

Molecular mechanisms of platelet activation and immune evasion

Louris Jakob Feitsma

Molecular mechanisms of platelet activation and immune evasion

PhD thesis, Utrecht University, the Netherlands

Colofon

Author: Louris Feitsma

Lay-out: Louris Feitsma & Trusanne Bos

Illustration: Trusanne Bos

Cover Design: Trusanne Bos

Printing: Proefschriftmaken.nl

ISBN: 978-90-393-6857-2

Illustrations IStock.com: 116971217 - Eraxion, 116971297 - Eraxion, 463286527 - royaltystockphoto, 475317618 - Ugreen, 544582082 - Rost-9D, 654815306 - Alex-LMX

© **Louris Feitsma, 2017, Utrecht, The Netherlands.** All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system, or transmitted in any form or by any means without prior permission of the author. The copyright of articles that have been published has been transferred to the respective journals.

Molecular mechanisms of platelet activation and immune evasion

Moleculaire mechanismen betrokken bij activatie van bloedplaatjes en
ontwijking van het immuunsysteem

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de
rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van
het college voor promoties in het openbaar te verdedigen op

woensdag 20 september 2017 des middags te 2.30 uur

door

Louris Jakob Feitsma

geboren op 31 oktober 1984 te Dantumadeel

Promotor

Prof. dr. P. Gros

Copromotor

Dr. E.G. Huizinga

Het in dit proefschrift beschreven onderzoek werd gefinancierd door de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO), binnen het wetenschapsterrein van de Chemische Wetenschappen (CW).

*To make a prairie it takes a clover and one bee,
One clover, and a bee.
And revery.
The revery alone will do,
If bees are few.*

- Emily Dickinson *1830-†1886



Leescommissie

Prof. dr. R.W. Farndale
Prof. dr. Ph. G. de Groot
Prof. dr. J.P.M. van Putten
Prof. dr. J.A.G. van Strijp
Dr. R.J. de Groot

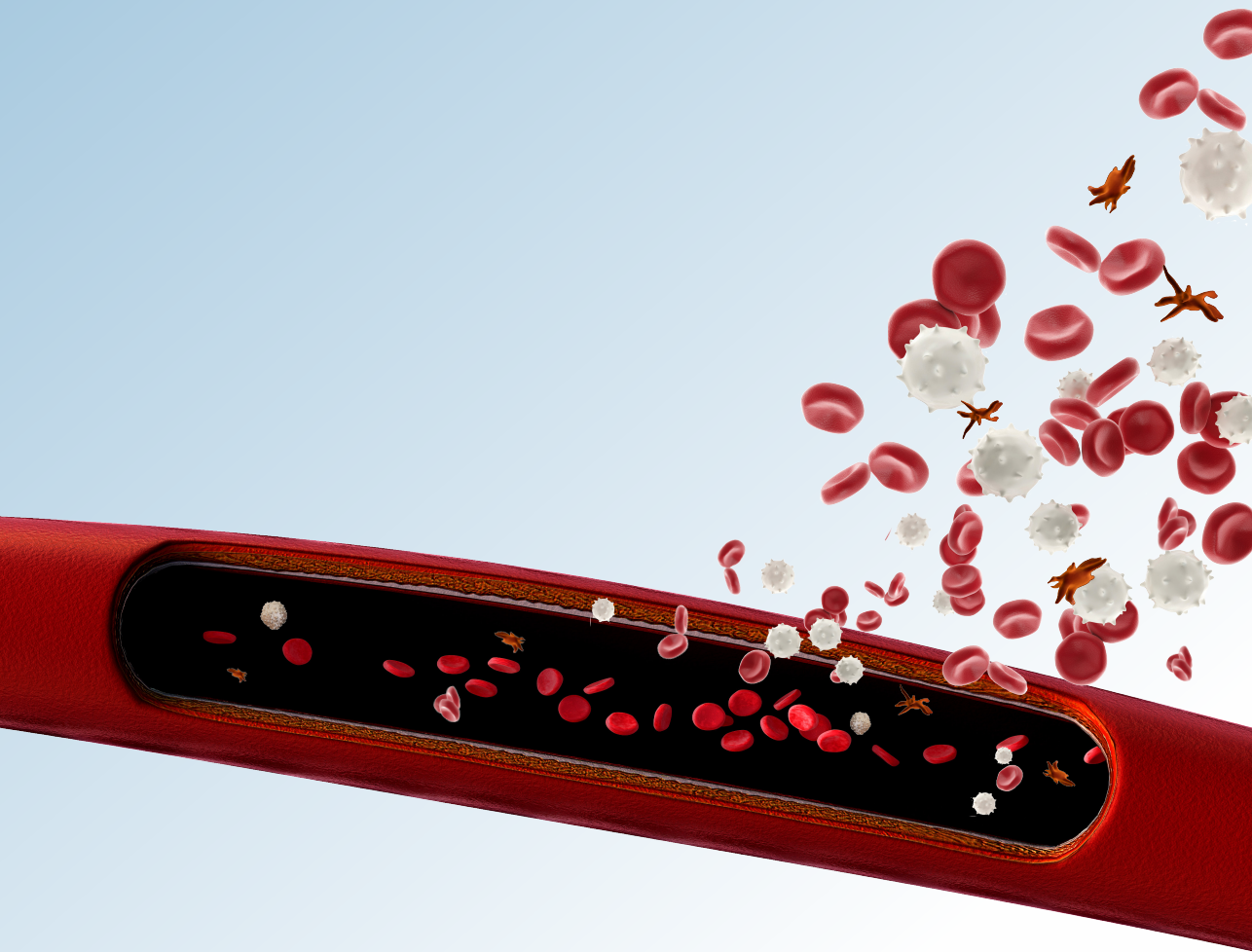
Paranimfen

Revina C. van Scherpenzeel
Deniz Ugurlar

Table of contents

Chapter 1	General introduction	9
Chapter 2	Structural insights into collagen-binding by platelet receptor GPVI <i>Manuscript submitted</i>	21
Chapter 3	Structural basis for inhibition of TLR2 by staphylococcal superantigen-like protein 3 (SSL3) <i>PNAS (2015) vol. 112 p. 11018–11023</i>	51
Chapter 4	Analysis of SSL-specific interactions with Fc-tagged receptor ectodomains using multichannel SPR <i>Manuscript in preparation</i>	75
Chapter 5	General discussion	103
Addendum	Nederlandse samenvatting Dankwoord List of publications Curriculum vitae (NL) Curriculum vitae (EN)	117

1



General introduction

Louris J. Feitsma

Crystal & Structural Chemistry, Bijvoet Center for Biomolecular Research, Department of Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

The human skin is a multi-layer interface providing the first line of defense against external factors. Tissue injury is typically accompanied by vascular damage and creates a sizable risk for excessive blood loss (hemorrhage) and permeation of pathogens (transmission). Incidence of skin-related wounds and lesions therefore demands for fast responses to close off the damaged endothelial or epithelial cell layer and to prevent pathogen proliferation. Exposure of the extracellular matrix of blood vessels triggers a cascade of hemostatic mechanisms that induces adhesion and aggregation of circulating blood platelets at the site of lesion. Initially formed platelet aggregates accelerate the coagulation cascade producing fibrin that cross-links platelets into a reinforced plug that stably seals off the lesion. Concurrently, diverse immune mechanisms facilitate the sensing of a repertoire of pathogen-associated molecular patterns by immune cells. Recognition of the pathogen subsequently provokes downstream immune processes that lead to leukocyte extravasation from the circulation, recruitment and adhesion to the site of infection, and neutralization of the pathogen. Opposing this, various bacteria do not passively undergo clearance and have developed complex immune inhibition mechanisms to evade their elimination. The *Staphylococcus aureus* bacterium secretes a collection of virulence factors that interferes in numerous immune activation processes and has thereby promoted its emergence as a pathogen. Additionally, increasing antibiotic-resistance supported development of difficult remediable variants, such as various MRSA-strains. In this chapter, I will evaluate recent insights into platelet and immune activation as well as evasion mechanisms of *S. aureus*.

Collagen-mediated activation of circulating blood platelets

Vascular strength and integrity is predominantly configured by fibrillar collagens of the extracellular matrix (ECM). In an intact cardiovascular system, an endothelial cell layer separates the blood stream from the extracellular matrix (ECM), and prevents adherence and activation of circulating blood platelets. Disruption of the endothelial cell layer upon vascular lesion exposes platelets to the ECM that functions as a prohemostatic and procoagulant surface¹. Under these conditions, collagens of the ECM promote deceleration, adhesion, and activation of circulating platelets (primary hemostasis)². At the same time, exposed subendothelial tissue factor induces a signaling cascade that leads to the polymerization of fibrin and clotting of blot (coagulation)¹. Platelet activation and coagulation are, initially, mutually reinforcing processes leading to the formation of collagen-attached platelet aggregates crosslinked by fibrin polymers that form a stable platelet plug that closes off the site of lesion.

The roles of collagens both in establishing vascular integrity and initiation of platelet adhesion emphasizes its important function in the circulatory system. Fibrillar collagens are the major proteinaceous ECM constituents and composed of polypeptide chains of approximately 1000 amino acid in length. Each polypeptide chain is build up from repetitive Gly-X-Y triads, in which X and Y are more often proline (P) and 4-hydroxyproline (O) than other natural occurring amino acids, and forms a right-handed helix. The most abundant vascular collagen types are type I (60%) and III (30%)³, that form respectively I(α 1 α 1 α 2)-type heterotypic and III(α 1 α 1 α 1)-type homotypic triple-helical structures, consisting of three equal collagen chains (III) or a two-plus-one combination of two homologous but not

identical chains (I). Triple helices multimerize into robust fibrillar structures^{4,5} that provide vascular strength. Platelet-adhesive mechanisms, however, are predominantly regulated by specific amino-acid sequences that in triple helical context constitute discrete sites for plasma Von Willebrand factor (VWF) and platelet receptors.

Platelet adhesion is mediated by a number of consecutive collagen-binding events. First, blood plasma protein Von Willebrand factor (VWF) binds with its A3-domain to a site constituted by the III(α I)RGQGVGMGF-sequence in collagen III⁶, or the degenerate I(α 1)RGQAGVMGF/(α 2)RGEONGIGF-sequence in collagen I⁷. Platelet receptor Glycoprotein Iba ($\text{GPIIb}\alpha$) of the GPI-IX receptor complex binds the A1-domain of immobilized VWF⁸, thereby reducing platelet velocity under conditions of high shear and establishing initial platelet tethering at the site of exposed collagen. Second, platelet-associated integrin $\alpha_2\beta_1$ binds to sites formed by GXX'GER- or GXX'GEN-motifs that scatter throughout the collagen I and III chains⁹⁻¹³, and provides stable platelet adhesion to collagen. Finally, Glycoprotein VI (GPVI) of the GPVI-Fc Receptor γ -chain (Fc-R γ) complex binds to GPO-rich regions in collagen¹⁴⁻¹⁶, thereby inducing platelet activation. The molecular details of GPVI-collagen binding and exact sequences in collagen required for GPVI-binding, however, remain elusive.

Whereas platelet adhesive and activation roles may not be fully distinguishable², signaling by the GPVI-FcR γ complex appears rapid and effective, and is therefore assumed to be primarily responsible for transmitting collagen-induced signals¹⁷. Similar to other Fc-R γ complexes that respond to other, receptor-specific stimuli, collagen-binding by GPVI provokes phosphorylation of the intracellular tyrosines of the ITAM-motif of Fc-R γ , which via a number of intermediate steps leads to an increase of the intracellular Ca^{2+} levels¹⁸. Cooperative interplay of three collagen-binding systems, VWF-GPI-IX-V, $\alpha_2\beta_1$, and GPVI-Fc-R γ , triggers the secretion of α -granules by activated platelets, which contain growth factors and chemokines that support vessel repair and recruit endothelial cell progenitors, respectively¹⁹.

Blood platelets and tightly regulated collagen-adhesive processes play a vital role in vascular repair. Malfunctioning of this system by diverse causes can lead to thrombotic or hemorrhagic complications. Vessel wall thickening or thinning by respectively accumulation of collagens and collagen degradation may cause thrombosis or aneurysm²⁰. GPVI-deficiency is linked to a mild bleeding disorder²¹, although it is rarely found in patients²². The physiological importance of GPVI in hemostasis is therefore not exactly clear, but its role in platelet activation makes it an interesting target for anti-thrombotic drug development.

Immune activation by recognition of pathogen-associated molecular patterns

The innate immune system functions as a first-line defense against invading pathogens, and is equipped with a repertoire of mechanisms to establish immediate immune responses upon the recognition of conserved microbial patterns. The existence of such pattern-recognition receptors (PRRs) has been suggested for the first time in 1989²³, but has not been demonstrated until Lemaitre and Hoffmann detected upregulated expression of the anti-fungal peptide Drosomycin via *Drosophila melanogaster* development factor,

termed 'Toll', in 1996^{24,25} – work for which the Nobelprize was awarded in 2011²⁶. Shortly after Toll's role in *Drosophila* was revealed, similar roles in mammalian immunity were identified for the human homolog 'hToll'²⁷, later renamed to Toll-like receptor 4 (TLR4)²⁸, which mediates sensing of bacterial cell wall constituent lipopolysaccharide (LPS)^{29–31}. The molecular details of the LPS-recognition mechanisms of TLR4, via binding to co-receptor myeloid differentiation factor 2 (MD2), were unraveled in the years after that^{32,33}, parallel with the discovery of novel PRRs and interrelated pathogen-associated molecular patterns (PAMPs)^{34,35}. The occurrence of conserved recognition mechanisms in both the animal and plant kingdom³⁶ reveals that archetype processes of host defense must have emerged during early development of eukaryotes. Substantial progression in understanding of pattern recognition in innate immunity, however, has been achieved within only a remarkably short and recent timeline³⁷.

Central component in the detection of PAMPs is the family of ten TLRs. Stimulation of all TLRs entails molecular patterns that are essential for survival of the microbe and stain the fellowship of non-self microorganisms inside the host. The TLR-family is

Table 1. Toll-like receptors (TLRs) in innate immunity

TLR-dimer	Location	Co-receptor(s)	Recognized PAMP(s)	Crystal structures (PDB-IDs)	Intracellular signalling pathway(s)
TLR2-TLR1	Plasma membrane	CD36, RP105	Triacyl lipopeptides ⁴⁰ (Gram ⁻ bacteria)	2z7x ⁵⁴ (Pam ₃ CSK ₄)	MYD88 → NF-κB → TNFs/ILs
TLR2-TLR6	Plasma membrane	CD36, RP105	Diacyl lipopeptides ³⁹ (Gram ⁺ bacteria)	3a79 ⁵³ (Pam ₂ CSK ₄)	MYD88 → NF-κB → TNFs/ILs
TLR3	Endosomal membrane		dsRNA (viruses) ⁴⁵ ssRNA (West Nile virus) ⁵¹	3ciy ⁷⁰ (dsRNA)	TRIF → IRF3/7 → IFNs TRIF → NF-κB → TNFs/ILs
TLR4	Plasma/ Endosomal membrane	MD2, CD14, LBP, RP105	Lipopolysaccharide ³¹ (LPS; Gram ⁻ bacteria)	3fxi (MD2-LPS) ³² 2z64 (MD2) ⁷¹	TRAM → TRIF → IRF3/7 → IFNs TRAM → TRIF → NF-κB → TNFs/ILs MYD88 → NF-κB → TNFs/ILs
TLR5	Plasma membrane		Flagellin ⁴¹ (Flagellated bacteria)	3v47 (Salmonella flagellin) ⁷²	MYD88 → NF-κB → TNFs/ILs
TLR7	Endosomal membrane		ssRNA ⁴⁷ (viruses)	5gmf (guanosine-polyU) ⁷³	MYD88 → NF-κB → TNFs/ILs MYD88 → IRF7 → IFNs
TLR8	Endosomal membrane		ssRNA ⁴⁸ (viruses)	4r08 (ssRNA) ⁷⁴ 3w3l (R848) ⁷⁵	MYD88 → NF-κB → TNFs/ILs MYD88 → IRF7 → IFNs
TLR9	Endosomal membrane		dsDNA CpG-motifs ^{43,44} (bacteria and viruses)	3wpc (dsDNA) ⁷⁶	MYD88 → NF-κB → TNFs/ILs MYD88 → IRF7 → IFNs
TLR10			-	-	Immune regulation ⁷⁷

Abbreviations used: CD14, CD36, cluster of differentiation 14 and 36; dsDNA, double-stranded DNA; dsRNA, double-stranded RNA; IFN, interferon; IL, interleukin; IRF3/7, interleukin-3/7 receptor-associated kinase; LBP, lipopolysaccharide binding protein; LPS, lipopolysaccharide; MD2, myeloid differentiation factor 2; MyD88, myeloid differentiation primary response protein 88; Pam₂, dipalmitoyl lipopeptide; Pam₃, tripalmitoyl lipopeptide; polyU, poly-uracil; NF-κB, nuclear factor κB; R848, resiquimod (imidazoquinoline compound⁷⁸); RP105, Radio Protective protein 105; ssRNA, single-stranded RNA; TNF, tumor necrosis factor.

generally subdivided into groups of plasma transmembrane receptors, that sense and respond to extracellular stimuli, and endosome transmembrane receptors recognizing and reacting to endosomal components³⁸. TLR2, together with TLR6 and 1, associates with bacterial lipopeptides^{39,40}, TLR4 with LPS³¹, and TLR5 with bacterial flagellin⁴¹ (Table 1); all constituents that, not including LPS-internalization via co-receptor CD14⁴², primarily remain outside the host cell. By contrast, endosomal TLR3 and TLR7-9 bind non-self nucleic acid patterns⁴³⁻⁵¹, that are host-internalized, predominantly by viruses, to promote pathogen proliferation. Hallmark of TLRs is the formation of active homodimers when exposed to microbial patterns⁵². One exception to this is TLR2, which forms heterodimers with either TLR6⁵³ or TLR1⁵⁴ that discriminate between diacylated lipopeptides from Gram-negative bacteria and triacylated lipopeptides from Gram-positive bacteria.

