPI in PBS for 3 minutes at room temperature. Coverslips were finally washed twice with PBS (37°C, pH 7.4) for 5 minutes, dried, and mounted on microscope slides overnight at 4°C with ProLong Diamond Antifade Mountant (Invitrogen #P3696). The slides were imaged using a Leica TCS SC8 confocal microscope (Leica).

4.5.10. Data processing

SDS-PAGE and western blot images were processed with Image Studio (LI-COR). Graphs were made with Prism 10 (GraphPad). Structural models and complexes were generated with AlphaFold2 [29,30]. Immunofluorescence microscopy images were processed with FIJI (ImageJ).

4.6. References

1. Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, Goodlett DR, et al. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature. 2001;410. doi:10.1038/35074106

2. Medzhitov R. Toll-like receptors and innate immunity. Nat Rev Immunol. 2001;1: 135–145.

3. Takeda K, Akira S. Toll-like receptors in innate immunity. Int Immunol. 2005;17: 1–14.

4. Reichhart J-M. TLR5 takes aim at bacterial propeller. Nat Immunol. 2003;4: 1159–1160.

5. Feng T, Cong Y, Alexander K, Elson CO. Regulation of Toll-like Receptor 5 Gene Expression and Function on Mucosal Dendritic Cells. PLoS One. 2012;7: e35918.

6. Lu Y, Li X, Liu S, Zhang Y, Zhang D. Toll-like Receptors and Inflammatory Bowel Disease. Front Immunol. 2018;9. doi:10.3389/fimmu.2018.00072

7. Clasen SJ, Bell MEW, Borbón A, Lee D-H, Henseler ZM, de la Cuesta-Zuluaga J, et al. Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. Sci Immunol. 2023;8: eabq7001.

8. Song DH, Lee J-O. Sensing of microbial molecular patterns by Toll-like receptors. Immunol Rev. 2012;250: 216–229.

9. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat Immunol. 2010;11: 373–384.

10. Akira S, Takeda K. Toll-like receptor signalling. Nat Rev Immunol. 2004;4: 499–511.

11. Hayden MS, Ghosh S. Signaling to NF-κB. Genes Dev. 2004;18: 2195–2224.

12. Balka KR, De Nardo D. Understanding early TLR signaling through the Myddosome. J Leukoc Biol. 2019;105: 339–351.

13. Tallant T, Deb A, Kar N, Lupica J, de Veer MJ, DiDonato JA. Flagellin acting via TLR5 is the major activator of key signaling pathways leading to NF-κB and proinflammatory gene program activation in intestinal epithelial cells. BMC Microbiol. 2004;4: 1–24.

14. Ve T, Williams SJ, Kobe B. Structure and function of Toll/interleukin-1 receptor/resistance protein (TIR) domains. Apoptosis. 2014;20: 250–261.

15. Javmen A, Zou J, Nallar SC, Szmacinski H, Lakowicz JR, Gewirtz AT, et al. TLR5-Derived, TIR-Interacting Decoy Peptides to Inhibit TLR Signaling. J Immunol. 2023;210: 1419–1427.

16. Ivison SM, Kahn MAS, Graham, NR, Berales CQ, Kaleem A, Tirling CO, Cherkasov A, Steiner TS. A phosphorylation site in the Toll-like receptor 5 TIR domain is required for inflammatory signalling in response to flagellin. Biochem Biophys Res Commun. 2007;352: 936–941.

17. Zhou K, Kanai R, Lee P, Wang H-W, Modis Y. Toll-like receptor 5 forms asymmetric dimers in the absence of flagellin. J Struct Biol. 2012;177: 402–409.

18. O'Neill LAJ, Bowie AG. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. Nat Rev Immunol. 2007;7: 353–364.

19. Dunne A, Ejdeback M, Ludidi PL, O'Neill LAJ, Gay NJ. Structural complementarity of Toll/interleukin-1 receptor domains in Toll-like receptors and the adaptors Mal and MyD88. J Biol Chem. 2003;278: 41443–41451.

20. Nyman T, Stenmark P, Flodin S, Johansson I, Hammarström M, Nordlund P. The crystal structure of the human toll-like receptor 10 cytoplasmic domain reveals a putative signaling dimer. J Biol Chem. 2008;283: 11861–11865.

21. O'Shea EK, Klemm JD, Kim PS, Alber T. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. Science. 1991;254: 539–544.

22. Woolfson DN. The design of coiled-coil structures and assemblies. Adv Protein Chem. 2005;70. doi:10.1016/S0065-3233(05)70004-8

23. Pellman D, Garber EA, Cross FR, Hanafusa H. An N-terminal peptide from p60src can direct myristylation and plasma membrane localization when fused to heterologous proteins. Nature. 1985;314: 374–377.

24. Sarson-Lawrence KTG, Hardy JM, Iaria J, Stockwell D, Behrens K, Saiyed T, et al. Cryo-EM structure of the extracellular domain of murine Thrombopoietin Receptor in complex with Thrombopoietin. Nat Commun. 2024;15. doi:10.1038/s41467-024-45356-2

25. Chrudinová M, Kirk NS, Chuard A, Venugopal H, Zhang F, Lubos M, et al. A viral insulin-like peptide inhibits IGF-1 receptor phosphorylation and regulates IGF1R gene expression. Molecular Metabolism. 2024;80. doi:10.1016/j.molmet.2023.101863

26. Baltrusaitis EE, Ravitch EE, Fenton AR, Perez TA, Holzbaur ELF, Dominguez R. Interaction between the mitochondrial adaptor MIRO and the motor adaptor TRAK. J Biol Chem. 2023;299. doi:10.1016/j.jbc.2023.105441

27. Ohto U, Ishida H, Shibata T, Sato R, Miyake K, Shimizu T. Toll-like Receptor 9 Contains Two DNA Binding Sites that Function Cooperatively to Promote Receptor Dimerization and Activation. Immunity. 2018;48. doi:10.1016/j.immuni.2018.03.013

28. Kumar H, Kawai T, Akira S. Toll-like receptors and innate immunity. Biochem Biophys Res Commun. 2009;388: 621–625.

29. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.

30. Evans R, O'Neill M, Pritzel A, Antropova N, Senior A, Green T, et al. Protein complex prediction with AlphaFold-Multimer. bioRxiv. 2022. p. 2021.10.04.463034. doi:10.1101/2021.10.04.463034

Chapter 5: Towards a structure-based understanding of human TLR5 and the TLR5-flagellin complex

5.1. Introduction

Toll-like receptor 5 (TLR5) is a key component of the human innate immune system, forming part of the Toll-like receptor family of Type-1 transmembrane pattern recognition receptors (PRRs). TLR5 specifically binds bacterial flagellin, the protein monomer of the bacterial flagellum [1,2]. Flagellin binding occurs at the TLR5 extracellular domain (ECD), which is a 'C' shaped structure composed of 24 leucine-rich repeats (LRRs). This domain is highly conserved throughout PRRs, and contains the MAMP-binding interfaces [3–5]. Following flagellin binding, TLR5 is believed to undergo a conformational change, triggering the NF-κB signalling pathway and resulting in a cytokine response [5,6].

To understand the mechanisms behind TLR5 signalling, a detailed understanding of TLR5-flagellin binding is needed. Mutational studies have identified residues in both flagellin and TLR5 integral to facilitating an immune response, largely focusing on the canonical TLR5 agonist *Salmonella typhimurium* FliC (*St*FliC) [6–9]. Structural studies have supported these findings, providing molecular-level details of interactions key to TLR5-flagellin complex formation [10,11]. The most informative structure comes from X-ray crystallography on zebrafish TLR5b (zTLR5b) in complex with a truncated FliC from *Salmonella enterica* subspecies *enterica* serovar Dublin (*Sd*FliC). This structure identified direct interactions between TLR5 and flagellin at two binding interfaces, supporting findings from mutagenic studies, and found evidence for a conformational change in the TLR5 LRR-9 loop following flagellin binding [10].

However, the applicability of this zTLR5b:*Sd*FliC structure to a human context is limited [10]. Generating a crystal of this complex required the truncation of both TLR5 and *Sd*FliC. Consequently, the structure provides no information on flagellin D0-domain interactions and subsequent conformational changes in TLR5, which are thought to have a significant role in initiating TLR5 signalling [6,8]. The use of zTLR5b instead of a human TLR5 construct also raises the possibility of species-specific differences in TLR5, as variations have been described in TLR5 expression, function, and the response to individual flagellins [12–14].

To build a comprehensive understanding of TLR5-flagellin binding in a human context, the structure of human TLR5 and a TLR5-flagellin complex is essential. Currently, only one experimental human TLR5 structure exists, but it has a very low (26Å) resolution, limiting its usefulness for interpreting direct interactions [5]. No structures have yet resolved the TLR5 transmembrane (TM) and cytoplasmic Toll/Interleukin-1 Receptor (TIR) domains, which are crucial for downstream signal transduction through the recruitment of MyD88 [15,16]. While structural modelling and superposition complexes, as used in Chapter 3, can partially bridge this gap, they cannot identify specific interactions essential for TLR5-flagellin binding and signalling.

In this study, I aim to fill these knowledge gaps by determining the structure of full-length human TLR5 (TLR5-FL) and the complete TLR5 ectodomain (TLR5-ECD), both as monomers and in complex with *St*FliC, using Cryo-electron microscopy (Cryo-EM). By obtaining the structure of TLR5 in both monomer and complex forms, I aim to identify ligand-induced conformational changes caused by flagellin binding and receptor dimerisation. With Cryo-EM, I can image unmodified flagellin and TLR5 constructs, which would otherwise require modification for X-ray crystallography. Cryo-EM may also allow the visualisation of different stages of TLR5-flagellin heterodimer formation, which would help clarify mechanisms involved in this process.

I successfully produced and purified both TLR5-FL and TLR5-ECD proteins at high concentrations, and confirmed their ability to bind *St*FliC. Negative stain and Cryo-EM imaging of the TLR5-ECD was performed by my collaborators in the Möller lab, at the Universität Osnabrück, and resolved 2D classes displaying the expected ‘C’ conformation. However, these were of insufficient clarity to resolve a high-resolution structure. I further attempted several methods to generate TLR5-ECD:*St*FliC complexes. While none resolved clear complexes in 2D class averages, promising directions for continued research have been identified.

Overall, I made substantial progress toward resolving an experimental structure for human TLR5 and the TLR5-flagellin complex, and identified clear avenues for achieving this in the near future. A comprehensive structural understanding of human TLR5-flagellin interactions would enhance our understanding of host-pathogen interactions, particularly regarding how pathogens manipulate TLR5 signalling to facilitate stimulatory, silent, and evasive immune responses. Such insights would be invaluable in the design of novel therapeutics aimed at modulating TLR5 activity, benefiting those suffering from conditions related to TLR5 dysregulation. Lastly, clarifying the exact mechanisms of TLR5 signal transduction would allow for a broader understanding of TLR signalling pathways, contributing more generally to our understanding of innate immunity and its relationship to with the gut microbiome.

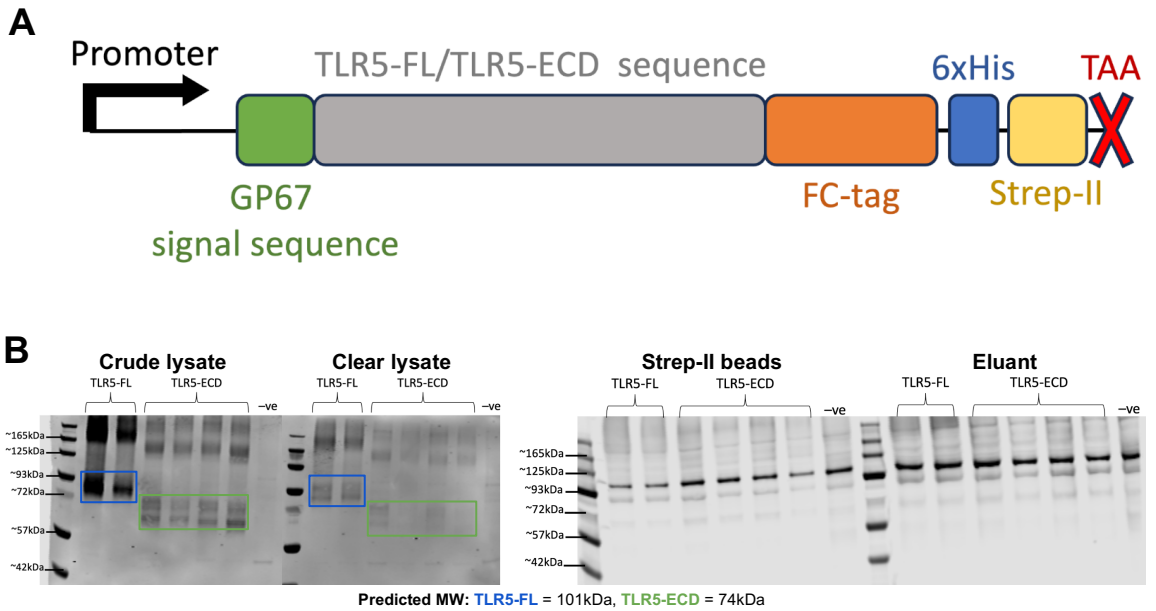


Figure 5.1 – Transfection and purification of TLR5-FL and TLR5-ECD constructs in SF9 insect cells.