TLR-activating signals are relayed, independent of its subcellular location, via the conserved cytoplasm-exposed Toll/Interleukin-1 receptor (TIR) domains and trigger the myeloid differentiation primary response protein 88 (MyD88) pathway and/or the TIR-domain-containing adapter-inducing interferon- β (TRIF) pathway³⁸. All TLRs except for TLR3 activate the MyD88-dependent signaling cascade provoking expression of nuclear factor NF- κ B⁵⁵. Activation of the TRIF-dependent pathway by TLR3, as well as TLR4, also leads to expression of transcription factor NF- κ B, but additionally induces expression of interferon regulatory factors (IRFs)³⁸. Both pathways initiate transcription of pro-inflammatory chemokines and cytokines such as interleukins responsible for further immune signaling, but fine-specificity and ultimate survival or programmed death (apoptosis) of the infected cell is tightly controlled by cross-talk between the pathways³⁸. The secretion of pro-inflammatory interleukins bridges the innate and adaptive immune system and triggers downstream immune activity: leukocyte extravasation from the circulation, migration through the extracellular matrix, chemoattraction to the site of infection, and ultimate killing of the pathogen⁵⁶.

Despite their important role in microbial sensing, TLRs have also been linked to several diseases where immune responses cannot be directly attributed to bacterial or viral infection, and TLR-sensing may fail to discriminate between self and non-self. Aberrant activation or inadequate regulation of TLR-signaling is implicated in the pathogenesis of autoimmunity^{57,58}, and suggested to contribute to the development of rheumatoid arthritis (TLR2-4)⁵⁹⁻⁶¹, multiple sclerosis (TLR2 and 4)⁶²⁻⁶⁴, systemic lupus erythematosus (production of auto-antibodies; linked to TLR7-9)⁶⁵⁻⁶⁷. Furthermore, TLR-signaling is associated with the progression of various cancers⁶⁸. Overall, the distinct TLR-pathways are therefore mentioned as putative targets for the development of novel drugs, predominantly for remedy of inflammatory diseases⁶⁹.

Immune evasion mechanisms of *Staphylococcus aureus*

The gram-positive *Staphylococcus aureus* bacterium is the common cause of skin and soft-tissue infections in the human population⁷⁹. It is continuously colonized in a growing percentage of the population and resides as a commensal bacterium on the skin, in the nares and in the digestive system. Infection of the circulatory system or lungs, however, frequently progresses into more serious diseases, such as endocarditis⁸⁰, sepsis⁸¹, pneumonia⁸². Although mild skin and bloodstream infections appear recurrent in 8-33% of

the cases⁸³, the immune activation cascade is generally sufficient for successful elimination of this pathogen in individuals with a properly functioning immune system. High-risk groups, including children, elderly, and patients with immunodeficiency or autoimmune diseases, however, are more susceptible to *Staphylococcus* infections. Emergence of increasingly antibiotic-resistant variants, such as various methicillin-resistant *S. aureus* (MRSA) strains in hospitals and in the community, is a world-wide growing concern, and necessitates the development of novel vaccines and therapeutics⁸⁴.

Hallmark of *S. aureus* is its capacity to modulate immune signaling of the host organism at various activation stages of the innate and adaptive immune system. More so than other bacteria, *S. aureus* secretes a repertoire of virulence factors, of which an increasing number of components has been linked to vital immune processes⁸⁵. Immune evasion functions comprise a variety of inhibitory mechanisms that interfere in neutrophil extravasation, chemotaxis, complement activation, phagocytosis, and neutrophil-mediated killing of the bacteria, but also include killing of the host cell (recently reviewed in *Thammavongsa et al.*⁸⁶). Among the *S. aureus* secretome, the family of 14 staphylococcal superantigen-like protein (SSLs) appears an important determinant for interference in early stages of immune activation, as derived from the SSL-functions unraveled thus far (Table 2). A subset of the SSL-family encompassing SSL2-6 and SSL11 also exhibits the capacity to associate with sialyl-Lewis^x (sLe^x) glycan structures⁸⁷, which are abundantly present on immune cells⁸⁸. Since most of the factors targeted by SSLs of this subset are situated on the immune cell surface, sLe^x-binding capacities may assist in cell surface enrichment and target-binding.

Unremitted colonization of *S. aureus* and recurrent invasion of the host organism has been strongly promoted by the development of immune evasion mechanisms during millions of years of co-evolution. The total *S. aureus* secretome has previously been estimated to consist of 200 to 600 proteins⁸⁹, and suggests that even more proteins than currently described are involved in yet unidentified immune modulating processes. This is illustrated by SSL2, SSL9, and SSL12-14, which have not been linked to a specific host target, and whether or not these proteins also contribute to immune evasion therefore remains elusive. Opposite to that, some SSLs, such as SSL5, have been associated with multiple targets in both immunity and hemostasis. Here, additional studies are required to address specificity of these host-microbe interactions, and to discriminate between target-specific binding and interactions solely relying on glycan-binding. Overall, the development of novel therapeutics that raise protective immunity against *S. aureus* infections is a challenging task, and must combat the multifaceted capacity of this bacterium to evade immunity.

Conclusion and scope of this thesis

All vertebrates are equipped with vital signal transduction mechanisms that establish vascular repair and neutralization of transmitted pathogens. Inappropriate, disproportionate, or deficient responses in hemostasis and immunity frequently lead to thrombotic or inflammatory diseases. Moreover, immune inhibition by external factors, such as components of the *S. aureus* secretome, is a regular cause of mild or more threatening infection diseases. Modern clinical interference demands for novel therapeutic strategies

Table 2. Family of Staphylococcus superantigen-like proteins and their proposed targets and functions

Name	Gene (location)	Putative Target(s)	Proposed Function	Cell surface	sLe ^x binding?
SSL1	<i>ssl1</i> (vSaα)	MMPs ⁹⁰	Inhibition of neutrophil chemotaxis and migration	Interstitial space	No
SSL2	<i>ssl2</i> (vSaα)	-			Yes
SSL3	<i>ssl3</i> (vSaα)	TLR2 ^{91,92}	Inhibition of neutrophil activation	Immune cell	Yes
SSL4	<i>ssl4</i> (vSaα)	TLR2 ^{91,92}	Inhibition of neutrophil activation	Immune cell	Yes
SSL5	<i>ssl5</i> (vSaα)	PSGL-1 ⁹³	Inhibition of neutrophil migration	Immune cell	Yes
		GPCRs ⁹⁴	Inhibition of leukocyte activation	Immune cell	
		MMPs ^{90,95}	Inhibition of neutrophil chemotaxis and migration	Interstitial space	
		GPIIbα ⁹⁶ , GPVI ⁹⁷	Platelet activation	Blood platelet	
SSL6	<i>ssl6</i> (vSaα)	CD47 ⁹⁸	(Early) induction of phagocytosis	Immune cell	Yes
SSL7	<i>ssl7</i> (vSaα)	IgA ^{99,100}	Inhibition of phagocytosis	<i>S. aureus</i>	No
		C5 ¹⁰⁰⁻¹⁰²	Inhibition of complement activation	<i>S. aureus</i>	
SSL8	<i>ssl8</i> (vSaα)	Tenascin C ¹⁰³	Inhibition of keratinocyte motility	Extracellular matrix	No
SSL9	<i>ssl9</i> (vSaα)	-			No
SSL10	<i>ssl10</i> (vSaα)	IgG1 ^{104,105}	Inhibition of complement activation	<i>S. aureus</i>	No
		Complement C4 ¹⁰⁶	Inhibition of complement activation	<i>S. aureus</i>	
		CXCR4 ¹⁰⁷	Inhibition of neutrophil migration	Immune cell	
		Factor Xa ¹⁰⁸	Inhibition of blood coagulation	Blood plasma	
SSL11	<i>ssl11</i> (vSaα)	FcαRI ¹⁰⁹	Inhibition of phagocytosis	Immune cell	Yes
		PSGL-1 ¹⁰⁹	Inhibition of neutrophil migration	Immune cell	
SSL12	<i>ssl12</i> (IEC2)	-			No
SSL13	<i>ssl13</i> (IEC2)	-			No
SSL14	<i>ssl14</i> (IEC2)	-			No

Abbreviations used: C4 & C5, complement 4 and 5; CD47, Cluster of differentiation 47; CXCR4, C-X-C type chemokine receptor 4; FcαRI, Fc (Fragment crystallizable region)-receptor for IgA; GPCR, G-protein coupled receptor; GPIIbα, Glycoprotein Ib α-subunit; GPVI, Glycoprotein VI; IEC2, immune evasion cluster 2; IgA & IgG1, Immunoglobulin A and G1; MMP, Matrix metalloproteinase; PSGL-1, P-selectin glycoprotein ligand-1; sLe^x, tetrasaccharide sialyl Lewis^x; TLR2, Toll-like receptor 2; vSaα, pathogenicity island α of *S. aureus* virulence factors; Xa, Coagulation Factor Xa.

that act on the level of tight cellular regulation. However, fundamental understanding of the underlying processes linked to these diseases is required prior to the development of new generations of drugs and antibiotics. This thesis describes novel molecular and mechanistic insights into processes of platelet activation, immune activation, and immune inhibition. In **chapter 2**, we present crystal structures of two GPVI-collagen complexes that provide the molecular details of GPVI-mediated platelet activation. We show that the most N-terminal Ig-like domain (D1-domain) encompasses a conserved collagen-binding site located across the D1 β-sheet formed by strands C, D, F, and F'. Interactions with collagen involve two of the three chains forming a triple helix and require a six-residue GPVI-binding site constituted by the canonical OGPOGP-sequence. Also small amino acids may be tolerated at the first or last position. In **chapter 3**, we present the crystal structure of the SSL3-TLR2 complex, which reveals the molecular and mechanistic basis for SSL3-mediated inhibition of TLR2-activation. SSL3 binds near the lipopeptide pocket

of TLR2 and blocks TLR2-activation via inhibition of lipopeptide-binding and dimerization with TLR6 or TLR1. In **chapter 4**, we present a novel method to elucidate target-specific and SSL-glycan contributions to binding of *S. aureus* SSLs to host targets in immunity and hemostasis. Together with the crystal structure presented in chapter 3, we show that TLR2-inhibition by SSL3 is independent of the TLR2-glycans. SSL6-CD47 binding comprises mixed interactions, confirming target-specific binding to CD47. Opposite to that, SSL5-binding to GPVI is mediated by interactions with GPVI-glycans only, and GPVI is therefore not a specific target of SSL5. The results presented in these three chapters are summarized in **chapter 5**. Also discussed are implications following from our work for current progress in the development anti-thrombotic, anti-inflammatory, and anti-bacterial drugs.

References

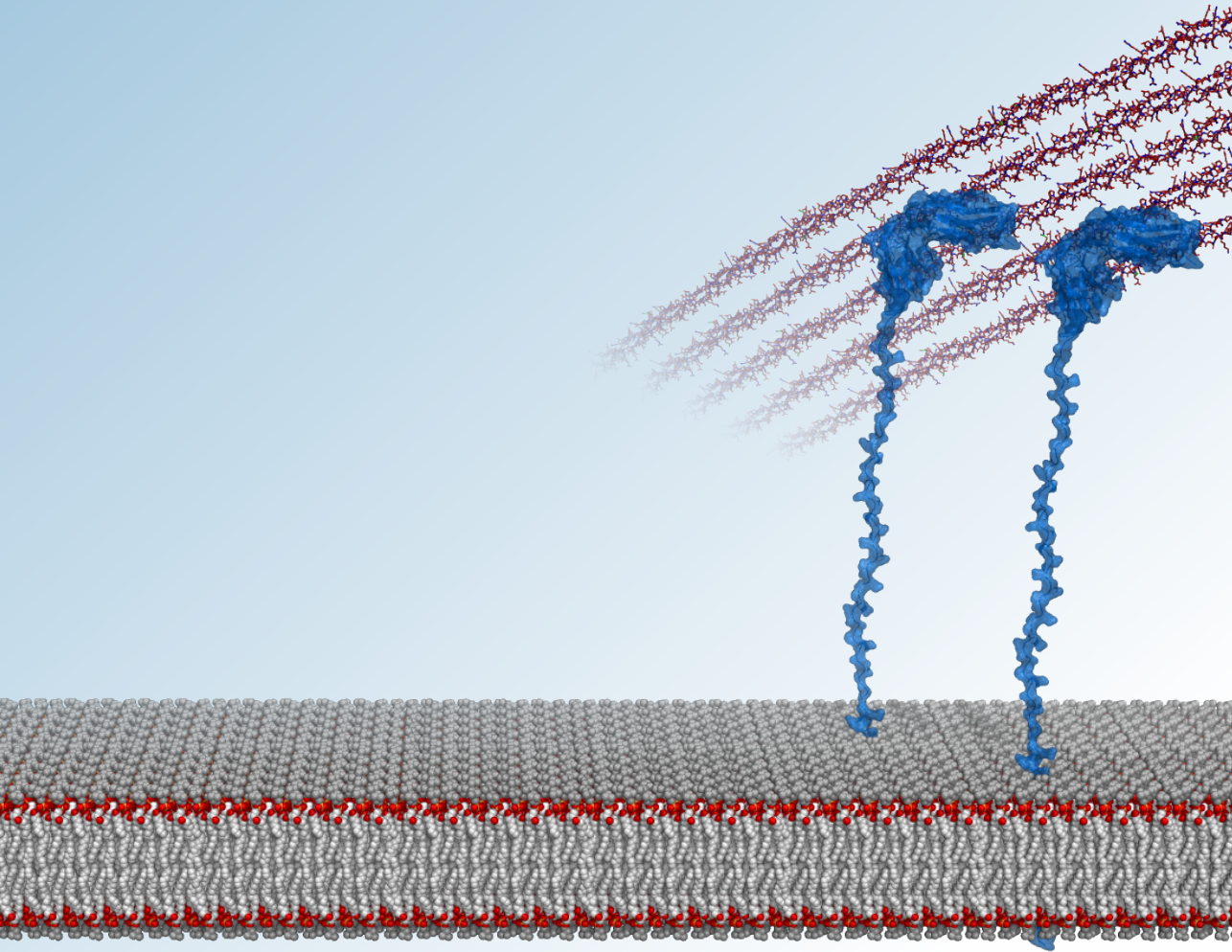
1. Versteeg, H. H. *et al.* New fundamentals in hemostasis. *Physiol. Rev.* **93**, 327–58 (2013).
2. Farndale, R. W., *et al.* The role of collagen in thrombosis and hemostasis. *J. Thromb. Haemost.* **2**, 561–73 (2004).
3. Howard, P. S. & Macarak, E. J. Localization of collagen types in regional segments of the fetal bovine aorta. *Lab. Invest.* **61**, 548–55 (1989).
4. Orgel, J. P. R. O. *et al.* Microfibrillar structure of type I collagen in situ. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9001 (2006).
5. Perumal, S., Antipova, O. & Orgel, J. P. R. O. Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 2824–9 (2008).
6. Lisman, T. *et al.* A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. *Blood* **108**, 3753–6 (2006).
7. Brondijk, T. H. C. *et al.* Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3 domain complex. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5253–5258 (2012).
8. Huizinga, E. G. *et al.* Structures of glycoprotein Ibalpha and its complex with von Willebrand factor A1 domain. *Science* **297**, 1176–9 (2002).
9. Raynal, N. *et al.* Use of synthetic peptides to locate novel integrin $\alpha_2\beta_1$ -binding motifs in human collagen III. *J. Biol. Chem.* **281**, 3821–31 (2006).
10. Emsley, J. *et al.* Structural basis of collagen recognition by integrin alpha2beta1. *Cell* **101**, 47–56 (2000).
11. Knight, C. G. Identification in Collagen Type I of an Integrin $\alpha_2\beta_1$ -binding Site Containing an Essential GER Sequence. *J. Biol. Chem.* **273**, 33287–33294 (1998).
12. Knight, C. G. *et al.* The Collagen-binding A-domains of Integrins $\alpha_1\beta_1$ and $\alpha_2\beta_1$ recognize the Same Specific Amino Acid Sequence, GFOGER, in Native (Triple-helical) Collagens. *J. Biol. Chem.* **275**, 35–40 (2000).
13. Siljander, P. R. M. *et al.* Integrin activation state determines selectivity for novel recognition sites in fibrillar collagens. *J. Biol. Chem.* **279**, 47763–47772 (2004).
14. Smethurst, P. A. *et al.* Structural basis for the platelet-collagen interaction: the smallest motif within collagen that recognizes and activates platelet Glycoprotein VI contains two glycine-proline-hydroxyproline triplets. *J. Biol. Chem.* **282**, 1296–304 (2007).
15. Jarvis, G. E. *et al.* Identification of a major GpVI-binding locus in human type III collagen. *Blood* **111**, 4986–4996 (2008).
16. Jung, S. M. *et al.* Constitutive dimerization of glycoprotein VI (GPVI) in resting platelets is essential for binding to collagen and activation in flowing blood. *J. Biol. Chem.* **287**, 30000–30013 (2012).
17. Nieswandt, B. & Watson, S. P. Platelet-collagen interaction: Is GPVI the central receptor? *Blood* **102**, 449–461 (2003).
18. Watson, S. P. *et al.* GPVI and integrin $\alpha_{IIb}\beta_3$ signaling in platelets. *J. Thromb. Haemost.* **3**, 1752–1762 (2005).
19. Smyth, S. S. *et al.* Platelet functions beyond hemostasis. *J. Thromb. Haemost.* **7**, 1759–1766 (2009).
20. Manon-Jensen, T. *et al.* Collagen-mediated hemostasis. *J. Thromb. Haemost.* **14**, 438–448 (2016).
21. Moroi, M., Jung, S. M., Okuma, M. & Shinmyozu, K. A patient with platelets deficient in glycoprotein VI that lack both collagen-induced aggregation and adhesion. *J. Clin. Invest.* **84**, 1440–5 (1989).
22. Nurden, A. & Nurden, P. Advances in our understanding of the molecular basis of disorders of platelet function. *J. Thromb. Haemost.* **9 Suppl 1**, 76–91 (2011).
23. Janeway, C. A. Approaching the Asymptote? Evolution and Revolution in Immunology. *Cold Spring*