(A) Sequence map annotated with the key features of TLR5-FL and TLR5-ECD constructs. (B) Anti-His western blots from transfected SF9 V0 cultures. Different stages of the purification are displayed, with bands corresponding to TLR5-FL and TLR5-ECD highlighted by blue and green boxes respectively. Negative controls are from non-transfected SF9 cells. (C) SDS-PAGE gels from a His column purification (Left) and SEC on the eluent fraction (Right). Bands corresponding to TLR5-FL are highlighted.

5.2. Results

5.2.1. Expression and purification of recombinant TLR5 proteins

Full-length TLR5 (TLR5-FL) and TLR5 ectodomain (TLR5-ED) sequences were codon optimised and produced by IDT. TLR5 constructs were cloned into pLIB plasmids, with an N-terminal GP67 secretion signal sequence and a C-terminal 6xHis tag (Fig. 5.1A) [17,18]. Baculoviruses were generated following the established biGBac system, with SF9 transfection confirmed by western blot (Fig. 5.1B) [19]. Optimal expression conditions of 1:1000 (v/v) inoculation with baculovirus and 72 hour expression time were identified.

Purification of the recombinant TLR5 constructs from the pellet targeted the 6xHis tag. Clear bands were identified in cell lysates, but not in beads or the eluant (Fig. 5.1B). Purification on a cell pellet from a 50 mL culture showed substantial non-target protein bands in SDS-PAGE analyses, which could not be separated from TLR5-FL or TLR5-ECD by size exclusion chromatography (SEC) (Fig. 5.1C).

Consequently, TLR5 was instead purified from the cell supernatant, where it had been secreted due to the GP67 secretion signal. Clear bands corresponding to TLR5-FL and TLR5-ECD were observed following His-affinity purification (Fig. 5.2A). SEC purifications using the HiLoad® 16/600 Superdex® 200 column (Cytiva) provided the best separation, but had a substantial dilution effect which limited the final concentration. Attempts were made to concentrate the sample, but the SEC-purified TLR5 constructs rapidly precipitated each time. A final strategy was established, relying on a high SEC input concentration (1 mL protein at ~5 mg/mL) and a Superdex® 200 10/300 GL column, which provided sufficient separation and a reduced dilution effect (Fig. 5.2B). Through this method, approximately 1 mg/mL and 0.7 mg/mL of SEC purified protein were obtained for TLR5-ECD and TLR5-FL respectively.

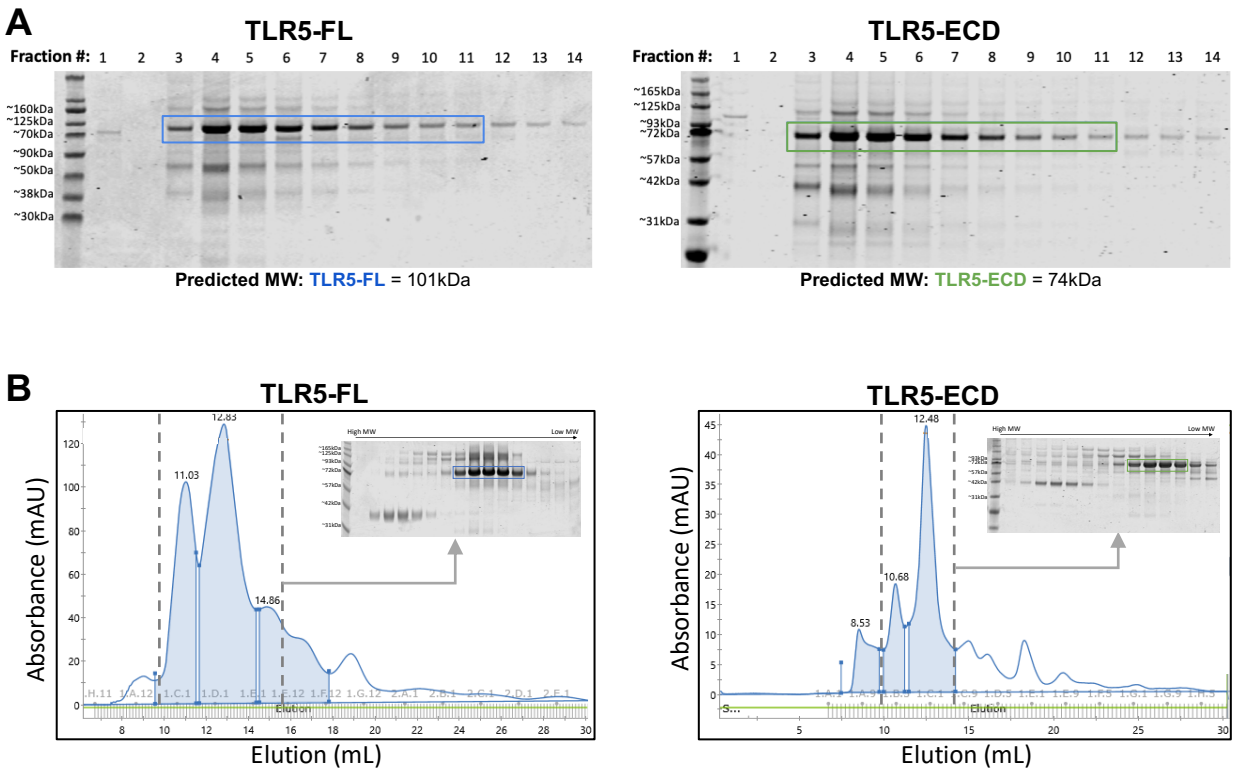


Figure 5.2 – Representative His-affinity and SEC purifications of TLR5-FL and TLR5-ECD constructs from Hi-5 insect cells.

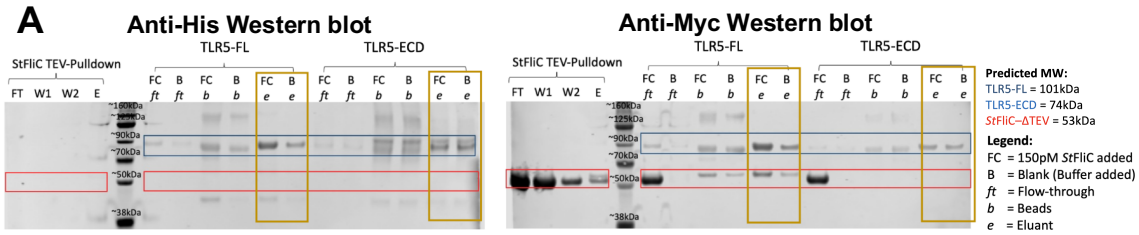
(A) SDS-PAGE blots of fractions from TLR5-FL (Left) and TLR5-ECD (Right) Äkta Start Ni-affinity purifications. Bands corresponding to the TLR5-FL and TLR5-ECD samples pooled and concentrated for SEC purification are highlighted. (B) Representative SEC chromatograms from TLR5-FL (Left) and TLR5-ECD (Right). The range of fractions analysed in the inset SDS-PAGE blots are indicated by the dotted lines. Fractions containing TLR5-FL and TLR5-ECD used for structural analyses are highlighted.

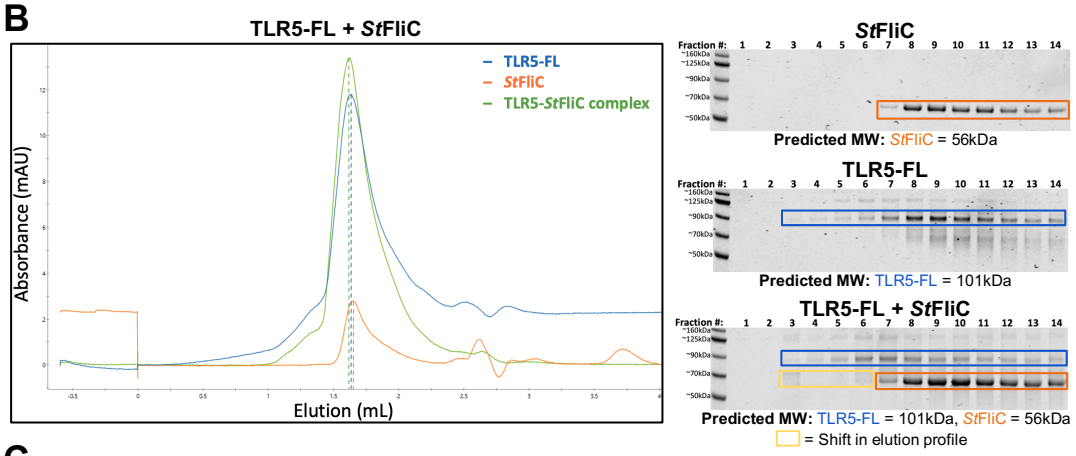
5.2.2. Functional validation of recombinant TLR5 proteins

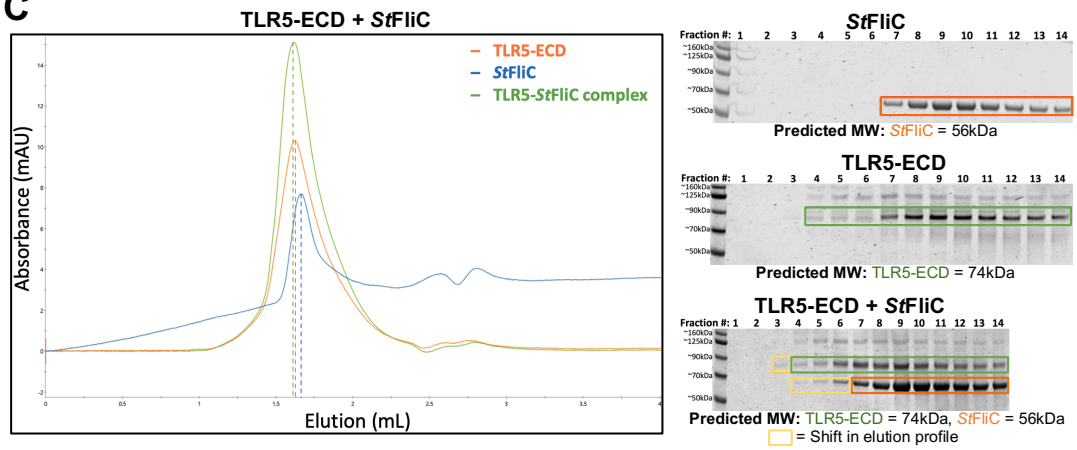
To confirm that the recombinant TLR5 proteins could interact with flagellins, I performed several functionality assays. The His tag was cleaved from the recombinant *St*FliC protein using a TEV protease site, and confirmed by Anti-His and Anti-Myc western blots (Fig. 5.3A). Clear bands corresponding to *St*FliC were observed in pulldowns with TLR5-FL, but not TLR5-ECD, implying the former displayed a stronger binding ability.

Analytical SECs performed on both recombinant TLR5 proteins indicated both TLR5 constructs interacted with *St*FliC. Samples with TLR5 and *St*FliC constructs co-incubated showed a clear shift in the elution curve compared to single-protein reference peaks (Fig. 5.3B-C). SDS-PAGE analysis of TLR5-FL binding curves showed clear shifts in *St*FliC elution from fraction 7 to 3 (Fig. 5.3B). A smaller shift was also observed in the TLR5-ECD from fractions 7 to 4 (Fig. 5.3C). In both cases, a reduced salt buffer, containing 150 mM NaCl instead of 300 mM, was required to create a substantial shift, though the effect remained most apparent in TLR5-FL.

Finally, a competition assay performed by Dr. Sara Clasen showed a reduction in TLR5 signalling activity corresponding to the concentration of both TLR5-FL and TLR5-ECD. The TLR5-ECD showed a marginally higher competitive ability, though significance could not be determined due to the low sample number (Fig. 5.3D). Alongside prior results, I concluded that that both TLR5-FL and TLR5-ECD were able to bind *St*FliC, and should be capable of forming TLR5:*St*FliC complexes.


A
Anti-His Western blot
Anti-Myc Western blot
TLR5-FL
TLR5-ECD
StFliC TEV-Pulldown
FT W1 W2 E
FC B FC B FC B FC B FC B FC B
ft ft b b e e
~160kDa
~125kDa
~90kDa
~70kDa
~50kDa
~38kDa
Predicted MW:
TLR5-FL = 101kDa
TLR5-ECD = 74kDa
StFliC–ΔTEV = 53kDa
Legend:
FC = 150pM StFliC added
B = Blank (Buffer added)
ft = Flow-through
b = Beads
e = Eluant


B
TLR5-FL + StFliC
TLR5-FL
StFliC
TLR5-StFliC complex
Absorbance (mAU)
Elution (mL)
StFliC
Fraction #: 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Predicted MW: StFliC = 56kDa
TLR5-FL
Predicted MW: TLR5-FL = 101kDa
TLR5-FL + StFliC
Predicted MW: TLR5-FL = 101kDa, StFliC = 56kDa
= Shift in elution profile


C
TLR5-ECD + StFliC
TLR5-ECD
StFliC
TLR5-StFliC complex
Absorbance (mAU)
Elution (mL)
StFliC
Fraction #: 1 2 3 4 5 6 7 8 9 10 11 12 13 14
Predicted MW: StFliC = 56kDa
TLR5-ECD
Predicted MW: TLR5-ECD = 74kDa
TLR5-ECD + StFliC
Predicted MW: TLR5-ECD = 74kDa, StFliC = 56kDa
= Shift in elution profile

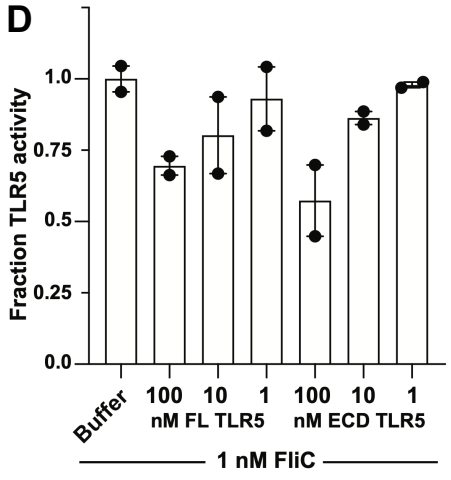

D
Fraction TLR5 activity
1.0
0.75
0.5
0.25
0.0
Buffer
100 10 1 100 10 1
nM FL TLR5
nM ECD TLR5
1 nM FliC

Figure 5.3 – Functional validation assays for TLR5-FL and TLR5-ECD.
(A) Anti-His (Left) and Anti-Myc (Right) western blots of *St*FliC pulldowns using both TLR5 constructs. *St*FliC samples were cleaved with TEV, to remove the His tag. Gold boxes indicate rows expected to contain *St*FliC. (B-C) Analytical SPRs examining shifts in the curve in TLR5+*St*FliC complexes, with separate peaks indicated by dotted lines of the appropriate colour. SDS-PAGE gels on eluted fractions highlight shifts in samples with co-incubated TLR5 and *St*FliC. (D) Results from 2 replicates of a competition assay with TLR5-FL and TLR5-ECD constructs performed in HEK-Blue TLR5 reporter cells. This data was generated by Dr. Sara Clasen.

5.2.3. TLR5-ECD imaging by Negative staining and Cryo-EM

The majority of negative stain and Cryo-EM imaging was performed by my collaborator Jan-Hannes Schäfer, from the Möller lab at the Universität Osnabrück. The recombinant TLR5-ECD construct was used for these experiments, as it was purified at higher concentration. Class averages from preliminary negative stains were inconsistent, but some displayed the expected 'C' shaped conformation (Fig. 5.4A). Larger oligomers were also observed, though none clearly corresponded to a 1:1 TLR5-ECD homodimer complex.

Cryo-EM on the TLR5-ECD also resolved class-averages in the expected 'C' shaped conformation, though substantial noise prevented the refinement of a high-resolution density map (Fig. 5.4B). This noise was thought to be caused by protein aggregation, which was observed in some micrographs. Modifying the buffer to add detergents resulted in a moderate reduction in aggregation, and were applied for subsequent Cryo-EM analyses.

TLR5-ECD and *St*FliC were co-incubated in order to stabilise the flexible TLR5-ECD, and determine if 1:1 TLR5-ECD:*St*FliC complexes could be visualised. Despite trialling several modifications to the buffer and incubation conditions, clear complexes could not be resolved (Fig. 5.4C). However, densities corresponding to un-complexed TLR5 and *St*FliC were visualised at low resolution (Fig. 5.4D).

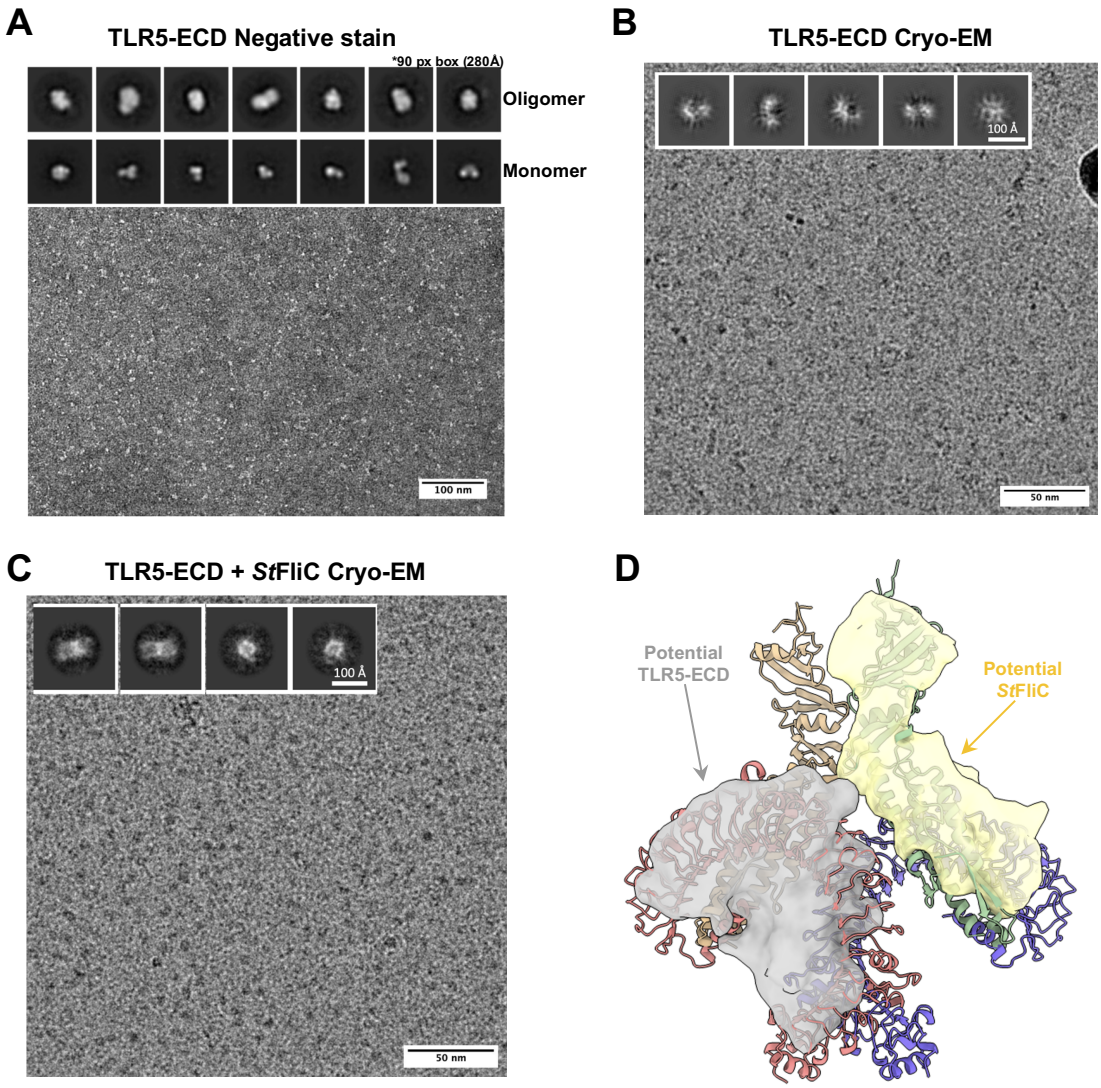


Figure 5.4 – Negative stain and Cryo-EM analysis of samples containing TLR5-ECD and TLR5-ECD + *St*FliC.

All data shown in this figure was generated by the Möller Lab at Universität Osnabrück. (A) Class averages and a representative micrograph from a TLR5-ECD negative stain. Class averages with potential monomers and larger oligomers are shown. (B) Class-averages and a representative micrograph from a preliminary TLR5-ECD Cryo-EM sample. Class averages of the expected 'C' conformation are observed, but display substantial noise. (C) Class-averages and a representative micrograph from a preliminary TLR5-ECD + *St*FliC Cryo-EM sample. Class averages show potential monomers or oligomers, but showed substantial noise. (D) Low resolution density maps extracted from preliminary TLR5-ECD+*St*FliC Cryo-EM data, overlaid on an experimental complex (PDB ID #3v47 from [10]). No clear evidence of densities in complex were identified.

5.2.4. TLR5-ECD + *St*FliC crosslinking and Negative Staining

Crosslinking the TLR5-ECD to *St*FliC was trialled to improve the consistency of complex formation, in a rapid and buffer-independent manner. Optimal crosslinking conditions were identified through an activity screen (Fig. 5.5A). Western blot analyses of SEC purified samples showed bands corresponding predicted monomers and potential TLR5-ECD:*St*FliC complexes. However, as both the TLR5-ECD and *St*FliC constructs contained His tags, the presence of the TLR5-ECD in higher molecular-weight samples could not be confirmed (Fig. 5.5B). Preliminary negative staining was performed on SEC-purified crosslinked samples, and distinct proteins were clearly visible in the micrograph (Fig. 5C). However, I have not yet been able to resolve clear class averages showing monomers or a potential complex.