- Harb. Symp. Quant. Biol.* **54**, 1–13 (1989).
24. Lemaitre, B. *et al.* The dorsoventral regulatory gene cassette spatzle/Toll/Cactus controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973–983 (1996).
 25. Lemaitre, B. The road to Toll. *Nat. Rev. Immunol.* **4**, 521–527 (2004).
 26. Nobel Media AB. Jules A. Hoffmann - Facts. http://www.nobelprize.org/nobel_prizes/medicine/laureates/2011/hoffmann-facts.html (2014). Available at: http://www.nobelprize.org/nobel_prizes/medicine/laureates/2011/hoffmann-facts.html. (Accessed: 1st May 2017)
 27. Medzhitov R, Preston-Hurlburt P & Janeway CA Jr. A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
 28. Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. a & Bazan, J. F. A family of human receptors structurally related to *Drosophila* Toll. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 588–593 (1998).
 29. Poltorak, A. *et al.* Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* **282**, 2085–8 (1998).
 30. Qureshi, S. T. *et al.* Endotoxin-tolerant mice have mutations in Toll-like receptor 4 (Tlr4). *J. Exp. Med.* **189**, 615–25 (1999).
 31. Hoshino, K. *et al.* Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hypo-responsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. *J. Immunol.* **162**, 3749–52 (1999).
 32. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* **458**, 1191–5 (2009).
 33. Ohto, U., Fukase, K., Miyake, K. & Shimizu, T. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 7421–6 (2012).
 34. Akira, S., Uematsu, S. & Takeuchi, O. Pathogen recognition and innate immunity. *Cell* **124**, 783–801 (2006).
 35. Kawai, T. & Akira, S. The roles of TLRs, RLRs and NLRs in pathogen recognition. *Int. Immunol.* **21**, 317–337 (2009).
 36. Whitham, S. *et al.* The product of the tobacco mosaic virus resistance gene N: Similarity to toll and the interleukin-1 receptor. *Cell* **78**, 1101–1115 (1994).
 37. O'Neill, L. A. J., Golenbock, D. & Bowie, A. G. The history of Toll-like receptors - redefining innate immunity. *Nat. Rev. Immunol.* **13**, 453–60 (2013).
 38. Gay, N. J., Symmons, M. F., Gangloff, M. & Bryant, C. E. Assembly and localization of Toll-like receptor signalling complexes. *Nat. Rev. Immunol.* **14**, 546–558 (2014).
 39. Takeuchi, O. *et al.* Discrimination of bacterial lipoproteins by Toll-like receptor 6. *Int. Immunol.* **13**, 933–940 (2001).
 40. Takeuchi, O. *et al.* Cutting Edge: Role of Toll-Like Receptor 1 in Mediating Immune Response to Microbial Lipoproteins. *J. Immunol.* **169**, 10–14 (2002).
 41. Hayashi, F. *et al.* The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* **410**, 1099–103 (2001).
 42. Zanoni, I. *et al.* CD14 Controls the LPS-Induced Endocytosis of Toll-like Receptor 4. *Cell* **147**, 868–880 (2011).
 43. Hemmi, H. *et al.* A Toll-like receptor recognizes bacterial DNA. *Nature* **408**, 740–5 (2000).
 44. Bauer, S. *et al.* Human TLR9 confers responsiveness to bacterial DNA via species-specific CpG motif recognition. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9237–42 (2001).
 45. Alexopoulou, L., Holt, A. C., Medzhitov, R. & Flavell, R. A. Recognition of double-stranded RNA and activation of NF- κ B by Toll-like receptor 3. *Nature* **413**, 732–8 (2001).
 46. Lund, J. M. *et al.* Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc Natl Acad Sci U S A* **101**, 5598–5603 (2004).
 47. Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S. & Reis e Sousa, C. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* **303**, 1529–31 (2004).
 48. Heil, F. *et al.* Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* **303**, 1526–9 (2004).
 49. Hornung, V. *et al.* Sequence-specific potent induction of IFN- α by short interfering RNA in plasmacytoid dendritic cells through TLR7. *Nat. Med.* **11**, 263–70 (2005).
 50. Pirher, N., Ilicak, K., Pohar, J., Bencina, M. & Jerala, R. A second binding site for double-stranded RNA in TLR3 and consequences for interferon activation. *Nat. Struct. Mol. Biol.* **15**, 761–3 (2008).
 51. Wang, T. *et al.* Toll-like receptor 3 mediates West Nile virus entry into the brain causing lethal encephalitis. *Nat. Med.* **10**, 1366–73 (2004).
 52. Kang, J. Y. & Lee, J.-O. Structural biology of the toll-like receptor family. *Annu. Rev. Biochem.* **80**, 917–941 (2011).
 53. Kang, J. Y. *et al.* Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 hetero-

- dimer. *Immunity* **31**, 873–84 (2009).
54. Jin, M. S. *et al.* Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **130**, 1071–82 (2007).
 55. Kawai, T. & Akira, S. Signaling to NF- κ B by Toll-like receptors. *Trends Mol. Med.* **13**, 460–469 (2007).
 56. Ley, K., Laudanna, C., Cybulsky, M. I. & Nourshargh, S. Getting to the site of inflammation: the leukocyte adhesion cascade updated. *Nat. Rev. Immunol.* **7**, 678–89 (2007).
 57. Fischer, M. & Ehlers, M. Toll-like receptors in autoimmunity. *Ann. N. Y. Acad. Sci.* **1143**, 21–34 (2008).
 58. Mohammad Hosseini, A., Majidi, J., Baradaran, B. & Yousefi, M. Toll-Like Receptors in the Pathogenesis of Autoimmune Diseases. *Adv. Pharm. Bull.* **5**, 605–614 (2015).
 59. Radstake, T. R. D. J. *et al.* Expression of toll-like receptors 2 and 4 in rheumatoid synovial tissue and regulation by proinflammatory cytokines interleukin-12 and interleukin-18 via interferon-gamma. *Arthritis Rheum.* **50**, 3856–65 (2004).
 60. Brentano, F. *et al.* RNA released from necrotic synovial fluid cells activates rheumatoid arthritis synovial fibroblasts via Toll-like receptor 3. *Arthritis Rheum.* **52**, 2656–65 (2005).
 61. Ospelt, C. *et al.* Overexpression of toll-like receptors 3 and 4 in synovial tissue from patients with early rheumatoid arthritis: toll-like receptor expression in early and longstanding arthritis. *Arthritis Rheum.* **58**, 3684–92 (2008).
 62. Sloane, J. A. *et al.* Hyaluronan blocks oligodendrocyte progenitor maturation and remyelination through TLR2. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11555–60 (2010).
 63. Reynolds, J. M., Martinez, G. J., Chung, Y. & Dong, C. Toll-like receptor 4 signaling in T cells promotes autoimmune inflammation. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13064–9 (2012).
 64. Miranda-Hernandez, S. & Baxter, A. G. Role of toll-like receptors in multiple sclerosis. *Am. J. Clin. Exp. Immunol.* **2**, 75–93 (2013).
 65. Marshak-Rothstein, A. Toll-like receptors in systemic autoimmune disease. *Nat. Rev. Immunol.* **6**, 823–35 (2006).
 66. Lafyatis, R. & York, M. Innate immunity and inflammation in systemic sclerosis. *Curr. Opin. Rheumatol.* **21**, 617–22 (2009).
 67. Celhar, T. *et al.* TLR7 and TLR9 in SLE: when sensing self goes wrong. *Immunol. Res.* **53**, 58–77 (2012).
 68. Chen, K. *et al.* Toll-like receptors in inflammation, infection and cancer. *Int. Immunopharmacol.* **7**, 1271–1285 (2007).
 69. Hennessy, E. J., Parker, A. E. & O’Neill, L. A. J. Targeting Toll-like receptors: emerging therapeutics? *Nat. Rev. Drug Discov.* **9**, 293–307 (2010).
 70. Liu, L. *et al.* Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* **320**, 379–81 (2008).
 71. Kim, H. M. *et al.* Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. *Cell* **130**, 906–17 (2007).
 72. Yoon, S. *et al.* Structural basis of TLR5-flagellin recognition and signaling. *Science* **335**, 859–64 (2012).
 73. Zhang, Z. *et al.* Structural Analysis Reveals that Toll-like Receptor 7 Is a Dual Receptor for Guanosine and Single-Stranded RNA. *Immunity* **45**, 737–748 (2016).
 74. Tanji, H. *et al.* Toll-like receptor 8 senses degradation products of single-stranded RNA. *Nat. Struct. Mol. Biol.* **22**, 109–15 (2015).
 75. Tanji, H., Ohto, U., Shibata, T., Miyake, K. & Shimizu, T. Structural reorganization of the Toll-like receptor 8 dimer induced by agonistic ligands. *Science* **339**, 1426–9 (2013).
 76. Ohto, U. *et al.* Structural basis of CpG and inhibitory DNA recognition by Toll-like receptor 9. *Nature* **520**, 702–5 (2015).
 77. Jiang, S., Li, X., Hess, N. J., Guan, Y. & Tapping, R. I. TLR10 Is a Negative Regulator of Both MyD88-Dependent and -Independent TLR Signaling. *J. Immunol.* **196**, 3834–41 (2016).
 78. Hemmi, H. *et al.* Small anti-viral compounds activate immune cells via the TLR7 MyD88-dependent signaling pathway. *Nat. Immunol.* **3**, 196–200 (2002).
 79. van Belkum, A. *et al.* Co-evolutionary aspects of human colonisation and infection by *Staphylococcus aureus*. *Infect. Genet. Evol.* **9**, 32–47 (2009).
 80. Fernández Guerrero, M. L., González López, J. J., Goyenechea, A., Fraile, J. & de Górgolas, M. Endocarditis caused by *Staphylococcus aureus*: A reappraisal of the epidemiologic, clinical, and pathologic manifestations with analysis of factors determining outcome. *Medicine (Baltimore)*. **88**, 1–22 (2009).
 81. Bone, R. C. Gram-positive organisms and sepsis. *Arch. Intern. Med.* **154**, 26–34 (1994).
 82. David, M. Z. & Daum, R. S. Community-associated methicillin-resistant *Staphylococcus aureus*: Epidemiology and clinical consequences of an emerging epidemic. *Clin. Microbiol. Rev.* **23**, 616–687 (2010).

83. Kallen, A. J. Health Care–Associated Invasive MRSA Infections, 2005–2008. *JAMA* **304**, 641 (2010).
84. Liu, G. Y. Molecular pathogenesis of *Staphylococcus aureus* infection. *Pediatr. Res.* **65**, 71R–77R (2009).
85. Koymans, K. J., Vrieling, M., Gorham, R. D. & van Strijp, J. A. G. in *Life Science Journal* **6**, 23–27 (2015).
86. Thammavongsa, V., Kim, H. K., Missiakas, D. & Schneewind, O. Staphylococcal manipulation of host immune responses. *Nat. Rev. Microbiol.* **13**, 529–43 (2015).
87. Baker, H. M. *et al.* Crystal structures of the staphylococcal toxin SSL5 in complex with sialyl Lewis X reveal a conserved binding site that shares common features with viral and bacterial sialic acid binding proteins. *J. Mol. Biol.* **374**, 1298–308 (2007).
88. Munro, J. M. *et al.* Expression of sialyl-Lewis X, an E-selectin ligand, in inflammation, immune processes, and lymphoid tissues. *Am. J. Pathol.* **141**, 1397–1408 (1992).
89. Economou, A. Bacterial secretome: the assembly manual and operating instructions (Review). *Mol. Membr. Biol.* **19**, 159–169 (2002).
90. Koymans, K. J. *et al.* Staphylococcal superantigen-like protein 1 and 5 (SSL1 & SSL5) limit neutrophil chemotaxis and migration through MMP-inhibition. *Int. J. Mol. Sci.* **17**, 1–16 (2016).
91. Bardeel, B. W. *et al.* Evasion of Toll-like receptor 2 activation by staphylococcal superantigen-like protein 3. *J. Mol. Med.* **90**, 1109–1120 (2012).
92. Yokoyama, R. *et al.* Staphylococcal superantigen-like protein 3 binds to the Toll-like receptor 2 extracellular domain and inhibits cytokine production induced by *Staphylococcus aureus*, cell wall component, or lipopeptides in murine macrophages. *Infect. Immun.* **80**, 2816–25 (2012).
93. Bestebroer, J. *et al.* Staphylococcal superantigen-like 5 binds PSGL-1 and inhibits P-selectin-mediated neutrophil rolling. *Blood* **109**, 2936–43 (2007).
94. Bestebroer, J. *et al.* Staphylococcal SSL5 inhibits leukocyte activation by chemokines and anaphylatoxins. *Blood* **113**, 328–337 (2009).
95. Itoh, S. *et al.* Staphylococcal superantigen-like protein 5 inhibits matrix metalloproteinase 9 from human neutrophils. *Infect. Immun.* **78**, 3298–3305 (2010).
96. de Haas, C. J. C. *et al.* Staphylococcal superantigen-like 5 activates platelets and supports platelet adhesion under flow conditions, which involves glycoprotein Ibalpha and alpha IIb beta 3. *J. Thromb. Haemost.* **7**, 1867–74 (2009).
97. Hu, H. *et al.* GPVI and GPIb α mediate staphylococcal superantigen-like protein 5 (SSL5) induced platelet activation and direct toward glycans as potential inhibitors. *PLoS One* **6**, e19190 (2011).
98. Fevre, C. *et al.* *Staphylococcus aureus* proteins SSL6 and SEIX interact with neutrophil receptors as identified using secretome phage display. *Cell. Microbiol.* **16**, 1646–1665 (2014).
99. Ramsland, P. A. *et al.* Structural basis for evasion of IgA immunity by *Staphylococcus aureus* revealed in the complex of SSL7 with Fc of human IgA1. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15051–6 (2007).
100. Langley, R. *et al.* The staphylococcal superantigen-like protein 7 binds IgA and complement C5 and inhibits IgA-Fc alpha RI binding and serum killing of bacteria. *J. Immunol.* **174**, 2926–33 (2005).
101. Laursen, N. S. *et al.* Structural basis for inhibition of complement C5 by the SSL7 protein from *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3681–6 (2010).
102. Bestebroer, J. *et al.* Functional basis for complement evasion by staphylococcal superantigen-like 7. *Cell. Microbiol.* **12**, 1506–1516 (2010).
103. Itoh, S. *et al.* Staphylococcal superantigen-like protein 8 (SSL8) binds to tenascin C and inhibits tenascin C–fibronectin interaction and cell motility of keratinocytes. *Biochem. Biophys. Res. Commun.* **433**, 127–132 (2013).
104. Patel, D., Wines, B. D., Langley, R. J. & Fraser, J. D. Specificity of staphylococcal superantigen-like protein 10 toward the human IgG1 Fc domain. *J. Immunol.* **184**, 6283–92 (2010).
105. Itoh, S. *et al.* Staphylococcal superantigen-like protein 10 (SSL10) binds to human immunoglobulin G (IgG) and inhibits complement activation via the classical pathway. *Mol. Immunol.* **47**, 932–938 (2010).
106. Patel, D. *Functional and Structural Characterization of Staphylococcal Superantigen-Like Protein 10 (SSL10)*. Thesis, (The University of Auckland (Thesis), 2011).
107. Walenkamp, A. M. E. *et al.* Staphylococcal superantigen-like 10 inhibits CXCL12-induced human tumor cell migration. *Neoplasia* **11**, 333–44 (2009).
108. Itoh, S. *et al.* Staphylococcal superantigen-like protein 10 (SSL10) inhibits blood coagulation by binding to prothrombin and factor Xa via their γ -carboxyglutamic acid (Gla) domain. *J. Biol. Chem.* **288**, 21569–80 (2013).
109. Chung, M. C. *et al.* The crystal structure of staphylococcal superantigen-like protein 11 in complex with sialyl Lewis X reveals the mechanism for cell binding and immune inhibition. *Mol. Microbiol.* **66**, 1342–55 (2007).

2



Structural insights into collagen-binding by platelet receptor GPVI

Louris J. Feitsma^a, T. Harma C. Brondijk^a, Gavin Jarvis^b, Dominique Hagemans^a, Dominique Bihan^b, Natasia Jerah^b, Marian Versteeg^a, Richard W. Farndale^b, and Eric G. Huizinga^a

^a Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Department of Chemistry, Faculty of Science, Utrecht University, Utrecht, The Netherlands

^b Department of Biochemistry, University of Cambridge, United Kingdom

Significance statement

Thrombus formation is an important mechanism in hemostasis at sites of vascular injury. Platelet receptor glycoprotein VI (GPVI) supports activation by binding to exposed collagen. Crystal structures of GPVI bound to collagen peptides now reveal its primary collagen binding site across the D1-domain β -sheet. GPVI binds sites in collagen formed by two of the triple helix chains and canonical OGPOGP sequence motifs. These novel insights may contribute to the development of novel anti-thrombotic drugs to reduce thrombus deposition.

Keywords: Blood; Platelets; GPVI; collagen; hemostasis; thrombosis.

Authorship contribution

L.J.F., T.H.C.B., G.J., D.H., D.B., N.J., and M.V. performed research; L.J.F., T.H.C.B., G.J., R.W.F., and E.G.H. designed experiments; L.J.F., R.W.F., and E.G.H. wrote the manuscript; R.W.F., and E.G.H. performed supervision.

Data deposition: The atomic coordinates and structure factors for GPVI Δ PAVS-PAPYKN, the GPVI-(GPO)₅ complex and for the GPVI-(GPO)₃ complex have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID codes 5OU7, 5OU8 and 5OU9).

Glycoprotein VI (GPVI) plays a vital role in normal hemostasis by mediating collagen-induced platelet activation after vascular damage, but also contributes to the onset of thrombosis, heart attack and stroke. Animal models of thrombosis have identified GPVI as a promising target for antithrombotic therapy. Although for many years the crystal structure of GPVI has been known, the essential details of its interaction with collagen have remained elusive. Here, we present crystal structures of the GPVI ectodomain bound to triple-helical collagen peptides, which reveal a collagen-binding site across the β -sheet of the D1-domain. Mutagenesis and binding studies confirm the observed binding site and identify Trp76, Arg38 and Glu40 as essential residues for binding to fibrillar collagen I and collagen-related peptide (CRP). GPVI binds a site on collagen comprising two collagen chains. The core of this site is formed by the sequence motif OGPOGP. Potent GPVI-binding peptides from Toolkit III all contain OGPOGP-motifs, but some motifs arise from peptide design and are not present in collagen III itself. Alanine-scanning of peptide III-30 showed the sequence motif AGPOGP also contributes to GPVI-binding. We further show that the D2-dimer observed in crystals of GPVI is unlikely to be functionally relevant, being unable to mediate high-affinity binding to parallel-oriented collagen-helices in a fibril. Our structure confirms the previously suggested similarity in collagen-binding between GPVI and leukocyte-associated immunoglobulin-like receptor 1 (LAIR-1), but also indicate significant differences that may be exploited for the development of GPVI-specific therapeutics.

Introduction

Thrombus formation at sites of vascular injury is initiated by collagen. In response to vessel damage, platelet receptors GPIb α and integrin $\alpha_2\beta_1$ support firm platelet adhesion to collagens of the subendothelial extracellular matrix (ECM): GPIb α associates with collagen-bound von Willebrand factor (VWF) whereas the integrin $\alpha_2\beta_1$ binds directly to exposed collagen^{1,2}. Subsequent interaction between collagen and the Glycoprotein VI - Fc-receptor γ -chain (GPVI-FcR γ) complex³⁻⁵, triggers phosphorylation of FcR γ and ultimately results in full platelet activation⁶⁻⁹. GPVI itself is present in both monomeric and dimeric form in the platelet membrane, and activation by several ligands shifts the population towards the dimeric species¹⁰⁻¹². Deficiency or inhibition of GPVI results in decreased thrombus deposition and platelet aggregation¹³, emphasizing its role in platelet activation.

Fibrillar collagens provide strength to the arterial vasculature, whilst the role of exposed collagens in platelet adhesion and activation reflects its important hemostatic function¹⁴. The triple-helical structure of collagen requires the repeating Gly-X-Y sequence, where, most frequently, X is proline (P) and Y is the post-translationally modified 4-hydroxyproline (O). Toolkits-II and -III containing homotrimeric peptides that cover the full sequences of the collagen II and III triple helical regions, respectively, were used to identify collagenous binding motifs for several proteins¹⁵: integrin $\alpha_2\beta_1$ ¹⁶, ECM protein SPARC^{17,18}, discoidin domain receptor 2¹⁹, the VWF A3-domain^{20,21}, and recently the Osteoclast-associated receptor (OSCAR)²². The use of synthetic peptides including Toolkit-III identified peptides III-30, III-01, III-40 and collagen-related peptide (CRP)

inducing GPVI-mediated platelet activation, and revealed a strong relationship between (hydroxy)proline-content and binding of soluble GPVI²³. The importance of collagen peptide (hydroxy)proline content for GPVI-binding is also underlined by the increasing binding capacity of peptides containing consecutive GPO-repeats, ranging from (GPO)₂ to (GPO)₁₀; the latter being the collagenous domain in the CRP-peptide^{24,25}. These studies, however, could not unambiguously characterize a GPVI-binding motif^{11,23}.

Gpvi is located in chromosome 19 in the leukocyte receptor complex (LRC)⁵ alongside many other immune receptor genes including the structurally-related but functionally separate collagen-binding receptors, OSCAR²⁶ and Leukocyte-Associated Immunoglobulin-like Receptors 1 and 2 (LAIR-1 and -2)²⁷. Activatory receptors GPVI and OSCAR share the same extracellular domain architecture, having two C2-type Ig-like domains (D1 and D2). Collagen binds both domains of OSCAR²², with D2 having higher affinity, whereas modeling and limited site-directed mutagenesis suggested that collagen binds GPVI in a groove located between β -strands C' and E of the D1-domain^{28,29}. LAIRs are immune regulating receptors and have only one extracellular Ig-like domain (D1) that contains a collagen-binding site across the β -sheet formed by strands FCC'³⁰. A few LAIR-1-binding Toolkit-peptides also bind either GPVI or OSCAR, but not both, suggesting similarities in collagen-binding mechanism but discrete identities for each receptor^{11,23,26,31}.

Here, we present crystal structures of GPVI in complex with collagen-like peptides that provide the structural basis for collagen binding to the canonical GPO sequence of collagen and which largely account for the propensity of GPVI to bind diverse sites in collagen.

Results

Crystal structure determination of GPVI collagen-peptide complexes

To obtain crystals of GPVI-collagen peptide complexes, it proved essential to truncate two loops in GPVI D2 that are flexible, judged from the published crystal structure (PDB ID 2GI7; Figure S2a)²⁸. From eight mutants tested, we selected Δ PAVS- Δ PAPYKN that lacks residues 102-105 and 131-136 (Figure S2b-d), which showed least decrease in expression yield (20%) and greatest increase in thermal stability (2.5 degrees). Its collagen-binding affinity is reduced about 5-fold, measured using dimeric Fc-fusions of the mutant and wild-type protein, as in Methods.

We obtained structures of the GPVI Δ PAVS- Δ PAPYKN mutant alone (1.9 Å resolution; Figure S3) and in complex with collagen peptides (GPO)₅ and (GPO)₇, respectively (both 2.5 Å resolution, Figure 1a-b; Table S2). Despite the different lengths and hydroxyproline-content of the peptides (Table 1), both complexes exhibit the same crystal form: an assembly of two GPVI molecules that bind a single triple-helical peptide. The two GPVI molecules do not contact each other and bind the peptide with the same region of D1. D2-domains are not involved; the truncated loops are located so that even the longer loops in the wild-type protein could not reach the collagen peptide. Therefore, the mildly reduced collagen-binding of the mutant used for crystallization cannot be explained from the crystal structures.

The GPVI binding site on collagen peptides

A GPVI binding site comprises two of the three collagen chains and has an area of about 470 Å². In the GPVI-(GPO)₅ complex, one GPVI molecule binds near the peptide N-terminus to a continuous surface comprised of Pro2-Hyp6 and Pro2-Hyp9 residues of the middle (*M*) and trailing (*T*) chain, respectively; the other molecule binds to an equivalent site close to the C-terminus comprising residues Pro8-Hyp12 of the leading (*L*) chain and Pro8-Hyp15 of the middle chain (Figure 1c). In the GPVI-(GPO)₃ complex, one GPVI-molecule binds to the L+M chains; a second molecule to the T+L chains (Figure 1d).

Interestingly, the latter binding site has a proline originating from the C-terminal (GPP)₂-extension of the peptide (see Table 1 and Figure 1d) at a position that is occupied by hydroxyproline in the other (GPO)₃- and both (GPO)₅-binding sites. In these sites the hydroxyl-group does not contact GPVI, and therefore the presence of proline at this position likely does not affect binding. Due to the peculiarities arising from chain stagger, binding sites formed by L+M and M+T chain combinations can have no residue other than hydroxyproline at this position, because – in the context of the amino acid sequence – it is equivalent to the Hyp-residue of the central Pro-Hyp patch (underlined O in POGOGP). In the POGOGPPGP sequence that forms the observed T+L binding site of the GPVI-(GPO)₃ complex, the two positions are not equivalent and can be occupied by different amino acids. In native collagen I, an α₁-α₂ heterotrimer, sites may also be constituted by a combination of collagen chains with differing amino acid sequences, allowing more possible alterations in the POGOGP-sequence.

Table 1. Collagen peptide sequences

Peptide	Sequence	T _m (°C)
(GPO) ₅	(GPO) ₅ -NH ₂	28.2
(GPO) ₃	(GPP) ₂ (GPO) ₃ (GPP) ₂ -NH ₂	41.1
CRP	GCO(GPO) ₁₀ GCOG-NH ₂	82.3
III30	GPC(GPP) ₅ GAOGLRGGAGPOGPEGGKAAGPOGPO(GPP) ₅ GPC-NH ₂	47.4
III30-A3	GPC(GPP) ₅ GA <u>A</u> GLRGGAGPOGPEGGKAAGPOGPO(GPP) ₅ GPC-NH ₂	39.0
III30-A5	GPC(GPP) ₅ GAO <u>G</u> ARGGAGPOGPEGGKAAGPOGPO(GPP) ₅ GPC-NH ₂	46.9
III30-A6	GPC(GPP) ₅ GAOGL <u>A</u> GGAGPOGPEGGKAAGPOGPO(GPP) ₅ GPC-NH ₂	44.9*
III30-A8	GPC(GPP) ₅ GAOGLR <u>G</u> AGPOGPEGGKAAGPOGPO(GPP) ₅ GPC-NH ₂	44.8*
III30-A11	GPC(GPP) ₅ GAOGLRGGAG <u>A</u> OPEGGKAAGPOGPO(GPP) ₅ GPC-NH ₂	41.3
III30-A12	GPC(GPP) ₅ GAOGLRGGAG <u>P</u> APEGGKAAGPOGPO(GPP) ₅ GPC-NH ₂	42.2
III30-A14	GPC(GPP) ₅ GAOGLRGGAGPOG <u>A</u> EGGKAAGPOGPO(GPP) ₅ GPC-NH ₂	43.3
III30-A15	GPC(GPP) ₅ GAOGLRGGAGPOG <u>P</u> AGGKAAGPOGPO(GPP) ₅ GPC-NH ₂	47.4
III30-A17	GPC(GPP) ₅ GAOGLRGGAGPOGPE <u>G</u> AKGAAGPOGPO(GPP) ₅ GPC-NH ₂	44.8*
III30-A18	GPC(GPP) ₅ GAOGLRGGAGPOGPEGG <u>A</u> GAAGPOGPO(GPP) ₅ GPC-NH ₂	45.5
III30-A23	GPC(GPP) ₅ GAOGLRGGAGPOGPEGGKAAG <u>A</u> OGP(O)(GPP) ₅ GPC-NH ₂	45.0
III30-A24	GPC(GPP) ₅ GAOGLRGGAGPOGPEGGKAAG <u>P</u> AGPO(GPP) ₅ GPC-NH ₂	44.4
III30-A26	GPC(GPP) ₅ GAOGLRGGAGPOGPEGGKAAGPOG <u>A</u> O(GPP) ₅ GPC-NH ₂	47.2
III30-A27	GPC(GPP) ₅ GAOGLRGGAGPOGPEGGKAAGPOG <u>P</u> A(GPP) ₅ GPC-NH ₂	45.3
III30-A24-27	GPC(GPP) ₅ GAOGLRGGAGPOGPEGGKAAG <u>P</u> AG <u>P</u> A(GPP) ₅ GPC-NH ₂	40.8*

* Melting temperatures were estimated.

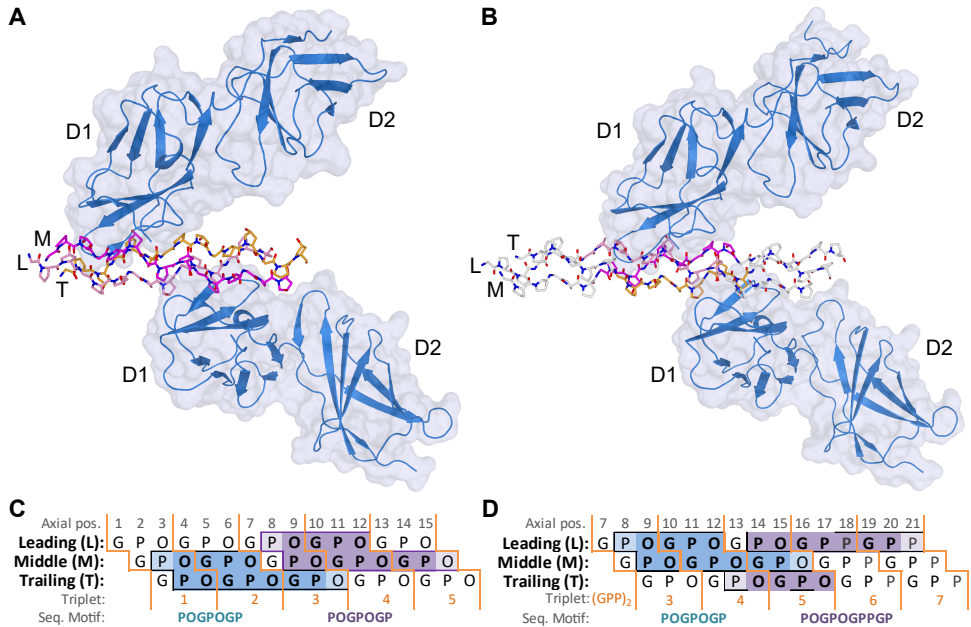


Figure 1: Crystal structures of the GPVI-collagen peptide complexes. (A-B) GPVI-(GPO)₅ (A) and GPVI-(GPO)₃ (B) complexes showing two GPVI-molecules (blue cartoon and surface representation) that bind the collagen peptide with their D1-domains. Chains of the collagen triple helix are shown in pink (Leading; L), magenta (Middle; M), and orange (Trailing; T) sticks. Two GPP-repeats at either end of the (GPO)₃-peptide are shown in white sticks. (C-D) Schematic representation of the (GPO)₅ (C) and (GPO)₃ (D) peptides; residues that are located within 4.5 Å of the first and second GPVI-molecule are shown in blue and purple, respectively. Shown in a lighter shade are residues at the extremes of the binding sites that contribute minimally to GPVI-binding, as evaluated using the FoldX-suite³² (see also Figure S4).

A crucial element of the GPVI-binding site as judged from the extent of interactions and the binding energy per residue calculated using the FoldX-suite³² is a centrally located Pro-Hyp patch (Figure S4). Contributing only marginally to binding are proline and hydroxyproline at both extremes of the binding site (light colors in Figure 1c-d). If these two residues are not included, formation of the core GPVI binding site in a homotrimeric collagen peptide requires the amino acid sequence POGPOGP for the L+M and M+T combinations of collagen chains. Due to the difference in chain stagger, a binding site formed by a T+L chain combination would require the longer sequence POGPOGPOGP, or, as is the case in the T+L binding site of the GPVI-(GPO)₃ complex, the sequence POGPOGPPGP.