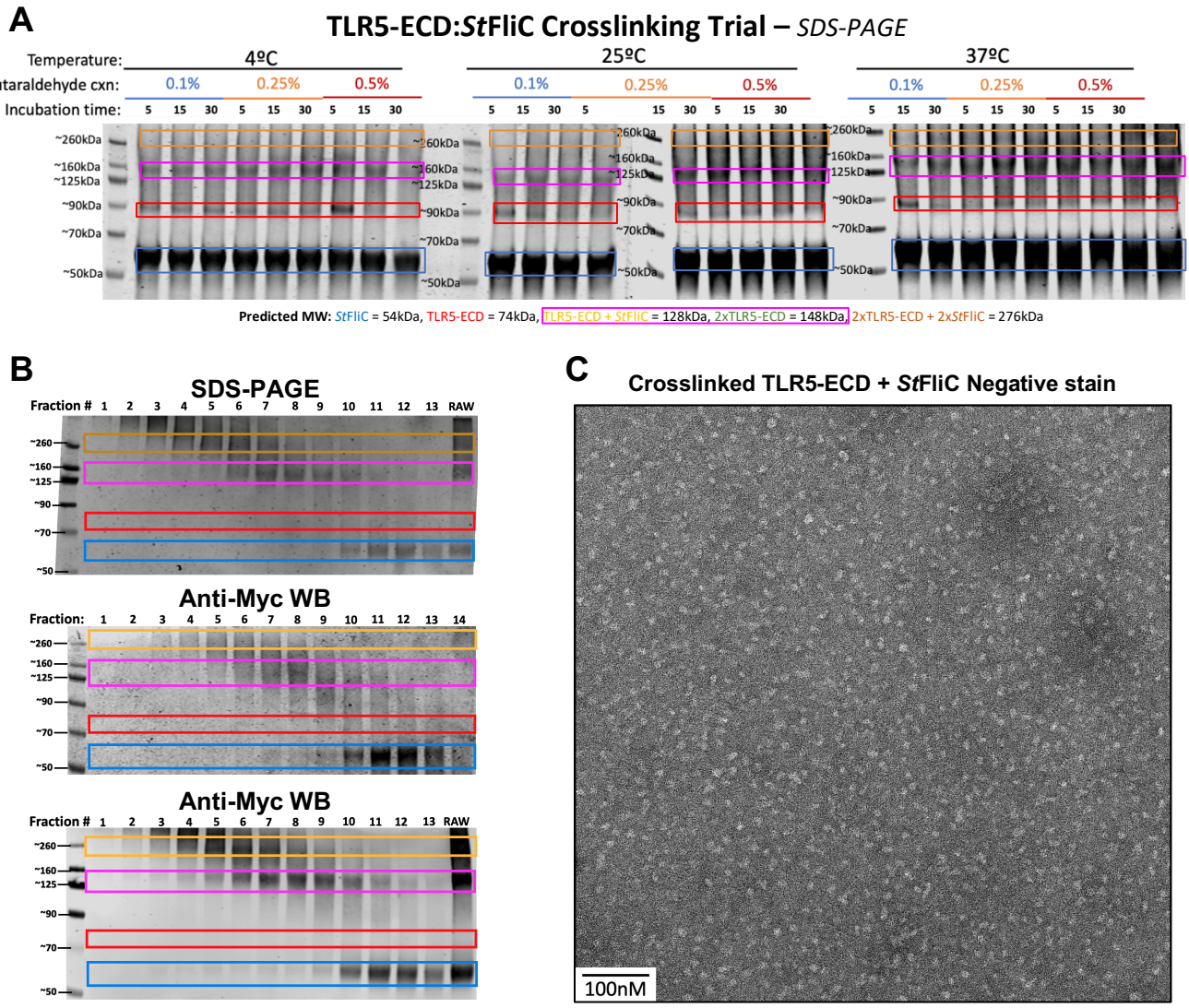


Figure 5.5 – Results from the preliminary TLR5-ECD + *St*FliC glutaraldehyde crosslinking trial.

(A) SDS-PAGE gel displaying the results of a cross-linking screen, used to identify the optimal crosslinking temperature, glutaraldehyde concentration, and incubation time. Bands corresponding to the MW of predicted TLR5-ECD:*St*FliC complexes are indicated. (B) Analytical SDS-PAGE and western blots on SEC fractions of cross-linked TLR5-ECD+*St*FliC. Bands corresponding to predicted complexes are highlighted. (C) Representative negative stain micrograph of the protein sample identified in fractions 8-9 from the SEC-purified cross-linked TLR5-ECD+*St*FliC sample shown in B.

5.3. Discussion

In this study, I aimed to solve the structure of TLR5 and the TLR5:*St*FliC complex, which is of pressing importance as no high-resolution human TLR5 structure has been determined. Our current understanding of TLR5-flagellin interactions relies instead on mutagenic studies and extrapolations from other TLRs, or non-human TLR5, limiting their applicability and creating uncertainty. Solving the structure of human TLR5 would enable detailed studies on its direct interactions with bacterial flagellin, and the identification of potential ligand-induced conformational changes, which were predicted in Chapter 2 [6].

I was able to successfully purify both full-length TLR5 (TLR5-FL) and the complete TLR5 ectodomain (TLR5-ECD) in sufficient concentrations for structural analysis, an achievement not documented in recent literature. However, these constructs were prone to precipitation after SEC purification, unlike the truncated $TLR5^{N14}$ construct used in Chapter 2 and 3 [6]. Both constructs could bind *St*FliC, but TLR5-FL displayed a higher apparent affinity. However, due to lower concentrations of TLR5-FL, the TLR5-ECD was prioritised in structural analyses.

The inability to resolve TLR5-ECD homodimers was not unexpected, given the purported role of flagellins in stabilising TLR5 complexes [5,6,20]. However, identifying 2D classes showing the TLR5-ECD in the expected 'C' conformation in both negative staining and Cryo-EM was very promising, despite the need for improvements to image quality. It should be noted that the size of this recombinant TLR5-ECD protein is at the lower-end of what Cryo-EM is currently able to resolve. This may also be influencing the quality of 2D classes, and is a further limitation of using this construct instead of TLR5-FL [21,22].

Pursuant to these factors, attempting to visualise a complex directly was pragmatic. The flexible TLR5-ECD should be stabilised by *St*FliC binding, and would form oligomers of a higher molecular weight, more suited to Cryo-EM analysis. Complex formation through native co-incubation and purification was unsuccessful, with potential causes identified as the poor binding ability of TLR5-ECD, TLR5 aggregation, and *St*FliC oligomerization. However, glutaraldehyde crosslinking showed initial promise [23]. Clear bands corresponded to several expected variations of TLR5-ECD:*St*FliC complexes, but the exact composition of these complexes could not be determined, and needs to be examined further. Negative stain micrographs of SEC-purified bands corresponding to 1:1 TLR5-ECD:*St*FliC complexes did show

distinct proteins, but I was unable to resolve clear 2D classes. This should be achievable through refining the particle picking and 2D classification parameters, and should be the immediate focus of continued work.

If 2D classes corresponding to the expected TLR5-flagellin complex are identified from these negative stains, Cryo-EM on these crosslinked complexes should be pursued. Otherwise, crosslinking should be avoided to prevent potential artefacts, and the focus should instead shift to improving buffer, co-incubation, and freezing conditions, in order to generate native complexes. Strategies such as co-expression and co-incubation of TLR5 and *St*FliC in insect cells could also be effective [24,25]. If none of these methods are successful, the focus should shift to improving the stability of both recombinant TLR5 proteins, which could be informed by $TLR5^{N14}$. This may improve the ability of TLR5-ECD to bind *St*FliC, and would facilitate the recovery of higher TLR5-FL concentrations. As TLR5-FL showed a higher affinity for *St*FliC, it should more readily form complexes, and its structure would also provide novel insight into the structure and function of the poorly studied TM and TIR domains.

5.4. Conclusion

Despite not solving the structures of TLR5 or TLR5-flagellin complexes, this study made significant progress towards this goal through the purification of high concentrations of recombinant TLR5-FL and TLR5-ECD proteins. Future efforts should prioritise optimising protein production and stability, particularly for the TLR5-FL construct, which would improve the likelihood of generating native TLR5:*St*FliC complexes. If successful, the structures of human TLR5 and the TLR5:*St*FliC complex would deepen our understanding of the mechanisms involved in TLR5-flagellin binding and signalling, substantiating many other findings from this dissertation. Ultimately, solving the TLR5-flagellin complex structure is crucial for developing methods to manipulate this receptor. It would also provide insights into the broader mechanisms underlying TLR immune responses, which will have significant implications in developing medical therapeutics targeting their complex signalling pathways.

5.5. Methods

5.5.1. Design and production of TLR5 constructs

The TLR5-FL consensus sequence was obtained from the literature (UniProt ID #O60602), and TLR5-ECD sequence identified based on the TLR5 sequence annotation performed in Chapter 4 (See Fig. 4.1A). An N-terminal GP67 secretion sequence identified in the literature was added to the N-terminus of each TLR5 [17]. Both sequences were codon optimised for *Drosophila melanogaster* and synthesised by IDT.

Primers for Gibson assembly were designed to clone the TLR5 sequences into pLIB plasmids containing C-terminal His tags. Gibson assembly was performed using NEBuilder HiFi DNA Assembly Master Mix (NEB #M5520) following the manufacturer's protocol, and integration was verified by sanger sequencing. Plasmids were transformed into NEB 5-alpha E. coli (NEB #C2987), cultured in 5 ml of LB with 100 µg/ml carbenicillin (Invitrogen), and purified using a ZymoPURE plasmid miniprep kit (Zymo Research # D4212).

5.5.2. SDS-PAGE and Western Blot Analyses

Unless otherwise stated, SDS-PAGE and western blot analyses were performed as described in Chapter 4.5.8. Anti-His western blots used 1:1000 diluted TetraHis (QIAGEN #34670) and Anti-Myc western blots used 1:2000 diluted Anti-Myc Tag Antibody clone 4A6 (Sigma-Aldrich #05-724). Both were stained with the secondary antibody IRDye 880CW (LI-COR #926-68070) diluted 1:10000, and imaged with an Odyssey CLx (LI-COR).

5.5.3. Insect cell transfection and expression trials

Baculoviruses were generated based on the biGBac system [19]. Chemically competent EMBacY cells were transformed with 5 µL plasmid, recovered for 6 hours at 37°C, then grown on LB-Agar plates with Kanamycin, Tetracycline, Gentamicin, XGal, and IPTG for 48 hours. Blue-white screening allowed the identification of transformed colonies, and white colonies were picked and grown in 3 mL LB with 100 µg/ml carbenicillin. These cultures were resuspended, lysed, and cleared with the ZymoPURE plasmid miniprep kit, then 700 µL supernatant

transferred to a 1.5mL tube with an equal volume of 100% isopropanol. DNA was precipitated on ice for 1 hour, then pelleted by centrifugation (20000 rcf, 15 min, 4°C). The pellet was washed twice in 700 µL 70% ethanol, then the media carefully removed from the pellet. The pellet was dried in a safety hood for 1 hour, and carefully resuspended in 30 µL sterile H_2O.

Sf9 insect cells in growth phase were split to a concentration of $1x10^6$/mL, and 2 mL were added to each well of a 6-well plate. A transfection mix (250 µL Sf-900 III SFM medium (Gibco #12658), 5 µL FuGENE HD Transfection Reagent (Promega #E231), 10 µL bacmid DNA) was incubated for 30 minutes, then added to each well. As controls, one well was transfected with a mix containing no plasmid, and one well was left untransfected. Transfections were incubated (27°C, 2 days), then verified by fluorescence microscopy. The supernatant from transfected cells and the untransfected control were expanded to 10 mL freshly resuspended Sf9 cells ($1x10^6$/mL) in a 10 cm dish, then incubated at 27°C for 2 days. The cells were resuspended by pipetting, transferred to a 15 mL tube, then pelleted by centrifugation (800 rpm, 5 min). The supernatant was removed and filtered into sterile 15 mL tubes, serving as the initial viral culture (V0), and stored at 4°C.

YFP expression indicative of transfection was verified by imaging cell pellets with a ChemiDoc. Protein expression in YFP-expressing pellets was verified by a His-tag pulldown. 4.5 mL Lysis buffer (50 mM HEPES (pH7.5), 300 mM NaCl, 5% Glycerol, 0.1% Tween-20, 2 mM Imidazole, 5 mM 2-Mercaptoethanol, 1:100 (v/v) Protease-inhibitor Mix HP PLUS (Serva #39107)) was added to the pellet, then sonicated (10 sec, 30% amplitude). Lysed cells were cleared by centrifugation (4°C, 20000 rcf, 1 hr), then incubated with 20 µL Ni-NTA Agarose (QIAGEN #302) at 4°C for 30 minutes. Beads were washed thrice in 500 µL Lysis buffer, then incubated in Elution buffer (Lysis buffer with 250 mM Imidazole) for 5 minutes. Beads were pelleted by centrifugation (500 rcf, 10 sec), and the supernatant taken as the eluent. Samples from the purification were analysed by SDS-PAGE and western blot to confirm expression.