The collagen binding site of GPVI

The collagen-binding site of GPVI is located across the D1 β-sheet, comprising strands C, D, F, and F' (Figure 2a). Its position differs from the collagen-binding site predicted from the crystal structure of GPVI alone²⁸, but is consistent with mapping of the collagen-binding site in LAIR1³⁰. Eight central residues are conserved in GPVI orthologues (Figure S5a). Among them, Trp76 forms extensive hydrophobic interactions with T:Pro5⁵ and a hydrogen bond with the backbone carbonyl of T:Hyp3 (Figure 2b). Arg38 and Arg67 form hydrogen bonds with carbonyl oxygens of the (GPO)₅ backbone. Gln71 and Glu40

⁵ We used the one-letter prefixes L, M, and T to refer to amino acids in the leading, middle, and trailing chain of (GPO)₅ and (GPO)₃, respectively.

are hydrogen bonded to the hydroxyl groups of *M*:Hyp3 and *M*:Hyp6, respectively (two times O in POGPQGP), and appear definite for the hydroxyproline-specificity of collagen substrates. The Tyr47 side chain is positioned close to *M*:Pro5 and *M*:Hyp6 and buries a substantial hydrophobic surface. Ser69 and Leu42 have hydrophobic interactions with *T*:Pro5 and *T*:Pro8, respectively. Finally, four non-conserved residues, Leu36, Asp49, Ser74, and Gln82, provide minor interactions at the periphery of the binding interface.

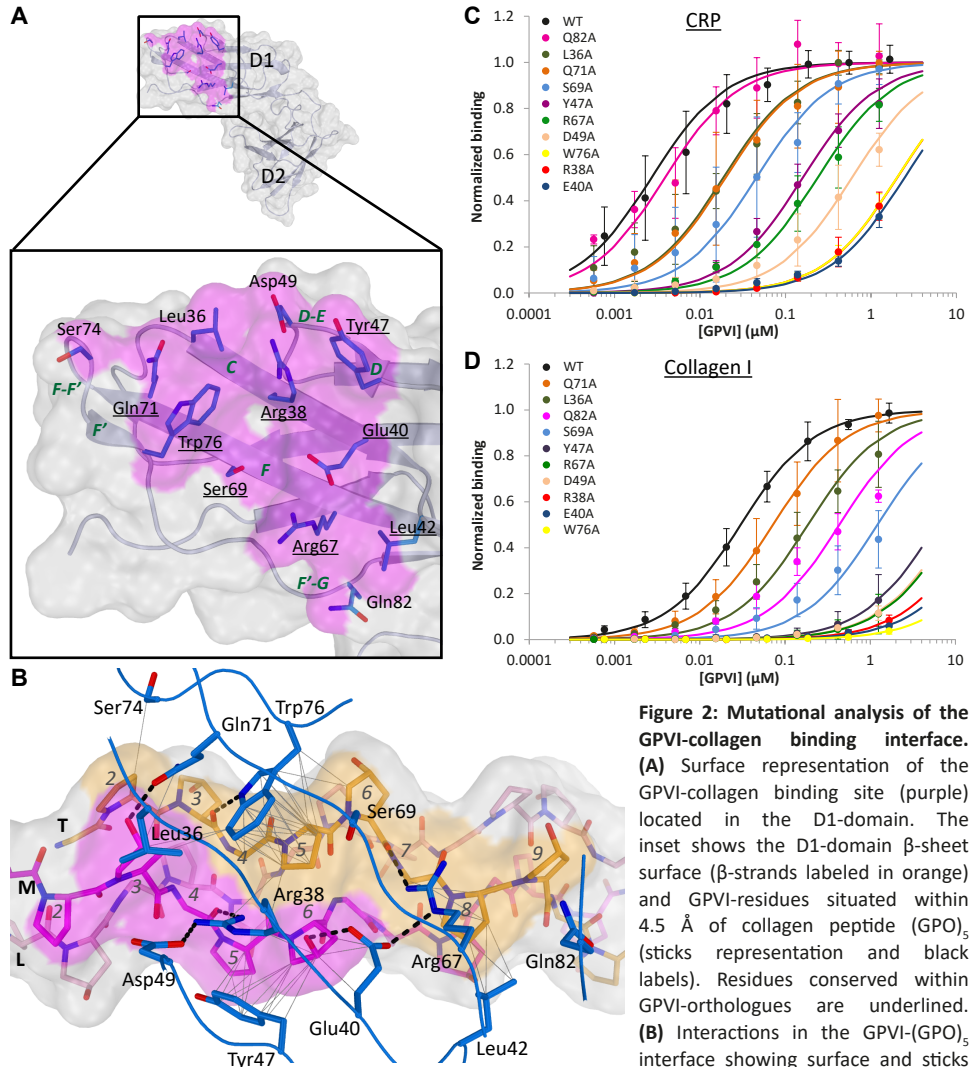


Figure 2: Mutational analysis of the GPVI-collagen binding interface. (A) Surface representation of the GPVI-collagen binding site (purple) located in the D1-domain. The inset shows the D1-domain β -sheet surface (β -strands labeled in orange) and GPVI-residues situated within 4.5 Å of collagen peptide ($(GPO)_5$) (sticks representation and black labels). Residues conserved within GPVI-orthologues are underlined. (B) Interactions in the GPVI- $(GPO)_5$ interface showing surface and sticks

representation of $(GPO)_5$, as well as GPVI-residues (blue sticks and black labels) that are located within 4.5 Å of $(GPO)_5$. Highlighted are the surface areas of the middle (*M*; magenta) and trailing (*T*; orange) chains of $(GPO)_5$, respectively, and the corresponding residues are numbered. Hydrogen bonds and salt bridges are shown in dashed lines, intermolecular carbon-carbon interactions within 4.5 Å in straight lines. (C-D) Binding of GPVI-mutants (colored series) to CRP- (C) and Collagen I-binding (D) as measured in solid-state assays. Data measured as A_{490} is normalized to the binding of wild type GPVI (black series). All data points represent the mean \pm SD of at least three independent experiments. Binding curves are fitted to the equation $Abs = (B_{max} \times c) / (K_{d,app} + c)$ using non-linear regression in SigmaPlot, where c is the GPVI concentration (μM).

Effect of GPVI mutations on collagen-binding

In solid state binding assays, wild type GPVI-Fc dimers bind with nanomolar affinity to both fibrillar collagen I ($K_D = 30$ nM) and the synthetic GPVI agonist CRP (Table 1; $K_D = 2.7$ nM). Binding to these ligands is drastically reduced by nearly all alanine-substitutions of conserved collagen-binding residues (Figure 2c-d and Table 2). Profound reductions were observed for mutants of Trp76, Arg38 and Glu40, showing greater than 600-fold increase in K_D relative to wild type protein. Moderate reductions in affinity (60- to 350-fold) were observed for the Arg67, Tyr47, and Asp49 mutants. Asp49 itself has little direct contact with collagen, but, by forming a salt bridge, appears to fix Arg38 optimally for interaction with collagen, thus explaining the substantial effect of Asp49 mutation.

Minor effects on CRP-binding were observed for mutants of Ser69 (17-fold), Leu36 (7-fold), and Gln82 (1.4-fold), in line with their peripheral location and limited contacts with collagen. Two mutants, Ser69 and Gln82, affected collagen I binding more substantially (40- and 14-fold, respectively), which may be related to the heterogeneity in GPVI-binding sequences in collagen I. Mutation of Gln71 slightly reduced both CRP- and collagen-binding (7- and 2-fold, respectively), indicating that formation of the Gln71-*M*:Hyp3 hydrogen bond contributes relatively little to binding. We also mutated residues of the proposed collagen binding site inside the C'-E groove of D1²⁸. All mutations are neutral to collagen I binding (Figure S6 and Table S3) except Lys59Ala, situated at the D1-surface just outside the C'-E groove. Lys59Ala caused a small decrease in binding (2.5-fold); similar to that previously observed for the Lys59Glu mutant²⁹. We cannot exclude that Lys59 as well as Arg60 and Arg166, two other more remote residues previously implicated by mutagenesis^{29,33,34}, are part of a low-affinity secondary site. Overall, these data confirm that the observed crystal structure of the GPVI-collagen complexes reveals the authentic binding site for collagen I and CRP.

Table 2. Binding of wild type GPVI and mutants to CRP, Collagen-I, and Toolkit-peptide III-30

Mutant	CRP			Fibrillar Collagen I			Toolkit-peptide III-30		
	$K_{D,app}$ (nM)	95% confid. interval (nM)	Fold increase	$K_{D,app}$ (nM)	95% confid. interval (nM)	Fold increase	$K_{D,app}$ (nM)	95% confid. interval (nM)	Fold increase
WT	2.71	2.2 - 3.2		30.1	26 - 34		10.9	9.7 - 12.0	
E40A	2.5×10^3	$2.1 - 2.8 \times 10^3$	913	25×10^3	$22 - 28 \times 10^3$	823	74×10^3	$57 - 91 \times 10^3$	6848
R38A	2.0×10^3	$1.7 - 2.3 \times 10^3$	748	18×10^3	$16 - 20 \times 10^3$	603	42×10^3	$38 - 45 \times 10^3$	3833
W76A	2.0×10^3	$1.8 - 2.2 \times 10^3$	744	43×10^3	$37 - 50 \times 10^3$	1440	21×10^3	$14 - 27 \times 10^3$	1905
D49A	599	454 - 744	221	9.2×10^3	$8.0 - 10 \times 10^3$	306	3.9×10^3	$3.5 - 4.2 \times 10^3$	357
R67A	238	187 - 289	88	10×10^3	$6.6 - 13 \times 10^3$	324	2.0×10^3	$1.9 - 2.2 \times 10^3$	187
Y47A	160	130 - 189	59	6.0×10^3	$4.4 - 7.6 \times 10^3$	199	10×10^3	$8.5 - 12 \times 10^3$	925
S69A	45.7	24 - 67	17	1.2×10^3	$0.9 - 1.5 \times 10^3$	40	6.5×10^3	$5.6 - 7.4 \times 10^3$	602
Q71A	19.9	11 - 29	7	72.2	54.8 - 89.5	2	9.15*	6.87 - 11.4	0.84
L36A	18.8	12 - 25	7	186	150 - 222	6	7.68*	5.77 - 9.58	0.70
Q82A	3.89	2.5 - 5.3	1.4	433	329 - 538	14	289	268 - 310	27

* Maximum III-30-binding of the Q71A and L36A mutants is 109% (104 - 114%) compared to wild type protein.

Comparison of GPVI and LAIR collagen-binding sites

The collagen binding site within the LAIR family is more variable than is the case for GPVI (Figure S5b). The only residues strictly conserved within and between both groups are Trp76, Arg38 and Glu40 (GPVI numbering). As is the case for GPVI, mutation of these residues in LAIR-1 diminishes collagen-binding drastically³⁰ and therefore, the Trp-Arg-Glu triad appears to be the core element in collagen-binding by both receptor-families. Additionally, conserved in human LAIR-1 and -2, but not in other LAIRs, are GPVI-residues Arg67, Tyr47, and Asp49. In line with the extent of sequence divergence, the collagen binding site of human GPVI more closely resembles crystal structures of human LAIR-1 (PDB-ID: 3KGR³⁰) than mouse LAIR-1 (PDB-ID: 4ESK³⁵) (Figure S5c-d). This observation together with the substantial differences in Toolkit and CRP binding properties of both receptors³¹ suggests that residues surrounding the Trp-Arg-Glu triad fine-tune collagen-binding affinity and site-specificity, which may be relevant for the development of receptor-specific therapeutics that minimize cross-reactivity between GPVI and LAIR-1.

Relevance of crystallographic GPVI-dimers for collagen-binding

High-affinity collagen binding has been suggested to be dependent on the formation of GPVI-dimers^{11,28}. All crystal structures presented here and previously²⁸ contain an extended β -sheet dimer formed by back-to-back D2-interactions of GPVI-molecules related by non-crystallographic symmetry (Figure S7). In our GPVI-peptide complexes, the dimer binds different copies of the collagen triple helix (Figure 3a). The inter-helix angle is about 80°, which is not compatible with the nearly parallel orientation of helices in a collagen fibril³⁶. Although we cannot exclude that such a dimer is present at the platelet surface, it appears highly unsuited to high-affinity binding to a collagen fibril.

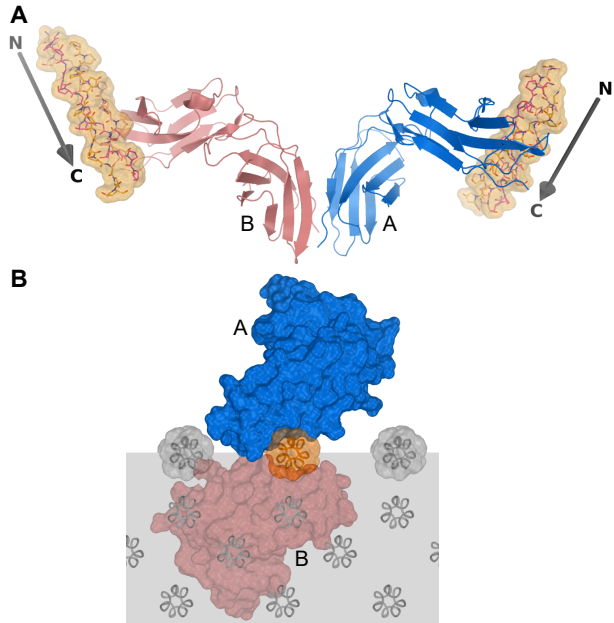
The crystal structures also reveal that two GPVI-monomers can bind in close proximity to a collagen helix without contacting each other (see Figure 1a-b). This binding mode involves nearly the full circumference of the helix (Figure 3b), and is impossible at the fibril surface without substantial displacement of the target helix. Preferred interactions during crystallization gave apparently rise to the formation of GPVI-dimers as well as the binding of two GPVI-molecules close together on one collagen helix. Both arrangements, however, are likely not relevant for binding of GPVI to fibrillar collagen.

GPVI binds to Toolkit-III peptides containing the six-residue OGPOGP-motif

The capacity of Toolkit-III peptides to bind GPVI strongly depends on their (hydroxy) proline-content^{11,23}, but exact sequence motifs are not known. We analyzed the correlation between published peptide-binding data¹¹ and the presence of the full or partial POGPOGP-sequence that forms the GPVI-binding site in our crystals (see *supplementary methods*). This analysis shows that the presence of a six-residue OGPOGP-motif best predicts GPVI-binding (Figure S8a). The correlation for the full seven-residue binding sequence is slightly worse, suggesting that the shorter sequence may be sufficient for GPVI binding. Out of a total of 57 Toolkit-peptides, only 11 contain the OGPOGP-motif, which occasionally overlaps the generic (GPP)₉-region at the C-terminus¹⁶ (Figure S8b). Among the twelve peptides that best support GPVI-binding ten contain the OGPOGP-motif; the 11th, peptide III-10, does not display significant GPVI-binding¹¹, which is unexpected. Seven out of nine remaining peptides that promote GPVI-binding better than GPP, contain a related, but a

Figure 3: Dimeric GPVI-arrangements inside crystals are not compatible with binding to fibrillar collagen.

(A) Back-to-back arrangement of two GPVI-molecules (salmon and blue) observed in crystals of the GPVI-peptide complexes. Each GPVI-molecule binds a different copy of the collagen triple helix (sticks and orange surface). The inter-helix angle is about 80°, and is therefore not compatible with a nearly parallel orientation of helices in a collagen fibril³⁶. **(B)** Superimposition of the GPVI₂-(GPO)₅ complex (coloring as in A) onto a model of the fibril surface³⁶ (gray surface and helices) viewed along the parallel helix axes. Together, the two GPVI-molecules would cover almost the full circumference of the projected helix, whereas only about half of it is exposed at the fibril surface.



slightly different XGPOGP- or OGPOGX-motif, with X being Ala, Pro, Ser, or Ile. In view of the extent of GPVI-collagen interactions observed in our crystals and the effect of GPVI-mutations on CRP-binding (see Figures 2 and S4), we expect that these modifications are compatible with GPVI-binding. Overall, strong GPVI-binding to Toolkit-peptides depends on the presence of an OGPOGP-motif; apparently minor modification at one of the extremes of this motif are tolerated, but result in weaker binding.

Binding to III-30 is mediated by one OGPOGP- and two AGPOGP-sites

Toolkit-peptide III-30 exhibits the highest capacity to bind GPVI-Fc, notably higher than that of other peptides containing one OGPOGP-motif¹¹. In our binding experiments, wild type GPVI-Fc dimers bind III-30 with 11 nM affinity (Figure 4a and Table 2), only about 4-fold weaker than binding to CRP. III-30 binding by GPVI-mutants generally follows a similar trend as binding of the same mutant-set to CRP. Deviating from this trend are mutations of Gln71 and Leu36, which have no effect on binding to III-30, but do affect CRP-binding. On the other hand, the peripheral Gln82Ala, Ser69Ala, and Tyr47Ala mutations reduce III-30 binding more than CRP-binding. Together these small, but significant differences in mutant-binding suggest that minor variations on the OGPOGP-sequence or residues outside the OGPOGP-motif modulate binding strength.

The high GPVI-binding capacity of III-30 could arise from the presence of a second binding site in addition to the site constituted by the OGPOGP-motif. To screen for additional sites in III-30, we analyzed the collagen III-derived region of III-30 by alanine-scanning every non-glycine position (Table 1). Of all single mutants, the Pro11Ala and Hyp12Ala mutations showed the strongest reductions in GPVI-Fc binding, about half of the total binding (Figure 4b). Pro11 and Hyp12 likely form a central Pro-Hyp patch in a binding site constituted by an AGPOGP-motif in III-30 (Figure 4c), which only deviates from the site observed in our crystals by the replacement of one hydroxyproline by alanine.