Verified V0 cultures were amplified twice to generate expression cultures (V2). 25 mL of fresh cells at $1x10^6$/mL were inoculated with 1:100 (v/v) V0 culture, and incubated (3 days, 27°C, 100 rpm). The supernatant was extracted (V1), and 1:100 (v/v) V1 used to inoculate another 25 mL of fresh cells. After 3 days, the supernatant was removed and filter purified for use as the expression culture (V2). Protein expression trials, used to identify optimal conditions, were

performed in Hi-5 insect cells. 50 mL Hi-5 cultures ($1x10^6$/mL) were inoculated with 1:10, 1:100, and 1:1000 (v/v) V2, and 10 mL samples taken at 48, 72, and 96 hours. Protein expression in pellets was measured through his-affinity pulldowns of cell lysates, as previously described. Secreted protein was measured with affinity pulldowns of the cell supernatant. Protein expression was visualised through SDS-PAGE and western blot.

5.5.4. TLR5 construct expression and purification

As His-tag pulldowns on cell lysates showed high non-target interactions, secreted recombinant TLR5 protein was instead purified from the supernatant. For both TLR5 constructs, 400 mL Hi-5 cells ($1x10^6$/mL) in fresh Sf-900 III SFM medium were induced with 1:1000 (v/v) V2, and grown for 72 hours (27°C, 100 rpm). Cells were pelleted by centrifugation (500 rcf, 4°C, 30 min), and the supernatant filtered through a 0.22 µm Millipore Steritop Vacuum Filter (Millipore #S2GPT). The supernatant was run through 2x 5 mL HisTrap HP columns (Cytiva #175248) on an ÄKTA start protein purification system (Cytiva). The columns were washed with 12 CV Lysis buffer (without protease inhibitor), then the sample eluted with 8 CV Elution buffer in 4mL fractions. Eluted fractions were analysed by SDS-PAGE, then those identified containing protein were pooled and concentrated to 1 mL with Pierce 10K MWCO protein concentrators (Thermo Scientific #8852), and quantified with a DS-11+ Spectrophotometer (DeNovix).

Concentrated samples were SEC purified using an ÄKTA Pure protein purification system (Cytiva) with a Superdex 200 10/300 GL column in SEC buffer (50 mM HEPES (pH 7.5), 300 mM NaCl, 5% Glycerol, 1 mM Tris(2-carboxyethyl)phosphin). Later purifications also contained 0.002% n-dodecyl ß-D-maltoside (DDM) or 0.0015% Lauryl maltose neopentyl glycol (LMNG) detergents to improve construct stability. Peaks were analysed by SDS-PAGE to identify fractions containing recombinant TLR5 protein, which were quantified, then stored at -20°C.

5.5.5. TLR5-Flagellin pull-down assay

Recombinant His+Myc-tagged *St*FliC used in this study was purified as described in Chapter 2 (See Chapter 2.4.4). 120 µL of 2.7 mg/mL *St*FliC was digested with 3 µL TEV Protease (NEB #P8112S) in 12.5 µL 10x TEV Protease Buffer (NEB) for 1 hr at 30°C), then added to a column

with 100 µL Ni-NTA agarose. The column was washed with 200 µL SEC buffer, then thrice thrice with 1.5 mL SEC buffer. the Flow-throughs were collected. His-Tag cleavage was confirmed Anti-His and Anti-Myc western blots of flow-through samples. The *St*FliC concentration was quantified with a DS-11+ Spectrophotometer.

50 µg TLR5-FL or TLR5-ECD were incubated (4°C, 2 hrs, shaking) with an equal molar concentration of TEV-cleaved *St*FliC or SEC buffer, and diluted with SEC buffer to a total volume of 360 µL. 50 µL Ni-NTA agarose, pre-equilibrated with Lysis buffer, were added to each sample, and incubated (4°C, 2 hrs, shaking). Beads were washed twice in 500 µL SEC Buffer, then incubated (4°C, 5 min) in 50 µL SEC buffer with 200 mM imidazole. The resin was pelleted by centrifugation (4°C, 500 rcf, 1 min), and the supernatant taken as eluant. Pull-downs were analysed using Anti-His and Anti-Myc western blots.

5.5.6. SEC interaction assay

30 µg of TLR5-FL or TLR5-ECD and an equal molar concentration of *St*FliC were co-incubated while shaking for 2 hours at 4°C. Protein samples were run through a Superose™ 6 Increase 5/150 GL column (Cytiva #29-0915) in a ÄKTA go™ protein purification system (Cytiva) with Low-Salt SEC buffer (SEC buffer with 150mM NaCl and without DDM or LMNG)). Individual proteins were run against co-incubated samples for comparison. Eluted fractions were analysed by SDS-PAGE.

5.5.7. HEK-Blue competition assay

HEK-Blue TLR5 reporter cells (Invivogen) were cultured as described previously (See Chapter 2.4.5). The HEK-Blue competition assay was performed in 24-well plates on cells at 70% confluence. 1 nM *St*FliC and 100, 10, or 1 nM of TLR5-FL or TLR5-ECD were added to the cells. A control was stimulated with 1 nM *St*FliC and Buffer. The plate was incubated for 1 hour (5% CO_2, 37°C), then 20 µL of the supernatant was taken for analysis with the Quanti-Blue assay (InvivoGen #rep-qbs), following the protocol described in chapter 4.5.5. Two biological replicates were performed for each sample. Each well was measured with three technical replicates, with the mean of these used as the final value of each biological replicate.

5.5.8. Glutaraldehyde crosslinking

To identify optimal crosslinking conditions, 25 µg of TLR5-ECD and *St*FliC were mixed and diluted to 100 µL with Low-salt SEC buffer, and crosslinked with 0.1, 0.25, or 0.5% Glutaraldehyde (Sigma-Aldrich #G5882). Samples were incubated at 4, 25, or 37°C, and crosslinking was halted after 5, 15, or 30 minutes by the addition of 11 µL TRIS-HCL (1 M, pH 8). 10 µL of each sample were used for SDS-PAGE analysis.

The Crosslinking SEC trial used 500 µg of TLR5-ECD and an equal molar concentration of *St*FliC, diluted to 1 mL in Low-salt SEC buffer, with 0.5% glutaraldehyde. The mix was incubated for 30 minutes at 37°C, then deactivated with 110 µL TRIS-HCL (1M, pH 8). The sample was dialysed for 30 minutes in Low-salt SEC buffer, then SEC was performed using a Superdex 200 10/300 GL column (Cytiva #17-5175-01) in Low-salt SEC buffer, with 150 µL fractions. The SEC peaks were analysed by SDS-PAGE, Anti-His, and Anti-Myc western blots.

5.5.9. Negative staining

Negative staining on TLR5-ECD co-incubated with *St*FliC was performed by Jan-Hannes Schäfer, from the Möller lab at the Universität Osnabrück. 100 nM of protein (TLR5-ECD:*St*FliC, 1:2) were co-incubated for 1 hour at room temperature in NS buffer (100 mM NaCl, 50 mM Tris (pH 7.5)). Negative staining was performed using 2% (w/v) Uranyl Formate following an established protocol [26,27].

I performed negative staining on crosslinked SEC samples. SEC fractions with bands corresponding to 1:1 TLR5-ECD:*St*FliC complexes (approx. 130 kDa) were quantified using a DS-11+ Spectrophotometer, and diluted to 50 µg/mL. Negative staining was performed as described previously (See Chapter 3.5.2).

5.5.10. Cryo-EM imaging

Cryo-EM imaging was performed exclusively by Jan-Hannes Schäfer, from the Möller lab at the Universität Osnabrück, following an existing protocol [27]. For TLR5-ECD monomer samples, 2 mg/mL of protein (with 0.002% DDM) were imaged. For TLR5-ECD complexes, 1 mg/mL TLR5-ECD (with 0.0015% LMNG) and 2 mg/mL *St*FliC were co-incubated for 2 hours at 4°C rotating, then frozen. All samples were applied onto freshly glow-discharged R1.2/1.3 400 mesh copper grids (Quantifoil) and plunge-frozen in liquid ethane with a Vitrobot Mark IV (Thermo Fisher Scientific) at 100% humidity and 4°C. Micrographs were recorded automatically using Titan Krios microscope (Thermo Fisher Scientific) at 300 kV, at a nominal magnification of 165k (0.68Å/px). A total of 8228 and 6208 movies were recorded for TLR5-ECD and TLR5-ECD:*St*FliC samples, respectively.

5.5.11. Data processing

SDS-PAGE and Western blot images were processed with Image Studio (LI-COR). SEC peaks were compared with the Unicorn 7 evaluation software (Cytiva). Graphs were made with Prism 10 (Graphpad). Structural models were generated with AlphaFold2 [28]. Negative stain and Cryo-EM micrographs were processed with Relion or Cryo-SPARC [29,30].

5.6. References

1. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat Immunol. 2010;11: 373–384.

2. Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, Goodlett DR, et al. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature. 2001;410: 1099–1103.

3. Li D, Wu M. Pattern recognition receptors in health and diseases. Signal Transduction and Targeted Therapy. 2021;6: 1–24.

4. Bell JK, Mullen GED, Leifer CA, Mazzoni A, Davies DR, Segal DM. Leucine-rich repeats and pathogen recognition in Toll-like receptors. Trends Immunol. 2003;24: 528–533.

5. Zhou K, Kanai R, Lee P, Wang H-W, Modis Y. Toll-like receptor 5 forms asymmetric dimers in the absence of flagellin. J Struct Biol. 2012;177: 402–409.

6. Clasen SJ, Bell MEW, Borbón A, Lee D-H, Henseler ZM, de la Cuesta-Zuluaga J, et al. Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. Sci Immunol. 2023;8: eabq7001.

7. Andersen-Nissen E, Smith KD, Bonneau R, Strong RK, Aderem A. A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. J Exp Med. 2007;204: 393–403.

8. Forstnerič V, Ivičak-Kocjan K, Plaper T, Jerala R, Benčina M. The role of the C-terminal D0 domain of flagellin in activation of Toll like receptor 5. PLoS Pathog. 2017;13: e1006574.

9. Ivičak-Kocjan K, Forstnerič V, Panter G, Jerala R, Benčina M. Extension and refinement of the recognition motif for Toll-like receptor 5 activation by flagellin. J Leukoc Biol. 2018;104: 767–776.

10. Yoon S-I, Kurnasov O, Natarajan V, Hong M, Gudkov AV, Osterman AL, et al. Structural basis of TLR5-flagellin recognition and signaling. Science. 2012;335: 859–864.

11. Song WS, Jeon YJ, Namgung B, Hong M, Yoon S-I. A conserved TLR5 binding and activation hot spot on flagellin. Sci Rep. 2017;7: 40878.

12. Tahoun A, Jensen K, Corripio-Miyar Y, McAteer S, Smith DGE, McNeilly TN, et al. Host species adaptation of TLR5 signalling and flagellin recognition. Sci Rep. 2017;7: 17677.

13. Faber E, Tedin K, Speidel Y, Brinkmann MM, Josenhans C. Functional expression of TLR5 of different vertebrate species and diversification in intestinal pathogen recognition. Sci Rep. 2018;8: 11287.

14. Forstnerič V, Ivičak-Kocjan K, Ljubetič A, Jerala R, Benčina M. Distinctive Recognition of Flagellin by Human and Mouse Toll-Like Receptor 5. PLoS One. 2016;11: e0158894.

15. Ve T, Williams SJ, Kobe B. Structure and function of Toll/interleukin-1 receptor/resistance protein (TIR) domains. Apoptosis. 2015;20: 250–261.

16. Balka KR, De Nardo D. Understanding early TLR signaling through the Myddosome. J Leukoc Biol. 2019;105: 339–351.

17. Whitford M, Stewart S, Kuzio J, Faulkner P. Identification and sequence analysis of a gene encoding gp67, an abundant envelope glycoprotein of the baculovirus Autographa californica nuclear polyhedrosis virus. J Virol. 1989;63: 1393.

18. Murphy C.I., Mcintire J.R., Davis D.V., Hodgdon H., Seals J.R., Young E. Enhanced Expression, Secretion, and Large-Scale Purification of Recombinant HIV-1 gp120 in Insect Cells Using the Baculovirus egt and p67 Signal Peptides. Protein Expr Purif. 1993;4: 349–357.

19. Weissmann F, Petzold G, VanderLinden R, Huis, Brown NG, Lampert F, et al. biGBac enables rapid gene assembly for the expression of large multisubunit protein complexes. Proceedings of the National Academy of Sciences. 2016;113: E2564–E2569.

20. Latz E, Verma A, Visintin A, Gong M, Sirois CM, Klein DCG, et al. Ligand-induced conformational changes allosterically activate Toll-like receptor 9. Nat Immunol. 2007;8: 772–779.

21. Wu M, Lander GC. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. Curr Opin Struct Biol. 2020;64: 9.

22. Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, et al. Breaking Cryo-EM Resolution Barriers to Facilitate Drug Discovery. Cell. 2016;165: 1698–1707.

23. Stark H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. Methods Enzymol. 2010;481. doi:10.1016/S0076-6879(10)81005-5

24. Rappsilber J. The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. J Struct Biol. 2011;173: 530.

25. Sokolenko S, George S, Wagner A, Tuladhar A, Andrich JMS, Aucoin MG. Co-expression vs. co-infection using baculovirus expression vectors in insect cell culture: Benefits and drawbacks. Biotechnol Adv. 2012;30: 766–781.

26. Gewering T, Januliene D, Ries AB, Moeller A. Know your detergents: A case study on detergent background in negative stain electron microscopy. J Struct Biol. 2018;203: 242–246.

27. Hofmann S, Januliene D, Mehdipour AR, Thomas C, Stefan E, Brüchert S, et al. Conformation space of a heterodimeric ABC exporter under turnover conditions. Nature. 2019;571: 580–583.

28. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.

29. Scheres SHW. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol. 2012;180: 519–530.

30. Punjani A, Rubinstein JL, Fleet DJ, Brubaker MA. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat Methods. 2017;14: 290–296.

Chapter 6: General Discussion

The gut microbiome plays a critical role in human health and disease, with roles in physiological processes including digestion, nutrient absorption, and the synthesis of essential vitamins [3,4,10,11]. Moreover, the gut microbiome is a key player in the immune system, helping to maintain homeostasis and protect against pathogens, with dysbiosis linked to a range of conditions such as inflammatory bowel disease (IBD), obesity, diabetes, and even mental health disorders [3,37,97,98]. Consequently, the modulation of Toll-like receptor (TLR) signalling pathways, a major avenue of host-microbe interactions, is the focus of numerous therapeutic studies [4,19,26,99,100].

Toll-like receptor 5 (TLR5), which specifically detects and responds to bacterial flagellin, is implicated in several inflammatory conditions including inflammatory bowel diseases (IBD) and cancer. However, no therapeutics have yet been developed, emphasising the need for a deeper understanding of this receptor. This dissertation aimed to elucidate the mechanisms underlying TLR5's discrimination between commensal and pathogenic flagellins, focusing on their structural interactions and the role of the D0 domain. By exploring how flagellin binding modulates TLR5 immune responses, this research contributes to a greater understanding of innate immune signalling, and informs the design of therapies targeting this receptor.

6.1. Characterising silent flagellin interactions with TLR5

In Chapter 2, we explored how TLR5 responded to flagellins from common commensal bacteria in our gut microbiome. Existing studies have focused almost entirely on human pathogens, and no large-scale comparative analyses of flagellin binding had been performed. Our screen showed that TLR5 binding at the D1 domain only correlated weakly with immune activity, challenging existing assumptions about this relationship. Furthermore, a class of flagellins termed ‘silent’ were identified, which displayed strong binding but elicited a weak immune response. Silent flagellins were widespread in the phyla *Lachnospiraceae*, which are highly abundant in the gut microbiome, suggesting it may constitute a widespread mechanism of immune evasion. Flagellins were found to modulate TLR5 activity through an interaction in the D0 domain, which facilitated flagellin binding to preformed TLR5 dimers and drastically increased the intensity of subsequent immune responses [101].

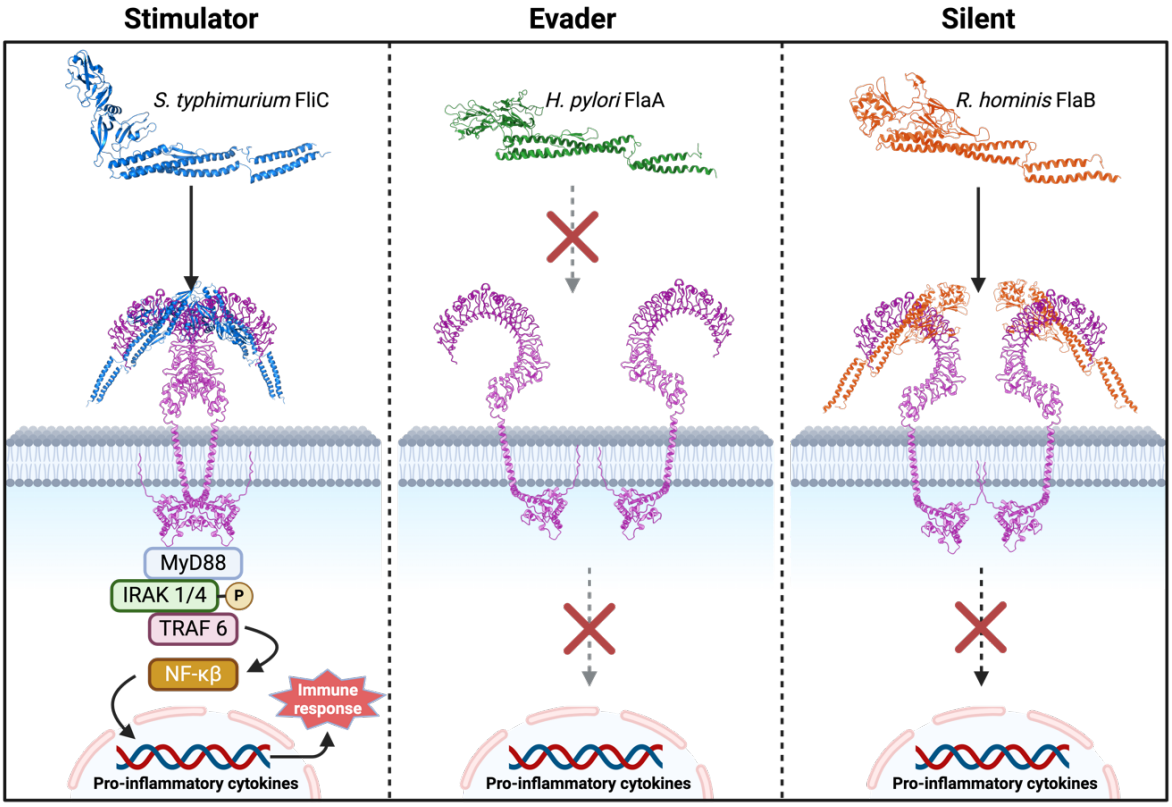


Figure 6.1 – Updated overview of potential TLR5-flagellin interaction outcomes. This illustrative overview shows the basic mechanisms of action in the three established classes of interaction between TLR5 and bacterial flagellins. The structures of representative flagellins and human TLR5 are shown. Stimulators are bound by TLR5 dimers, and mobilise a downstream cytokine response. Evaders are not bound by TLR5, thus evading TLR5 signalling. Silent flagellins are bound by TLR5, but can avoid stimulating a cytokine response. The graphic was made using Biorender, with structures generated with Alphafold2 [105].

Prior to this study, flagellin-TLR5 interactions were viewed through the lens of binding or evasion, where preventing recognition by TLR5, as seen in ε-proteobacteria, was the only mechanism to prevent a TLR5 immune response [83,85]. The discovery of silent flagellins fundamentally challenges this paradigm by identifying a third class of 'silent' interactions, which act separately to stimulation or evasion (Fig. 6.1). The TLR5 response was tuned by the cumulative effect of flagellin binding at multiple sites, which is in keeping with evidence of multiple TLR5-flagellin binding interfaces, flagellin-induced conformational changes observed in the TLR5-LRR9 loop, and evidence ligand-induced conformational change in TLR9 [40,41,58,82]. Silent interactions serve as a less energetically intensive mechanism of immune avoidance, compared to the flagella restructuring or the use of flagella sheaths commonly seen in TLR5 evading bacteria [79, 83-85]. Overall, our results provide a deeper understanding of the TLR5 signalling pathway and mechanisms of immune modulation, highlighting the importance of examining interactions between gut microbiome and our innate immune system.

6.2. The kinetic and structural basis for TLR5-flagellin binding

My first project (Chapter 2) found that both the D0 and D1 domains had a role in silent interactions, but the mechanism by which the D1 domain effected a silent response was unresolved. This was the aim of my second project 2 (Chapter 3), where I performed a novel analysis on the role of kinetics and structure in TLR5-flagellin binding. The stimulatory *S. typhimurium* FliC (*St*FliC) and silent *Roseburia hominis* FlaB (*Rh*FlaB) flagellins displayed distinct kinetic binding profiles when binding $TLR5^{N14}$. *St*FliC formed longer-lasting complexes with TLR5, despite a significantly lower association rate and binding strength. Novel roles for the TLR5 epitope were identified, with *Rh*FlaB was solely reliant on TLR5 epitope for association, while it had no impact on association for *St*FliC, instead playing a role in complex stability. Taken together, these findings suggest the *Rh*FlaB D1 domain facilitates a temporal mechanism of immune evasion, characterised by poor secondary binding site interactions and rapid dissociation. The structure of *Rh*FlaB, solved using Cryo-EM, revealed surface characteristics likely causal to this kinetic profile.

Binding kinetics have historically been very difficult to examine, but recent advancements and the widespread adoption of Surface Plasmon Resonance (SPR) has greatly simplified this process [102–104]. Consequently, my kinetic characterisation of TLR5-flagellin binding is the first of its kind, and can serve as a framework for further studies. The identification of a temporal immune evasion mechanism in *Rh*FlaB is sensible given the importance of cumulative binding interactions shown in Chapter 2, and is consistent with the probable requirement for conformational change in receptor activation. It may also explain how high concentrations of silent flagellins are still able to cause a TLR5 immune response, as the high volume of interactions could increase the likelihood of D0 domain binding despite weak secondary interface interactions [101]. Lastly, the identification of surface motifs potentially causal to binding kinetics is novel, and emphasises the importance of surface features in protein binding dynamics. These findings highlight the importance of binding kinetics and structural characteristics in understanding TLR5-flagellin interactions, inform the design of targeted therapies, and enhance our understanding of innate immune signalling.

6.3. Structural and mechanistic insights into TLR5 signalling

The final two projects (Chapter 4 and 5) both aimed to elucidate mechanisms by which flagellin binding facilitates TLR5 signalling, and obtain structural detail of such interactions. In Chapter 4, I examined how the rotational specificity of the TLR5-TIR dimers impacted their ability to elicit an immune response, with implications for the theory of flagellin-induced conformational change. I was ultimately unable to produce a TLR5-TIR dimer fully capable of stimulating TLR5 signalling as effectively as *St*FliC, though two similar orientations showed partial activation, and could warrant further investigation.

In Chapter 5, I aimed to solve the complete structure of TLR5 and TLR5-flagellin complexes, to directly visualise D0 domain binding and ligand-induced conformational change in TLR5. I designed and produced recombinant TLR5-FL and TLR5-ECD proteins, capable of binding *St*FliC, in sufficient concentrations for structural studies. Negative staining and Cryo-EM identified TLR5 ectodomain monomers in the expected 'C' shaped conformation, and Cryo-EM on co-incubated TLR5-ECD and *St*FliC showed separate low-resolution densities corresponding to each protein, but not in complex. Crosslinking on TLR5-ECD and *St*FliC resolved gel bands corresponding to TLR5-flagellin complexes, but these have yet to be resolved in negative stains.

The identification of partially active TLR5-TIR dimers supports the purported role of flagellin-induced conformational change in TLR5 signalling [41,101]. Furthermore, the inability of most constructs to signal may be a mechanism to prevent non-targeted stimulation by unliganded TLR5 dimers [62]. The production of high concentrations of functional TLR5-FL and TLR5-ECD is a key step in structural analysis, and provides a foundation for continued work. Collectively, these two projects have the potential to build a comprehensive understanding of the mechanics behind flagellin binding and signalling, which would be invaluable to the design of therapeutics targeting TLR5.