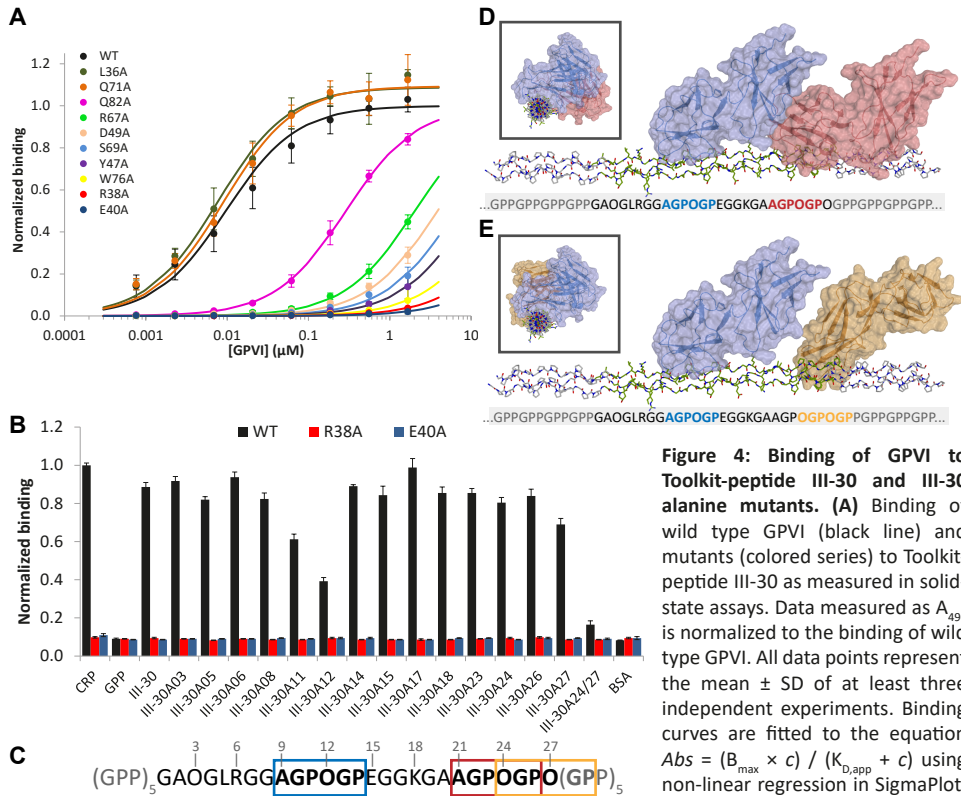


Figure 4: Binding of GPVI to Toolkit-peptide III-30 and III-30 alanine mutants. (A) Binding of wild type GPVI (black line) and mutants (colored series) to Toolkit-peptide III-30 as measured in solid-state assays. Data measured as A_{490} is normalized to the binding of wild type GPVI. All data points represent the mean \pm SD of at least three independent experiments. Binding curves are fitted to the equation $Abs = (B_{\text{max}} \times c) / (K_{\text{D,app}} + c)$ using non-linear regression in SigmaPlot,

where c is the GPVI concentration (μM). (B) Binding of wild type GPVI (blue) and mutants E40A (red) and R38A (green) to CRP, (GPP)₁₀, BSA, Toolkit III-30 and Toolkit III-30 derivatives, obtained by alanine-scanning of all non-glycine residues. Data measured as A_{490} is normalized to binding of wild type GPVI to CRP. All data points represent the mean \pm SD of at least three independent experiments. (C) Amino acid sequence of III-30 with three putative GPVI-binding motifs outlined: two AGPOGP-sites (blue and salmon) and one OGPOGP-site overlapping the generic (GPP)₅-terminus (yellow). (D-E) Binding poses of GPVI onto a 10/3 helix model (PDB-ID: 3DMW; residues 6-17³⁷) of peptide III-30 (see *supplementary methods*). GPVI-molecules were positioned by superimposition onto the two AGPOGP-sites (D) formed by the $L+M$ and $M+T$ chain combinations, respectively; and onto the first AGPOGP-site and the OGPOGP-site (E) formed by the $L+M$ and $T+L$ chain combinations, respectively.

A second AGPOGP-motif in III-30 comprises Ala21-Pro26 and overlaps the OGPOGP-motif (Figure 4c). Mutants of Hyp24, Pro26, and Hyp27 had only minor effects on GPVI-binding, and loss of binding was only observed upon mutation of both Hyp24 and Hyp27 (Figure 4b). The marginal effect of these single mutants suggests that there may be two binding sites that overlap and cannot be occupied by GPVI at the same time. In such a case, mutation of both sites would be required to reduce binding substantially, as is observed. Modeling of GPVI-binding to these sites shows that simultaneous binding of two GPVI-molecules would be possible, but only if the two molecules bind in the same relative orientation as observed in our crystals, covering nearly the full circumference of the helix. We consider it unlikely that such a binding mode is feasible on an immobilized peptide in solid state experiments. By contrast, simultaneous binding to the two AGPOGP-sites or the first AGPOGP-site and the OGPOGP-site is possible on one side of the triple helix without steric hindrance between the GPVI-molecules (Figures 4d-e and S8c). These data show that III-30 contains multiple GPVI binding sites: besides OGPOGP, also AGPOGP can serve as a GPVI-binding site.

Binding of two GPVI-molecules at the fibrillar surface

Peptide III-30 best supports GPVI-binding, followed by III-01 and III-40, but accessibility of the corresponding regions in fibrillar collagen III depends on exposure of helix-segments at the fibril surface (reviewed in Herr & Farndale³⁸). Two AGPOGP-sites in III-30 (the OGPOGP-site is absent in the native collagen III sequence) and two $\alpha_2\beta_1$ -sites³⁸ could be accessible by exposure of helix segment *D3*. Accessibility of the OGPOGP-sites in III-01 and III-40 requires exposure of N-terminal helix segments *D1* and *D4*, respectively. We assessed possible binding modes of GPVI at the fibril surface using 10₃ triple helical models of these Toolkit-peptides³⁷ (see supplementary methods) in combination with the *in situ* structure of collagen I³⁶.

First, binding of GPVI-molecules to the same helix requires a center-to-center distance between two binding sites of at least 32-40 Å, since the large D1-D2 angle causes sterical hindrance if two molecules bind closer together (Figure S9). The distance between the AGPOGP-sites is about 37 Å, which is sufficient for simultaneous binding of GPVI-molecules (Figure 5a). Of note, sterical considerations would allow for binding to the same helix of two $\alpha_2\beta_1$ -molecules of integrin $\alpha_2\beta_1$ to nearby GAOGER- and GMOGER-sites together with two GPVI-molecules (Figure S10). Although GPVI-binding sites in the other peptides are not exactly known, simultaneous binding of two GPVI-molecules on a single triple helix seems also possible on the basis of our model for sites in III-40 (Figure S8b-d), but likely not for III-01. A potential second site in III-01 overlapping the (GPP)₅-terminus is absent in collagen III itself.

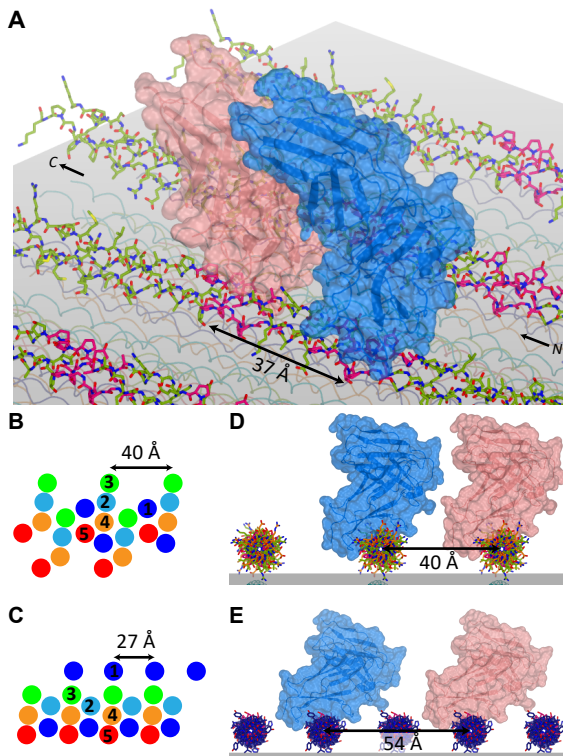


Figure 5: Hypothetical binding modes of GPVI and integrin $\alpha_2\beta_1$ to collagen fibrils. (A) Fibril surface model of collagen type III with parallel collagen helices around the *D3*-segment covered by Toolkit-III peptide 30. Superimposition of GPVI shows that two molecules (blue and salmon) can bind without sterical hindrance to the adjacent AGPOGP-sites. **(B-C)** Molecular packing of fibrillar collagen that would lead to exposure of the *D3*- **(B)** or the *D1*-segment **(C)** at the fibril surface, colored according to Herr *et al.*³⁸. **(D-E)** Cross-sections of the fibrillar surface model near the **(D)** *D3*- and **(E)** *D1*-segments showing parallel helices separated by respectively 40 and 27 Å. Two GPVI-molecules (blue and salmon) can bind adjacent III-30 sites at 40 Å distance without sterical hindrance, or adjacent III-01 sites at twice the helix-helix distance, 54 Å.

Second, two GPVI-molecules could bind to adjacent collagen helices. The complex architecture of the collagen I fibril³⁶ would result in exposure of parallel *D3*-periods in adjacent triple helices positioned 40 Å apart, or parallel *D1*-periods positioned 27 Å apart (Figure 5b-c)³⁸. The distance of 40 Å between equivalent *D3*-segments is sufficient for GPVI-binding to the same AGPOGP-sites of III-30 in directly neighboring helices (Figure 5d). Opposite to that, the 27 Å distance between *D1*-periods would be too short, requiring twice the helix distance (54 Å) for GPVI-binding to III-01 regions without steric hindrance (Figure 5e). All these models indicate possible options for binding of a second GPVI-molecule closely to the first monomer at the fibril surface, and may represent the mode of binding by platelet-associated GPVI-dimers at sites of vascular injury.

Discussion

Platelet receptor GPVI has a central role in hemostasis by binding to exposed collagens at sites of vascular injury. Our crystal structures of GPVI-collagen peptide complexes reveal for the first time the structural basis for receptor binding to the canonical collagen sequence (GPO)_n. As noted previously, all collagen binding domains on which structural data were available thus far engage a long side chain of a collagen residue: phenylalanine (VWF-A3²¹, SPARC¹⁸, and OSCAR²²), glutamate ($\alpha_2\beta_1$ ³⁹), or arginine (Hsp47⁴⁰). By contrast, our structures reveals a different collagen-binding mode in which GPVI associates with several modestly sized (hydroxy)proline residues in collagen, employing a rather flat surface across the *D1* β -sheet.

The role of GPVI dimers for high-affinity binding

Two studies indicate that formation of GPVI-dimers at the platelet surface does not require the engagement of collagen by GPVI or the presence of any other GPVI agonist^{11,12}. In effect, dimerization at the platelet surface co-localizes the collagen-binding domains and *O*-glycosylated stalks of two GPVI molecules; a situation that, to first approximation, can be mimicked by fusion of a GPVI extracellular fragment to a dimer-forming Fc domain. Compared to monomeric GPVI, such Fc-tagged GPVI-dimers are observed to bind 100- to 900-fold stronger to fibrillar collagen and collagen-like peptides *in vitro*^{11,41}. It has been suggested that the observed increase in binding affinity upon dimerization is too large to be explained by avidity effects alone and would require cooperative GPVI-binding arising from direct interactions between *D1D2* fragments within the dimer.

Our study unequivocally shows that the dimer observed in all crystal forms of GPVI obtained thus far²⁸ is not the species previously proposed to bind with high-affinity to fibrillar collagen because it cannot bind two parallel collagen helices (see Figure 3). Possibly, the observed dimer arrangement represents the conformation of GPVI dimers at the platelet surface in the absence of collagen. However, in view of the weakness of the interaction, which we do not even observe at high GPVI-concentration in solution, it appears unlikely that it is the driving force for dimer formation. The observation in our crystals of two GPVI molecules binding to one collagen triple-helix without any direct GPVI-GPVI contacts in itself argues against strong cooperative binding of GPVI, at least to a single collagen helix.

We did model binding of two GPVI-molecules to nearby sites on the same or two parallel helices at the surface of fibrillar collagen (Figure 5). The distances of 37-55 Å

between these sites seem sufficiently short to be bridged by a GPVI-Fc dimer or GPVI dimers at the platelet surface. Uncertainties in the precise local collagen conformation and the extent of mobility that exists at the fibril surface, however, prevents us from concluding whether binding of GPVI monomers to nearby sites would give rise to favorable GPVI-GPVI interactions, and thus cooperative binding. Of note, binding curves obtained in studies employing a broad concentration range of both monomeric and dimeric GPVI¹¹ do not suggest cooperative binding. Moreover, the observed increase in binding affinities of 100-900 fold does fall within the range of 5-to 1000-fold improved binding reported for the dimerization of Fab fragments by several different means or the conversion of single chain Fv's to IgG's (see recent review⁴²). Therefore, although cooperative binding arising from direct interactions between GPVI D1D2-domains cannot be excluded, high-affinity binding of GPVI dimers to collagen may arise exclusively from avidity effects.

Methods

Expression and purification of GPVI D1D2 WT and mutants

For crystallization, GPVI loop truncation mutants were made in a construct encoding D1 and D2 (Uniprot ID Q9HCN6-3; Gln21-Ser206), by QuikChange Site-directed mutagenesis and primers listed in Table S1. GPVI was expressed with N-terminal Cystatin-S signal peptide (Uniprot ID P01036; Met1-Ala20) followed by a TEV-cleavable His₆-StrepII₃-tag in HEK293-EBNA1-S cells (U-Protein Express). Proteins were purified on a StrepTactin-column (GE Healthcare) and eluted with a buffer containing 10 mM HEPES pH 7.5, 150 mM NaCl and 5 mM *d*-desthiobiotin. Purification tags were removed by overnight incubation with 1:30 mol/mol of His₆-TEV-protease/GPVI in the same buffer. Uncleaved protein and TEV-protease were removed by IMAC. Proteins were purified further by size-exclusion chromatography on a Superdex75 column (GE Healthcare) equilibrated in 10 mM HEPES pH 7.5, 150 mM NaCl and concentrated to 10 mg/mL.

For binding studies, the wild-type GPVI fragment described above and point mutants generated by QuikChange Site-directed mutagenesis were transiently expressed with N-terminal Cystatin-S signal peptide and a C-terminal TEV-cleavable immunoglobulin Fc-tag (Uniprot ID P01857; Glu99-Lys330) in HEK293-EBNA1 cells (U-Protein Express). GPVI-Fc dimers were purified by ProteinA-affinity chromatography (GE Healthcare), eluted with 0.1 M glycine pH 3.0; fractions were collected into 1/10 volume of neutralization buffer (1 M Tris-HCl pH 9.0). Protein purity was determined as >95% by SDS-PAGE. Proteins were stored in neutralized elution buffer at 277 K, or, for long-term storage, at 193 K after flash-freezing in liquid nitrogen.

Thermostability assay

GPVI loop truncation mutants were 3- to 10-fold diluted in PBS to end-concentrations of 0.25 mg/mL. 18.75 μ L of diluted protein solution was mixed with 6.25 μ L 40x Sypro Orange (Sigma-Aldrich) in a 96-well plate (Greiner 652201). Thermal-unfolding curves were measured as change in fluorescence as described⁴³. Sealed plates were heated from 289 to 368K with increments of 0.5K in an iCycler iQ Real Time PCR Detection System (Bio-Rad) and fluorescence was measured. Melting temperatures (T_m) were defined as the inflection point of the unfolding curves.

Peptide synthesis and melting point determination

Crystallization peptides (GPO)₅, (GPO)₃, Toolkit peptides III-30 and derivatives, and CRP were synthesized in an Applied Biosystems Pioneer solid-phase synthesizer, using Fmoc-chemistry as previously described^{16,23}. Peptides were purified by HPLC and verified using MALDI-TOF mass spectrometry. T_m was determined by polarimetry, confirming triple-helical conformation (Table 1).

Crystallization and data collection of GPVI alone and in complex with collagen peptides

Before crystallization, GPVI loop truncation mutant ΔPAVS-PAPYKN (10 mg/mL in gel filtration buffer) was deglycosylated overnight at 292 K by the addition 1:100 (v/v) EndoH_F (NEB; 1 × 10⁶ unit/mL). GPVI crystals were obtained in approximately one month by sitting-drop vapor diffusion with a well solution containing 0.1 M phosphate-citrate buffer pH 4.0 and 40% (v/v) PEG-300 at 292 K. Crystals were flash-frozen in liquid nitrogen without further cryo-protection. Diffraction data to 1.9 Å resolution were collected on the PX-beamline at the Swiss Light Source.

For co-crystallization of GPVI with collagen peptides, either (GPO)₅ or (GPO)₃ (Table 1), protein and trimeric peptide were mixed in a 1:1.2 molar ratio. Crystals of the GPVI-(GPO)₅ complex were grown at 292 K with crystallization buffer containing 0.1 M MES pH 6.0, 0.2 M MgCl₂, and 20% (w/v) PEG-6000. Crystals of the GPVI-(GPO)₃ complex were grown with buffer containing 0.1M MMT pH 9.0 and 25% (w/v) PEG-1500. Crystals were cryoprotected in crystallization buffer containing 20% (v/v) glycerol before flash-freezing in liquid nitrogen. Diffraction data were collected on the ID23-1 beamline at the European Synchrotron Radiation Facility (ESRF), with resolution limits of 2.5 Å and 2.7 Å for the GPVI-(GPO)₅ and GPVI-(GPO)₃ complex, respectively. Details about data processing and structure solving are provided in the supplementary methods.

Collagen-binding assay

We performed enzyme-linked immunosorbent assays (ELISA) as previously described²³ to measure the binding of dimeric GPVI-Fc mutants to human placental collagen type I fibrils, collagen-related peptide (CRP), Toolkit-peptide III-30, and III-30 derivatives. Additional details about measuring and data processing are provided in the supplementary methods.

Acknowledgements

We thank the Swiss Light Source (SLS), Villigen (CH) and European Synchrotron Radiation Facility (ESRF), Grenoble (FR) for providing data collection facilities and the beamline scientists for their help with data collection. This work was supported in part by ECHO grant 700.58.006 from the Council of Chemical Sciences of the Netherlands Organization for Scientific Research to E.G.H. and Medical Research Council project grant G0400701 and Wellcome Trust grant 068724/Z/02/Z to R.W.F.

References

1. Clemetson, K. J. & Clemetson, J. M. Platelet collagen receptors. *Thromb. Haemost.* **86**, 189–97 (2001).
2. Ruggeri, Z. M. Platelets in atherothrombosis. *Nat. Med.* **8**, 1227–1234 (2002).
3. Tsuji, M., Ezumi, Y., Arai, M. & Takayama, H. A novel association of Fc receptor γ-chain with glycopro-

- tein VI and their co-expression as a collagen receptor in human platelets. *J. Biol. Chem.* **272**, 23528–23531 (1997).
4. Gibbins, J. M., Okuma, M., Farndale, R., Barnes, M. & Watson, S. P. Glycoprotein VI is the collagen receptor in platelets which underlies tyrosine phosphorylation of the Fc receptor γ -chain. *FEBS Lett.* **413**, 255–259 (1997).
 5. Clemetson, J. M., Polgar, J., Magnenat, E., Wells, T. N. & Clemetson, K. J. The platelet collagen receptor glycoprotein VI is a member of the immunoglobulin superfamily closely related to Fc α R and the natural killer receptors. *J. Biol. Chem.* **274**, 29019–24 (1999).
 6. Watson, S. P. *et al.* The Role of ITAM- and ITIM-coupled Receptors in Platelet Activation by Collagen. *Thromb. Haemost.* **86**, 276–288 (2001).
 7. Watson, S. P., Auger, J. M., McCarty, O. J. T. & Pearce, a C. GPVI and integrin α IIb β 3 signaling in platelets. *J. Thromb. Haemost.* **3**, 1752–1762 (2005).
 8. Nieswandt, B. & Watson, S. P. Platelet-collagen interaction: Is GPVI the central receptor? *Blood* **102**, 449–461 (2003).
 9. Watson, S. P., Herbert, J. M. J. & Pollitt, a. Y. GPVI and CLEC-2 in hemostasis and vascular integrity. *J. Thromb. Haemost.* **8**, 1456–1467 (2010).
 10. Herr, A. B. Direct evidence of a native GPVI dimer at the platelet surface. *J. Thromb. Haemost.* **7**, 1344–6 (2009).
 11. Jung, S. M. *et al.* Constitutive dimerization of glycoprotein VI (GPVI) in resting platelets is essential for binding to collagen and activation in flowing blood. *J. Biol. Chem.* **287**, 30000–30013 (2012).
 12. Loyau, S. *et al.* Platelet glycoprotein VI dimerization, an active process inducing receptor competence, is an indicator of platelet reactivity. *Arterioscler. Thromb. Vasc. Biol.* **32**, 778–85 (2012).
 13. Moroi, M. & Jung, S. M. Platelet glycoprotein VI: its structure and function. *Thromb. Res.* **114**, 221–33 (2004).
 14. Farndale, R. W., Sixma, J. J., Barnes, M. J. & de Groot, P. G. The role of collagen in thrombosis and hemostasis. *J. Thromb. Haemost.* **2**, 561–73 (2004).
 15. Farndale, R. W. *et al.* Cell-collagen interactions: the use of peptide Toolkits to investigate collagen-receptor interactions. *Biochem. Soc. Trans.* **36**, 241–50 (2008).
 16. Raynal, N. *et al.* Use of synthetic peptides to locate novel integrin α 2 β 1-binding motifs in human collagen III. *J. Biol. Chem.* **281**, 3821–31 (2006).
 17. Giudici, C. *et al.* Mapping of SPARC/BM-40/osteonectin-binding sites on fibrillar collagens. *J. Biol. Chem.* **283**, 19551–60 (2008).
 18. Hohenester, E., Sasaki, T., Giudici, C., Farndale, R. W. & Bächinger, H. P. Structural basis of sequence-specific collagen recognition by SPARC. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 18273–7 (2008).
 19. Konitsiotis, A. D. *et al.* Characterization of high affinity binding motifs for the discoidin domain receptor DDR2 in collagen. *J. Biol. Chem.* **283**, 6861–8 (2008).
 20. Lisman, T. *et al.* A single high-affinity binding site for von Willebrand factor in collagen III, identified using synthetic triple-helical peptides. *Blood* **108**, 3753–6 (2006).
 21. Brondijk, T. H. C., Bihan, D., Farndale, R. W. & Huizinga, E. G. Implications for collagen I chain registry from the structure of the collagen von Willebrand factor A3 domain complex. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5253–5258 (2012).
 22. Zhou, L. *et al.* Structural basis for collagen recognition by the immune receptor OSCAR. *Blood* **127**, 529–537 (2016).
 23. Jarvis, G. E. *et al.* Identification of a major GpVI-binding locus in human type III collagen. *Blood* **111**, 4986–4996 (2008).
 24. Kehrel, B. *et al.* Glycoprotein VI is a major collagen receptor for platelet activation: it recognizes the platelet-activating quaternary structure of collagen, whereas CD36, glycoprotein IIb/IIIa, and von Willebrand factor do not. *Blood* **91**, 491–9 (1998).
 25. Smethurst, P. A. *et al.* Structural basis for the platelet-collagen interaction: the smallest motif within collagen that recognizes and activates platelet Glycoprotein VI contains two glycine-proline-hydroxyproline triplets. *J. Biol. Chem.* **282**, 1296–304 (2007).
 26. Barrow, A. D. *et al.* OSCAR Is a Receptor for Surfactant Protein D That Activates TNF- α Release from Human CCR2+ Inflammatory Monocytes. *J. Immunol.* **194**, 3317–3326 (2015).
 27. Meyaard, L. The inhibitory collagen receptor LAIR-1 (CD305). *J. Leukoc Biol* **83**, 799–803 (2008).

28. Horii, K., Kahn, M. L. & Herr, A. B. Structural basis for platelet collagen responses by the immune-type receptor glycoprotein VI. *Blood* **108**, 936–42 (2006).
29. Smethurst, P. A. *et al.* Identification of the primary collagen-binding surface on human glycoprotein VI by site-directed mutagenesis and by a blocking phage antibody. *Blood* **103**, 903–11 (2004).
30. Brondijk, T. H. C. *et al.* Crystal structure and collagen-binding site of immune inhibitory receptor LAIR-1: unexpected implications for collagen binding by platelet receptor GPVI. *Blood* **115**, 1364–73 (2010).
31. Lebbink, R. J. *et al.* Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. *Matrix Biol.* **28**, 202–10 (2009).
32. Schymkowitz, J. W. H. *et al.* Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10147–10152 (2005).
33. Lecut, C. *et al.* Identification of residues within human glycoprotein VI involved in the binding to collagen: evidence for the existence of distinct binding sites. *J. Biol. Chem.* **279**, 52293–9 (2004).
34. O'Connor, M. N., Smethurst, P. a, Farndale, R. W. & Ouwehand, W. H. Gain- and loss-of-function mutants confirm the importance of apical residues to the primary interaction of human glycoprotein VI with collagen. *J. Thromb. Haemost.* **4**, 869–73 (2006).
35. Sampathkumar, P., Bonanno, J., Fiser, A., Patskovsky, Y., Zencheck, W., Nathenson, S.G., Almo, S. C. Crystal structure of a strand-swapped dimer of Mouse Leukocyte-associated immunoglobulin-like receptor 1 IG-like domain (Unpublished data). *Protein data bank PDB-ID*, 4ESK
36. Orgel, J. P. R. O., Irving, T. C., Miller, A. & Wess, T. J. Microfibrillar structure of type I collagen in situ. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 9001 (2006).
37. Boudko, S. P. *et al.* Crystal structure of human type III collagen Gly991-Gly1032 cystine knot-containing peptide shows both 7/2 and 10/3 triple helical symmetries. *J. Biol. Chem.* **283**, 32580–9 (2008).
38. Herr, A. B. & Farndale, R. W. Structural insights into the interactions between platelet receptors and fibrillar collagen. *J. Biol. Chem.* **284**, 19781–5 (2009).
39. Emsley, J., Knight, C. G., Farndale, R. W., Barnes, M. J. & Liddington, R. C. Structural basis of collagen recognition by integrin $\alpha 2\beta 1$. *Cell* **101**, 47–56 (2000).
40. Widmer, C. *et al.* Molecular basis for the action of the collagen-specific chaperone Hsp47/SERPINH1 and its structure-specific client recognition. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13243–7 (2012).
41. Miura, Y., Takahashi, T., Jung, S. M. & Moroi, M. Analysis of the interaction of platelet collagen receptor glycoprotein VI (GPVI) with collagen. A dimeric form of GPVI, but not the monomeric form, shows affinity to fibrous collagen. *J. Biol. Chem.* **277**, 46197–204 (2002).
42. Rudnick, S. I. & Adams, G. P. Affinity and Avidity in Antibody-Based Tumor Targeting. *Cancer Biother. Radiopharm.* **24**, 155–161 (2009).
43. Ericsson, U. B., Hallberg, B. M. & Detitta, G. T. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal. Biochem.* **357**, 289–298 (2006).
44. Leslie, A. G. W. & Powell, H. R. *Evolving Methods for Macromolecular Crystallography. Evolving Methods for Macromolecular Crystallography* **245**, (Springer Netherlands, 2007).
45. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1204–14 (2013).
46. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235–42 (2011).
47. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
48. Okuyama, K. *et al.* Crystal Structures of Collagen Model Peptides with Pro-Hyp-Gly Repeating Sequence at 1.26 Å Resolution: Implications for Proline Ring Puckering. *Biopolymers* **76**, 367–377 (2004).
49. Vagin, A. a *et al.* REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2184–95 (2004).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 486–501 (2010).
51. van der Plas, R. M. *et al.* Binding of von Willebrand factor to collagen type III: role of specific amino acids in the collagen binding domain of vWF and effects of neighboring domains. *Thromb. Haemost.* **84**, 1005–11 (2000).

Supplementary data

Methods

Processing of diffraction data and structure solution

Diffraction data were integrated with MOSFLM⁴⁴ and scaled using Aimless⁴⁵ in the CCP4 suite⁴⁶. Phasing of the GPVI dataset by molecular replacement was performed with PHASER⁴⁷ and 4 copies of GPVI obtained from the previously solved structure of GPVI (PDB-ID: 2G17)²⁸. Because of an arrangement in two equal pairs of dimers related by non-crystallographic symmetry, chains B and D were restrained to chain A and C during refinement of the structure, for which tight restraints were applied on main chain atoms and medium restraints on side chain atoms.

Molecular replacement to solve the GPVI-(GPO)₅ complex involved initial placement of two copies of the refined model of GPVI truncation mutant and rigid body refinement using PHASER. The resulting *Fo*-*Fc* difference density map (Figure S1a) allowed unambiguous placement of one (GPO)₅-trimer, generated from the crystal structure of (GPO)₁₀ (PDB-ID: 1V4F⁴⁸; Figure S1b-d). The structure of the GPVI-(GPO)₃ complex in the same space group was obtained by rigid body refinement of the GPVI-(GPO)₅ complex (Figure S1e) and manual modification of the collagen peptide.

All structures were refined with REFMAC⁴⁹ alternated with model improvement using COOT⁵⁰ (Figure S1f-i). Refinement included TLS-refinement with 1 TLS-domain per GPVI chain and per collagen-trimer. In all cases, sufficient density was available to model the truncated loops (Figure S3b-c). Statistics of data processing and refinement for all structures are listed in Table S2.

Thermostability assay

GPVI loop truncation mutants were diluted 3- to 10-fold in PBS to 0.25 mg/mL. 18.75 μ L of diluted protein solution was mixed with 6.25 μ L 40x Sypro Orange (Sigma-Aldrich) in a 96-well plate (Greiner 652201). Thermal-unfolding curves were measured in sealed plates, heated from 289 to 368K with increments of 0.5K in an iCycler iQ Real Time PCR Detection System (Bio-Rad). The change in fluorescence was measured as described⁴³. Melting temperatures (T_m) were defined as the inflection point of the unfolding curves.

Collagen-binding assay and data processing

To obtain fibrillar collagen, human placenta collagen type I (Sigma-Aldrich cat. no. C7774) was dissolved in 50 mM acetic acid and dialyzed overnight against 10 mM sodium phosphate buffer pH 7.4⁵¹. Collagen peptides were dissolved in 10 mM acetic acid and coated directly from acidic solution. Coating of peptides (5 μ g/mL) and collagen fibrils (50 μ g/mL) on 96-well plates (Costar 2595; Corning) was performed overnight at RT. Wells were blocked with 3% (w/v) BSA in PBS for 2 hours at RT and then washed 4x with PBST (PBS + 0.1% Tween-20).

Serial 3-fold dilutions with concentrations ranging from 2.0 to 0.002 μ M GPVI-Fc in PBS + 0.1% (w/v) BSA were incubated for 2 hours at RT. After discarding excess GPVI-Fc and 4x washing, 100 μ L of 2000-fold diluted polyclonal rabbit anti-human IgG-HRP (DAKO, P0214) in PBS + 0.1% BSA was added to each well and incubated for 1 hour at RT. After 4x washing, color reactions were started by addition of 100 μ L HRP-substrates *ortho*-phenylenediamine (0.4 mg/mL) and H₂O₂ (0.08%, v/v) in phosphate-citrate buffer pH 5.0. Reactions were stopped by adding 50 μ L 1 M H₂SO₄ to each well, and absorbance was measured at 490 nm on a Model 680 Microplate Reader (Bio-Rad). Binding of each GPVI-Fc mutant was measured at least nine times in three independent sets of three measurements.

The average absorbance of three series of blanc measurements without GPVI was subtracted from each measurement. Binding curves are displayed relative to the maximum binding of GPVI-Fc WT. For each mutant, the apparent dissociation constant $K_{D,app}$ and relative $B_{max'}$ if evidently other than 1, were obtained by fitting the binding data to formula 1 using non-linear regression in SigmaPlot (Systat software).

$$(1) Abs = \frac{B_{max} \times c}{K_{D,app} + c} \text{ with } c \text{ is the GPVI concentration } (\mu\text{M})$$

Binding of wild GPVI-Fc and the E40A and R38A mutants to CRP, GPP, Toolkit III-30 and III-30 derivatives (Table 1) was measured as previously described²³. Binding to all peptides was measured at a fixed GPVI-Fc concentration of 20 μ g/mL and colored using anti-human IgG-HRP and HRP-substrate KPL TMB Sure Blue Reserve (Insight Biotechnology). Absorbance was measured at 450 nm on a Fluostar plate reader. Binding to these peptides was measured in at least three different experiments, and scaled to binding of wild type GPVI-Fc to CRP.

Analysis of the GPVI-(GPO)₅ complex using FoldX

The FoldX Suite³² was obtained under academic license from <http://foldxsuite.crg.eu/> (64-bit) and run on a desktop machine equipped with 48 processing cores and Debian 7 operating system. FoldX procedures described in this manuscript were all obtained by using the 'ignore waters' option, and default values for ionic strength, pH, and temperature/gas constant. Experiments performed using other options for water modeling did not lead to substantially different results and conclusions.

Optimization of energies of all residues in a pdb-files containing the coordinates of the GPVI-(GPO)₅ and GPVI-(GPO)₃ complex was performed using the *RepairPDB* function. Minimal changes were observed for the GPVI-collagen interfaces compared to the models obtained by crystallography.

Relative contributions of (GPO)₅- and (GPO)₃-residues to the total interaction energies were calculated using the *SequenceDetail* function, and performed on pdb-files containing coordinates of the collagen peptide only, and files containing coordinates of the complex between the peptide and one GPVI-molecule. The energy contribution to complex formation for (GPO)₅- or (GPO)₃-residue *i* was then calculated using formula 2, and results are shown (graph and surface rendering) in Figure S4.

$$(2) \Delta E_i = E_{\text{complex},i} - E_{\text{peptide},i}$$

Analysis of GPVI-binding sites in Toolkit-III peptides

To determine the sequence motif in Toolkit-III peptides that best correlates with the previously obtained binding data for GPVI-Fc dimers¹¹, we examined the presence and content in Toolkit-III peptides¹⁶ of the full POGPOGPO-sequence; the full site on (GPO)₅ that interacts with GPVI in our crystals, and which includes the ends that minimally contribute to GPVI-binding. In addition, we also searched for non-overlapping patterns of decreasing length starting with either P or O (G exists by default). Values for correlation between the two arrays containing pattern-content and binding data were calculated using Pearson's correlation coefficient and are shown in Figure S8a.