6.4. Final Conclusion and Future Perspectives

My work is the most comprehensive investigation into interactions between the gut microbiome and TLR5 to date, with a focus on novel interactions and mechanisms that were unexplored in a human context. The classification of silent interactions between TLR5 and flagellin fundamentally changes our understanding of flagellin-induced TLR5 signalling, and highlights the greater complexity of immune receptor-ligand interactions. My temporal investigation of TLR5-flagellin binding is the first of its kind, and the mechanism of temporal regulation suggested has not been shown in a host-microbe context. Lastly, my success in producing high-concentrations of functional TLR5 for structural analyses is a key step towards solving the elusive human TLR5 and TLR5-flagellin complex structures.

Despite my best efforts, numerous aspects of the TLR5 and flagellin interaction remain unresolved. An expanded study incorporating a wider set of silent flagellins is essential for determining if the mechanisms identified in Chapters 2 and 3 are widely conserved, or if other mechanisms can achieve the same outcome. A temporal investigation of binding between full-length flagellins and TLR5 will better inform the relationship between D1 and D0 domain binding, and clarify how the cumulative effect of multiple flagellin interactions facilitates TLR5 signalling. This could be achieved by modifying the TLR5-ECD constructs generated in Chapter 5 for use in SPR analyses. Identifying and solving the structure of a TLR5-TIR dimer capable of stimulating the TLR5 pathway could succeed following the recommendations in Chapter 4. This dimer would highlight the specificity of conformational change required for TLR5 signalling, and allow the structural characterisation of interactions between TLR5 and MyD88. Improving the stability and functionality of full-length TLR5 and TLR5 ectodomain constructs will facilitate

structural analysis, and can be informed by the design of the $TLR5^{N14}$ or other experimental TLR structures. This would provide the clearest possible view of binding in the human context, and be immensely informative for our understanding of TLR interactions overall.

Through my work, I have laid a foundation for further investigations into novel aspects of TLR5-flagellin binding and signalling. Understanding how microbes modulate TLR5 immune responses, and identifying structural mechanisms and motifs that impact binding kinetics, could prove invaluable for the design of therapeutic interventions targeting TLR5 and similar receptors. More broadly, this research highlights the critical influence of microbial interactions on our immune system, and I look forward to the great scientific and medical benefits which will arise from further studies of this field in the future.

6.5. General References

1. Sender R, Fuchs S, Milo R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS Biol. 2016;14. doi:10.1371/journal.pbio.1002533

2. Ursell LK, Metcalf JL, Parfrey LW, Knight R. Defining the Human Microbiome. Nutr Rev. 2012;70: S38.

3. Shreiner AB, Kao JY, Young VB. The gut microbiome in health and in disease. Curr Opin Gastroenterol. 2015;31: 69.

4. Hou K, Wu Z-X, Chen X-Y, Wang J-Q, Zhang D, Xiao C, et al. Microbiota in health and diseases. Signal Transduction and Targeted Therapy. 2022;7: 1–28.

5. Sonnenburg JL, Bäckhed F. Diet-microbiota interactions as moderators of human metabolism. Nature. 2016;535: 56–64.

6. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M, et al. Diversity of the Human Intestinal Microbial Flora. Science. 2005;308: 1635.

7. Savage DC. Microbial ecology of the gastrointestinal tract. Annu Rev Microbiol. 1977;31: 107–133.

8. Kennedy MS, Chang EB. The microbiome: Composition and locations. Prog Mol Biol Transl Sci. 2020;176. doi:10.1016/bs.pmbts.2020.08.013

9. Arumugam M, Raes J, Pelletier E, Le Paslier D, Yamada T, Mende DR, et al. Enterotypes of the human gut microbiome. Nature. 2011;473: 174–180.

10. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. The Human Microbiome Project. Nature. 2007;449: 804–810.

11. Singh RK, Chang H-W, Yan D, Lee KM, Ucmak D, Wong K, et al. Influence of diet on the gut microbiome and implications for human health. J Transl Med. 2017;15: 1–17.

12. Wilson AS, Koller KR, Ramaboli MC, Nesengani LT, Ocvirk S, Chen C, et al. Diet and the Human Gut Microbiome: An International Review. Dig Dis Sci. 2020;65: 723–740.

13. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, et al. Human gut microbiome viewed across age and geography. Nature. 2012;486: 222–227.

14. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. Science. 2012;336: 1268–1273.

15. Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, et al. Human genetics shape the gut microbiome. Cell. 2014;159: 789.

16. Gilbert JA, Quinn RA, Debelius J, Xu ZZ, Morton J, Garg N, et al. Microbiome-wide association studies link dynamic microbial consortia to disease. Nature. 2016;535: 94–103.

17. Honda K, Littman DR. The microbiota in adaptive immune homeostasis and disease. Nature. 2016;535: 75–84.

18. Chen F, Stappenbeck TS. Microbiome control of innate reactivity. Curr Opin Immunol. 2019;56: 107–113.

19. Thaiss CA, Zmora N, Levy M, Elinav E. The microbiome and innate immunity. Nature. 2016;535: 65–74.

20. Bonilla FA, Oettgen HC. Adaptive immunity. J Allergy Clin Immunol. 2010;125: S33–40.

21. Janeway CA Jr, Travers P, Walport M, Shlomchik MJ. Principles of innate and adaptive immunity. Immunobiology: The Immune System in Health and Disease 5th edition. Garland Science; 2001.

22. Kumar H, Kawai T, Akira S. Pathogen Recognition by the Innate Immune System. Int Rev Immunol. 2011. doi:10.3109/08830185.2010.529976

23. Parkin J, Cohen B. An overview of the immune system. Lancet. 2001;357: 1777–1789.

24. Iwasaki A, Medzhitov R. Control of adaptive immunity by the innate immune system. Nat Immunol. 2015;16: 343–353.

25. Janeway CA. How the immune system works to protect the host from infection: A personal view. Proceedings of the National Academy of Sciences. 2001;98: 7461–7468.

26. Robertson SJ, Goethel A, Girardin SE, Philpott DJ. Innate Immune Influences on the Gut Microbiome: Lessons from Mouse Models. Trends Immunol. 2018;39: 992–1004.

27. Takeuchi O, Akira S. Pattern recognition receptors and inflammation. Cell. 2010;140. doi:10.1016/j.cell.2010.01.022

28. Takeda K, Akira S. Toll-like receptors in innate immunity. Int Immunol. 2005;17: 1–14.

29. Fitzgerald KA, Kagan JC. Toll-like Receptors and the Control of Immunity. Cell. 2020;180: 1044–1066.

30. Kawai T, Akira S. The roles of TLRs, RLRs and NLRs in pathogen recognition. Int Immunol. 2009;21: 317–337.

31. Kumar H, Kawai T, Akira S. Toll-like receptors and innate immunity. Biochem Biophys Res Commun. 2009;388: 621–625.

32. O'Neill LAJ, Golenbock D, Bowie AG. The history of Toll-like receptors - redefining innate immunity. Nat Rev Immunol. 2013;13: 453–460.

33. Athman R, Philpott D. Innate immunity via Toll-like receptors and Nod proteins. Curr Opin Microbiol. 2004;7: 25–32.

34. Medzhitov R. Toll-like receptors and innate immunity. Nat Rev Immunol. 2001;1: 135–145.

35. Ishii KJ, Coban C, Akira S. Manifold Mechanisms of Toll-Like Receptor-Ligand Recognition. J Clin Immunol. 2005;25: 511–521.

36. Rock FL, Hardiman G, Timans JC, Kastelein RA, Fernando Bazan J. A family of human receptors structurally related to Drosophila Toll. Proc Natl Acad Sci U S A. 1998;95: 588.

37. Frosali S, Pagliari D, Gambassi G, Landolfi R, Pandolfi F, Cianci R. How the Intricate Interaction among Toll-Like Receptors, Microbiota, and Intestinal Immunity Can Influence Gastrointestinal Pathology. Journal of Immunology Research. 2015;2015. doi:10.1155/2015/489821

38. Himmel ME, Hardenberg G, Piccirillo CA, Steiner TS, Levings MK. The role of T-regulatory cells and Toll-like receptors in the pathogenesis of human inflammatory bowel disease. Immunology. 2008;125. doi:10.1111/j.1365-2567.2008.02939.x

39. Le Noci V, Bernardo G, Bianchi F, Tagliabue E, Sommariva M, Sfondrini L. Toll Like Receptors as Sensors of the Tumor Microbial Dysbiosis: Implications in Cancer Progression. Frontiers in Cell and Developmental Biology. 2021;9. doi:10.3389/fcell.2021.732192

40. Yoon S-I, Kurnasov O, Natarajan V, Hong M, Gudkov AV, Osterman AL, et al. Structural basis of TLR5-flagellin recognition and signaling. Science. 2012;335: 859.

41. Latz E, Verma A, Visintin A, Gong M, Sirois CM, Klein DC, et al. Ligand-induced conformational changes allosterically activate Toll-like receptor 9. Nat Immunol. 2007;8. doi:10.1038/ni1479

42. Gay NJ, Symmons MF, Gangloff M, Bryant CE. Assembly and localization of Toll-like receptor signalling complexes. Nat Rev Immunol. 2014;14: 546–558.

43. Balka KR, De Nardo D. Understanding early TLR signaling through the Myddosome. J Leukoc Biol. 2018;105: 339–351.

44. Akira S, Takeda K. Toll-like receptor signalling. Nat Rev Immunol. 2004;4: 499–511.

45. Akira S, Yamamoto M, Takeda K. Role of adapters in Toll-like receptor signalling. Biochem Soc Trans. 2003;31: 637–642.

46. Castaño-Rodríguez N, Kaakoush NO, Mitchell HM. Pattern-Recognition Receptors and Gastric Cancer. Front Immunol. 2014;5. doi:10.3389/fimmu.2014.00336

47. Husseinzadeh N, Davenport SM. Role of Toll-like receptors in cervical, endometrial and ovarian cancers: A review. Gynecol Oncol. 2014;135: 359–363.

48. Kim S-J, Chen Z, Chamberlain ND, Essani AB, Volin MV, Asif Amin M, et al. Ligation of TLR5 promotes myeloid cell infiltration and differentiation into mature osteoclasts in RA and experimental arthritis. J Immunol. 2014;193: 3902.

49. Stanislawowski M, Wierzbicki PM, Golab A, Adrych K, Kartanowicz D, Wypych J, et al. Decreased Toll-like receptor-5 (TLR-5) expression in the mucosa of ulcerative colitis patients. J Physiol Pharmacol. 2009;60.

50. Feng S, Zhang C, Chen S, He R, Chao G, Zhang S. TLR5 Signaling in the Regulation of Intestinal Mucosal Immunity. J Inflamm Res. 2023.

51. Lu Y, Li X, Liu S, Zhang Y, Zhang D. Toll-like Receptors and Inflammatory Bowel Disease. Front Immunol. 2018;9: 326971.

52. Hawn TR, Verbon A, Lettinga KD, Zhao LP, Li SS, Laws RJ, et al. A Common Dominant TLR5 Stop Codon Polymorphism Abolishes Flagellin Signaling and Is Associated with Susceptibility to Legionnaires' Disease. J Exp Med. 2003;198: 1563.

53. Carvalho FA, Koren O, Goodrich JK, Johansson MEV, Nalbantoglu I, Aitken JD, et al. Transient inability to manage proteobacteria promotes chronic gut inflammation in TLR5-deficient mice. Cell Host Microbe. 2012;12: 139–152.

54. Hayashi F, Smith KD, Ozinsky A, Hawn TR, Yi EC, Goodlett DR, et al. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature. 2001;410. doi:10.1038/35074106

55. Feuillet V, Medjane S, Mondor I, Demaria O, Pagni PP, Galán JE, et al. Involvement of Toll-like receptor 5 in the recognition of flagellated bacteria. Proc Natl Acad Sci U S A. 2006;103: 12487–12492.

56. Hajam IA, Dar PA, Shahnawaz I, Jaume JC, Lee JH. Bacterial flagellin—a potent immunomodulatory agent. Exp Mol Med. 2017;49: e373–e373.

57. Schwabe RF, Seki E, Brenner DA. Toll-like receptor signaling in the liver. Gastroenterology. 2006;130: 1886–1900.

58. Forstnerič V, Ivičak-Kocjan K, Plaper T, Jerala R, Benčina M. The role of the C-terminal D0 domain of flagellin in activation of Toll like receptor 5. PLoS Pathog. 2017;13: e1006574.

59. Khani M-H, Bagheri M, Dehghanian A, Zahmatkesh A, Moradi Bidhendi S, Salehi Najafabadi Z, et al. Effect of C-Terminus Modification in Salmonella typhimurium FliC on Protein Purification Efficacy and Bioactivity. Mol Biotechnol. 2018;61: 12–19.

60. Lu J, Sun PD. The Structure of the TLR5-Flagellin Complex: A New Mode of Pathogen Detection, Conserved Receptor Dimerization for Signaling. Sci Signal. 2012. doi:10.1126/scisignal.2002963

61. López-Yglesias AH, Zhao X, Quarles EK, Lai MA, VandenBos T, Strong RK, et al. Flagellin induces antibody responses through a TLR5- and inflammasome-independent pathway. J Immunol. 2014;192: 1587–1596.

62. Zhou K, Kanai R, Lee P, Wang H-W, Modis Y. Toll-like receptor 5 forms asymmetric dimers in the absence of flagellin. J Struct Biol. 2012;177: 402–409.

63. Haiko J, Westerlund-Wikström B. The Role of the Bacterial Flagellum in Adhesion and Virulence. Biology. 2013;2: 1242.

64. Duan Q, Zhou M, Zhu L, Zhu G. Flagella and bacterial pathogenicity. J Basic Microbiol. 2013;53: 1–8.

65. Ramos HC, Rumbo M, Sirard J-C. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. Trends Microbiol. 2004;12: 509–517.

66. Faulds-Pain A, Birchall C, Aldridge C, Smith WD, Grimaldi G, Nakamura S, et al. Flagellin redundancy in Caulobacter crescentus and its implications for flagellar filament assembly. J Bacteriol. 2011;193: 2695–2707.

67. Hu D, Reeves PR. The Remarkable Dual-Level Diversity of Prokaryotic Flagellins. mSystems. 2020;5. doi:10.1128/mSystems.00705-19

68. Najafi J, Shaebani MR, John T, Altegoer F, Bange G, Wagner C. Flagellar number governs bacterial spreading and transport efficiency. Science Advances. 2018;4. doi:10.1126/sciadv.aar6425

69. Tan A, Alsenani Q, Lanz M, Birchall C, Drage LKL, Picton D, et al. Evasion of toll-like receptor recognition by Escherichia coli is mediated via population level regulation of flagellin production. Front Microbiol. 2023;14: 1093922.

70. Kühn MJ, Schmidt FK, Farthing NE, Rossmann FM, Helm B, Wilson LG, et al. Spatial arrangement of several flagellins within bacterial flagella improves motility in different environments. Nat Commun. 2018;9: 1–12.

71. Bonifield HR, Hughes KT. Flagellar phase variation in Salmonella enterica is mediated by a posttranscriptional control mechanism. J Bacteriol. 2003;185: 3567–3574.

72. Beatson SA, Minamino T, Pallen MJ. Variation in bacterial flagellins: from sequence to structure. Trends Microbiol. 2006;14: 151–155.

73. Honko AN, Mizel SB. Effects of flagellin on innate and adaptive immunity. Immunol Res. 2005;33: 83–101.

74. Yamaguchi T, Toma S, Terahara N, Miyata T, Ashihara M, Minamino T, et al. Structural and Functional Comparison of Salmonella Flagellar Filaments Composed of FljB and FliC. Biomolecules. 2020;10. doi:10.3390/biom10020246

75. Steimle A, Menz S, Bender A, Ball B, Weber ANR, Hagemann T, et al. Flagellin hypervariable region determines symbiotic properties of commensal Escherichia coli strains. PLoS Biol. 2019;17: e3000334.

76. Malapaka RRV, Adebayo LO, Tripp BC. A deletion variant study of the functional role of the Salmonella flagellin hypervariable domain region in motility. J Mol Biol. 2007;365: 1102–1116.

77. Smith KD, Andersen-Nissen E, Hayashi F, Strobe K, Bergman MA, Barrett SLR, et al. Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. Nat Immunol. 2003;4: 1247–1253.

78. Geis G, Suerbaum S, Forsthoff B, Leying H, Opferkuch W. Ultrastructure and Biochemical Studies of the Flagellar Sheath of Helicobacter Pylori. J Med Microbiol. 1993;38: 371–377.

79. Rhee K-H, Park J-S, Cho M-J. Helicobacter pylori: Bacterial Strategy for Incipient Stage and Persistent Colonization in Human Gastric Niches. Yonsei Med J. 2014;55: 1453.

80. Subramanian N, Qadri A. Lysophospholipid sensing triggers secretion of flagellin from pathogenic salmonella. Nat Immunol. 2006;7: 583–589.

81. Gerstel U, Czapp M, Bartels J, Schröder J-M. Rhamnolipid-induced shedding of flagellin from Pseudomonas aeruginosa provokes hBD-2 and IL-8 response in human keratinocytes. Cell Microbiol. 2009;11: 842–853.

82. Andersen-Nissen E, Smith KD, Bonneau R, Strong RK, Aderem A. A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. J Exp Med. 2007;204: 393–403.

83. Kreutzberger MAB, Ewing C, Poly F, Wang F, Egelman EH. Atomic structure of the flagellar filament reveals how ε Proteobacteria escaped Toll-like receptor 5 surveillance. Proc Natl Acad Sci U S A. 2020;117: 16985–16991.

84. Gewirtz AT, Yu Y, Krishna US, Israel DA, Lyons SL, Peek RM. Helicobacter pylori Flagellin Evades Toll-Like Receptor 5-Mediated Innate Immunity. J Infect Dis. 2004;189: 1914–1920.

85. Andersen-Nissen E, Smith KD, Strobe KL, Barrett SLR, Cookson BT, Logan SM, et al. Evasion of Toll-like receptor 5 by flagellated bacteria. Proceedings of the National Academy of Sciences. 2005;102: 9247–9252.

86. Song WS, Jeon YJ, Namgung B, Hong M, Yoon S-I. A conserved TLR5 binding and activation hot spot on flagellin. Sci Rep. 2017;7: 40878.

87. Terashima H, Kawamoto A, Morimoto YV, Imada K, Minamino T. Structural differences in the bacterial flagellar motor among bacterial species. Biophysics and Physicobiology. 2017;14: 191.

88. Faber E, Tedin K, Speidel Y, Brinkmann MM, Josenhans C. Functional expression of TLR5 of different vertebrate species and diversification in intestinal pathogen recognition. Sci Rep. 2018;8. doi:10.1038/s41598-018-29371-0

89. Keestra AM, de Zoete MR, van Aubel RAMH, van Putten JPM. Functional characterization of chicken TLR5 reveals species-specific recognition of flagellin. Mol Immunol. 2008;45: 1298–1307.

90. Voogdt CGP, Wagenaar JA, van Putten JPM. Duplicated TLR5 of zebrafish functions as a heterodimeric receptor. Proc Natl Acad Sci U S A. 2018;115: E3221.

91. de Zoete MR, Keestra AM, Roszczenko P, van Putten JP. Activation of human and chicken toll-like receptors by Campylobacter spp. Infect Immun. 2010;78. doi:10.1128/IAI.00897-09

92. Forstnerič V, Ivičak-Kocjan K, Ljubetič A, Jerala R, Benčina M. Distinctive Recognition of Flagellin by Human and Mouse Toll-Like Receptor 5. PLoS One. 2016;11: e0158894.

93. Metcalfe HJ, La Ragione RM, Smith DGE, Werling D. Functional characterisation of bovine TLR5 indicates species-specific recognition of flagellin. Vet Immunol Immunopathol. 2014;157: 197–205.

94. Tahoun A, Jensen K, Corripio-Miyar Y, McAteer S, Smith DGE, McNeilly TN, et al. Host species adaptation of TLR5 signalling and flagellin recognition. Sci Rep. 2017;7: 1–13.

95. Pan X, Yang Y, Zhang J-R. Molecular basis of host specificity in human pathogenic bacteria. Emerg Microbes Infect. 2014;3: e23.

96. Bäumler A, Fang FC. Host Specificity of Bacterial Pathogens. Cold Spring Harb Perspect Med. 2013;3. doi:10.1101/cshperspect.a010041

97. Tuddenham S, Sears CL. The intestinal microbiome and health. Curr Opin Infect Dis. 2015;28: 464–470.

98. Sanna S, van Zuydam NR, Mahajan A, Kurilshikov A, Vich Vila A, Võsa U, et al. Causal relationships among the gut microbiome, short-chain fatty acids and metabolic diseases. Nat Genet. 2019;51: 600–605.

99. Hennessy EJ, Parker AE, O'Neill LAJ. Targeting Toll-like receptors: emerging therapeutics? Nat Rev Drug Discov. 2010;9: 293–307.

100. Farooq M, Batool M, Kim MS, Choi S. Toll-Like Receptors as a Therapeutic Target in the Era of Immunotherapies. Front Cell Dev Biol. 2021;9: 756315.

101. Clasen SJ, Bell MEW, Borbón A, Lee D-H, Henseler ZM, de la Cuesta-Zuluaga J, et al. Silent recognition of flagellins from human gut commensal bacteria by Toll-like receptor 5. Sci Immunol. 2023;8: eabq7001.

102. Janith GI, Herath HS, Hendeniya N, Attygalle D, Amarasinghe DAS, Logeeshan V, et al. Advances in surface plasmon resonance biosensors for medical diagnostics: An overview of recent developments and techniques. Journal of Pharmaceutical and Biomedical Analysis Open. 2023;2: 100019.

103. Capelli D, Scognamiglio V, Montanari R. Surface plasmon resonance technology: Recent advances, applications and experimental cases. Trends Analyt Chem. 2023;163: 117079.

104. Wang D, Loo JFC, Chen J, Yam Y, Chen S-C, He H, et al. Recent Advances in Surface Plasmon Resonance Imaging Sensors. Sensors. 2019;19. doi:10.3390/s19061266

105. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021;596: 583–589.

106. Yonekura K, Maki-Yonekura S, Namba K. Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. Nature. 2003;424: 643–650.

Appendix

Author contributions for Chapter 4 and Chapter 5

Author abbreviations:

Michael En Wei Bell[1,2]	MEWB
Sara J. Clasen[1,2]	SJC
Ruth E. Ley[1,2]	REL
Andrei Lupas[3]	AL
Arne Möller[4]	AM
Jan-Hannes Schäfer[4]	JS
Iris Koch[5]	IK

Author affiliations:

[1] Department of Microbiome Science, Max Planck Institute for Biology, Tübingen 72076, Germany.

[2] Cluster of Excellence EXC 2124 Controlling Microbes to Fight Infections, University of Tübingen, Tübingen, Germany.

[3] Department of Protein Evolution, Max Planck Institute for Biology, Tübingen 72076, Germany.

[4] Department of Structural Biology, Max Planck Institute of Biophysics, Frankfurt am Main, Germany.

[5] Electron Microscopy Facility, Max Planck Institute for Biology, Tübingen 72076, Germany.

Author Contributions – Chapter 4:

This project was conceptualized by MEWB, SJC, AL, and REL. GCN4-TIR constructs were designed by MEWB, with assistance from AL. HEK cell work was performed by MEWB, with advice from SJC. All proteins were produced by MEWB. All other experimental work was performed by MEWB. All figures and text were produced by MEWB.

Author Contributions – Chapter 5:

This project was conceptualized by MEWB, SJC, REL, AM, and JS. Plasmids were produced by MEWB and SJC. Recombinant flagellins were produced by MEWB and SJC. All other proteins were produced by MEWB. The HEK cell competition assay was performed by SJC, using TLR5 samples produced by MEWB. Negative staining and Cryo-EM analysis of TLR5-ECD and TLR5-ECD+*St*FliC samples were performed by JS on samples produced by MEWB. Negative stain images of cross-linked SEC-Purified TLR5-ECD:*St*FliC were captured by IK on grids produced by MEWB. All other experimental work was performed by MEWB. Figure 5.4 was adapted from figures made by JS. All other figures were made by MEWB. The text was written by MEWB.