Peptide model generation and superimposition of GPVI-molecules on the collagen helix

To determine possible sites on collagen that - depending on the helicity of the helix, taking 7₂ vs. 10₃ as extremes states - can be simultaneously occupied by two GPVI-molecules, we generated hypothetical models of dual binding using our GPVI-(GPO)₅ structure. The tightly folded 7₂-helix structure of (GPO)₁₀ was adapted from the infinite (GPO)_n-structure with PDB-ID 1V4F⁴⁸, similarly as shown in Figure S1b-c. The more loosely folded 10₃-helix was obtained from the structure of the G982-G1023-region of human type III collagen (PDB-ID: 3DMW³⁷, residues 6-17), which was manually modified to (GPO)₁₀. Models of Toolkit-peptides III-30, III-01, and III-40 were generated from the 10₃-helix, by manual modification of the amino acid sequence, thereby preserving the 10₃-type backbone structure.

Positioning of the GPVI-molecules on the different sites on both helix-types was performed by superposition of the first GPVI-molecule and the (GPO)₅-residues to which it binds (Hyp3-Hyp6 of the middle chain and Pro2-Pro8 of the trailing chain) onto consecutive sites on (GPO)₁₀-7₂ and (GPO)₁₀-10₃, formed by increasing (GPO)₁₀-residue numbers and other chain combinations (see Figure S9). Positioning of GPVI-molecules on the Toolkit-peptide models (Figures 4 and S8) was performed using the same method. All output models were filtered for solutions that are precluded by sterical hindrance.

Supplementary Tables & Figures

Table S1. Primer sequences for construction of GPVI loop-truncation mutants

Mutant	+/-	Deleted residues	Primer sequence
ΔPAVS	Fw	Pro122-Ser125	5'-CTCTCAGCCAGCCCGGCTCAGGAGGGGACGTAACCC-3'
ΔPAVS	Rv	Pro122-Ser125	5'-GGGTTACGTCCCCTCTGAGCCGGGCTGGGCTGAGAG-3'
ΔPAVSG	Fw	Pro122-Ser125, Gly127	5'-CTCTCAGCCAGCCCGGCTCAGGGGACGTAACCTACAGTGTC-3'
ΔPAVSG	Rv	Pro122-Ser125, Gly127	5'-GACACTGTAGGGTTACGTCCCCTGAGCCGGGCTGGGCTGAGAG-3'
ΔPAPYK	Fw	Pro151-Lys155	5'-GCTCTGTACAAGGAAGGGGACAATCCCAGAGATGGTACCGGG-3'
ΔPAPYK	Rv	Pro151-Lys155	5'-CCCGGTACCATCTCTCGGGATTGTCCTTCTTGTACAGAGC-3'
ΔPAPYKN	Fw	Pro151-Asn156	5'-GCTCTGTACAAGGAAGGGGACCCGAGAGATGGTACCGGGCTAG-3'
ΔPAPYKN	Rv	Pro151-Asn156	5'-CTAGCCGGTACCATCTCTCGGGTCCCCTTCTTGTACAGAGC-3'

Table S2. Data collection and refinement statistics

	GPVI ΔPAVS-PAPYKN	GPVI-(GPO) ₅	GPVI-(GPO) ₃
Data Collection			
Synchrotron	Swiss Light Source (SLS)	European Synchrotron Radiation Facility (ESRF)	European Synchrotron Radiation Facility (ESRF)
Beamline	PX	ID23-1	ID23-1
Wavelength (Å)	0.99999	0.97242	0.97242
Space group	P 2 ₁	P 4 ₁ 2 ₁ 2	P 4 ₁ 2 ₁ 2
Cell dimensions			
a, b, c (Å)	78.65, 44.05, 117.66	59.97, 59.97, 313.2	59.72, 59.72, 319.7
α, β, γ (°)	90, 104.7, 90	90, 90, 90	90, 90, 90
Resolution range (Å)*	52.38 – 1.90 (1.94 - 1.90)	59.97 - 2.50 (2.61 - 2.50)	58.70 - 2.50 (2.61 - 2.50)
No. unique reflections	58873 (3656)	20736 (2463)	21273 (2503)
Redundancy	2.1 (2.1)	7.9 (7.4)	6.4 (6.8)
R_{merge}	0.081 (1.08)	0.121 (1.03)	0.113 (0.971)
I/σ	5.0 (1.1)	9.7 (1.8)	8.5 (1.9)
Completeness (%)	95.1 (93.3)	99.5 (99.8)	99.9 (100.0)
CC(1/2)	0.990 (0.620)	0.995 (0.535)	0.984 (0.646)
Refinement			
R_{work} / R_{free}	0.2212 / 0.2597	0.2081 / 0.2574	0.2279 / 0.2603
Content of asymmetric unit	4× GPVI, 14× Cl ⁻ , 2× PO ₄ ³⁻ , 2× PEG ₆ , 344× H ₂ O	2× GPVI, 1× (GPO) ₅ , 2× Cl ⁻ , 2× PEG ₄ , 88× H ₂ O	2× GPVI, 1× (GPO) ₃ , 2× Cl ⁻ , 2× PEG ₄ , 41× H ₂ O
No. atoms	6101	3114	3210
Protein	5645	2973	3117
Water / other ligands	344 / 112	88 / 53	41 / 52
Average B / Wilson B (Å²)	33.2 / 25.6	38.1 / 34.2	69.5 / 56.4
RMS deviations			
Bond lengths (Å)	0.0111	0.0092	0.0082
Bond angles (°)	1.4800	1.6349	1.5635
Ramachandran Plot			
Favored (%)	98.5	97.4	96.0
Allowed (%)	1.37	2.63	3.71
Outliers (%)	0.14	0	0.25

* Numbers between brackets refer to the highest resolution shell.

Table S3. Collagen I binding of GPVI wild type and mutants with substitutions outside the primary binding site

Mutant	K _{D,app}	95% confidence interval	Increase	Max. binding*
WT	30.1 nM	26.8 ≤ K _D ≤ 33.5 nM		
K59A	72.4 nM	39.0 ≤ K _D ≤ 106 nM	2-fold	85% (76 - 95%)
K41A	35.2 nM	28.3 ≤ K _D ≤ 42.1 nM		119% (113 - 124%)
Q48A	19.3 nM	13.5 ≤ K _D ≤ 20.5 nM		109% (102 - 115%)
L62E	29.3 nM	22.4 ≤ K _D ≤ 36.2 nM		108% (103 - 114%)
R46E	47.7 nM	36.3 ≤ K _D ≤ 59.2 nM		105% (99 - 110%)

* Values represent the max. binding (B_{max}) and 95% confidence intervals, if other than 100%.

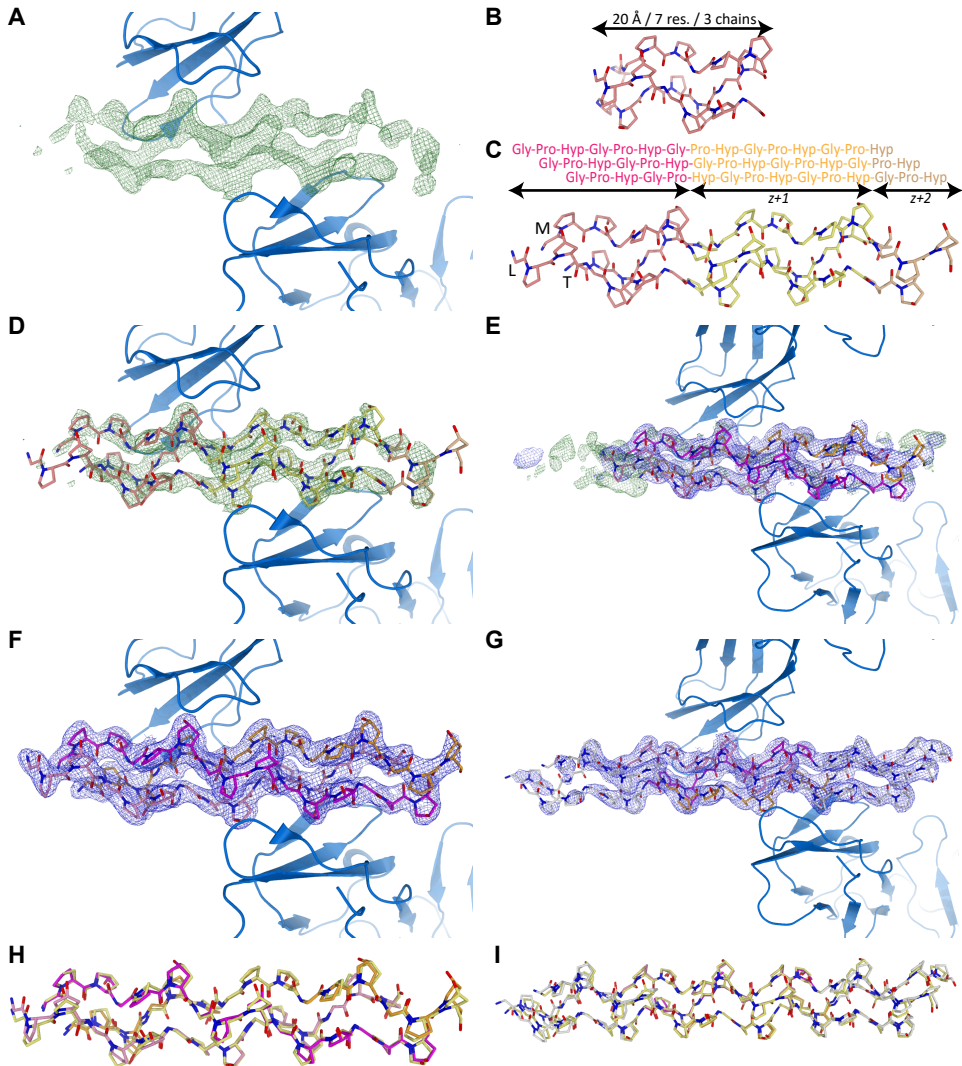


Figure S1. Model building of collagen peptides in the GPVI-(GPO)₅ and GPVI-(GPO)₃ complexes. (A) Cartoon representation of two molecules GPVI (marine-blue) placed in the asymmetric unit of the GPVI-(GPO)₅ complex after molecular replacement, and the *Fo*-*Fc* difference electron density map contoured at 2σ (green). (B) Structure of one 7_2 -helix turn of the infinite structure of (GPO)_n (PDB ID code: 1V4F)⁴⁸. (C) Structure of 3 one-residue staggered (GPO)₅-chains as constructed from the (GPO)_n-asymmetric unit (orange) and two interconnected helix turns (yellow and pink) that are obtained by translational symmetry along the z-axis. (D) Manual positioning of the (GPO)₅ triple helix in the *Fo*-*Fc* electron density map after two GPVI-molecules were placed using molecular replacement. (E) $2Fo$ -*Fc* (blue, 1.2σ contoured) and *Fo*-*Fc* (green, 2σ contoured) electron density maps after rigid body refinement of the GPVI-(GPO)₅ complex against the dataset of the GPVI-(GPO)₃ complex. (F) Cartoon representation of GPVI, sticks model of (GPO)₅ (tricolor), and the $2Fo$ -*Fc* electron density map contoured at 1.2σ within 1.6 \AA around all (GPO)₅-atoms after the final refinement of the GPVI-(GPO)₅ complex ($R_{\text{work}} = 0.2081$; $R_{\text{free}} = 0.2574$). (G) Cartoon representation of GPVI, sticks model of (GPO)₃ (tricolor; white (GPP)₂-termini), and the $2Fo$ -*Fc* electron density map contoured at 1.2σ within 1.6 \AA around all (GPO)₃-atoms after the final refinement of the GPVI-(GPO)₅ complex ($R_{\text{work}} = 0.2279$; $R_{\text{free}} = 0.2603$). (H) Superposition of the structures of (GPO)₅ from the GPVI-(GPO)₅ complex (tricolor) and (GPO)₅ as adapted from the infinite (GPO)_n-structure (yellow). (I) Superposition of the structures of (GPO)₃ from the GPVI-(GPO)₃ complex (tricolor; white (GPP)₂-termini) and (GPO)₃ as adapted from the infinite (GPO)_n-structure (yellow).

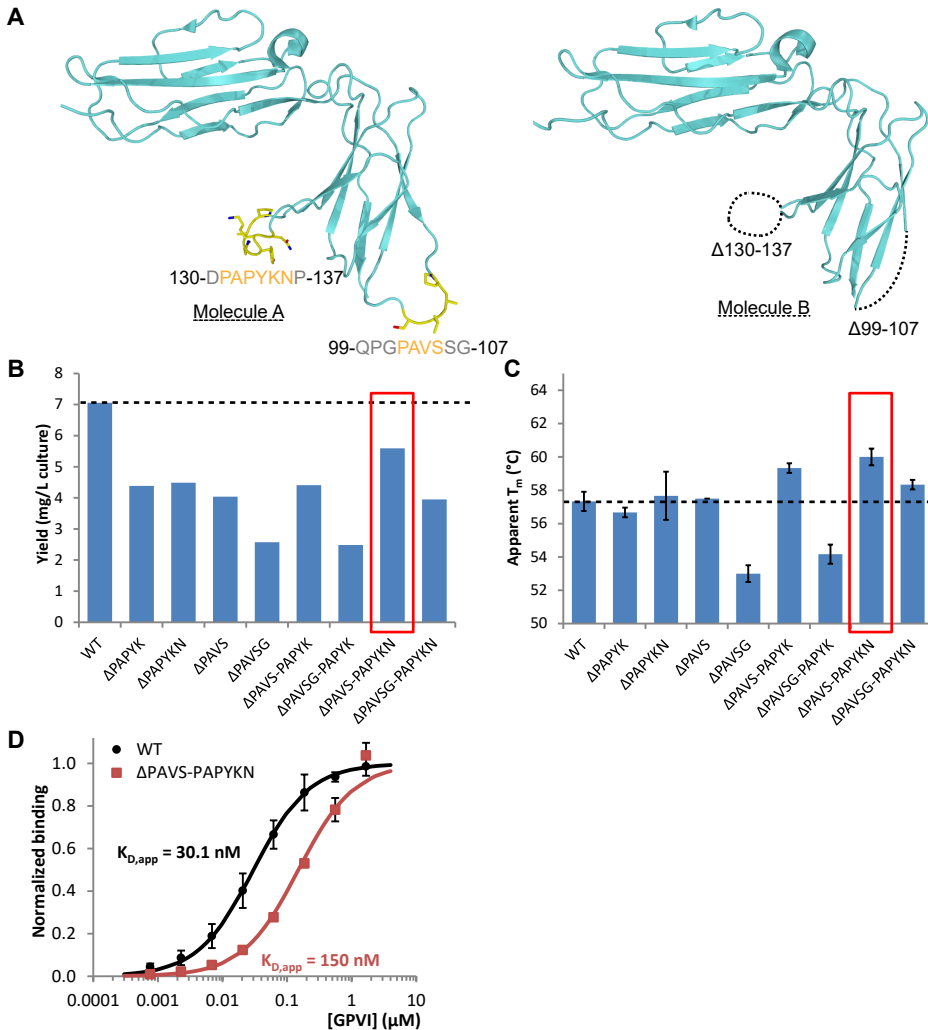


Figure S2. Construction of GPVI loop truncation mutants. (A) Previously solved GPVI-structure (PDB ID code: 2GI7²⁸; cyan cartoon representation) highlighting the flexible loops in the D2-domain (res. 99-107 and 130-137) that are present in molecule A (left, yellow sticks), but absent in molecule B (right, dashed lines). **(B)** Expression yields of wild type GPVI and eight loop truncation mutants showing moderate reductions for all mutants with respect to wild type protein (7.1 mg/L culture). The highest yield (5.6 mg/L culture) was obtained for the $\Delta\text{PAVS-}\Delta\text{PAPYKN}$ mutant (red box). **(C)** Melting temperatures (T_m) of wild type GPVI and loop truncation mutants by measuring thermal-unfolding curves⁴³ and determination of the inflection points of unfolding curves for all proteins. The highest melting temperature was obtained for the $\Delta\text{PAVS-}\Delta\text{PAPYKN}$ mutant (red box), which appeared more thermostable than wild type GPVI. **(D)** Collagen I-binding of wild type GPVI (black) and the $\Delta\text{PAVS-}\Delta\text{PAPYKN}$ mutant (red) as measured in a solid-state assay. Data measured as A490 is normalized to the binding of wild type protein. All data points represent the mean \pm SD of at least three independent experiments. Binding curves are fitted to the equation $Abs = (B_{max} \times c) / (K_{D,app} + c)$ for 1 to 1 binding using non-linear regression in SigmaPlot, where c is the GPVI concentration (μM).

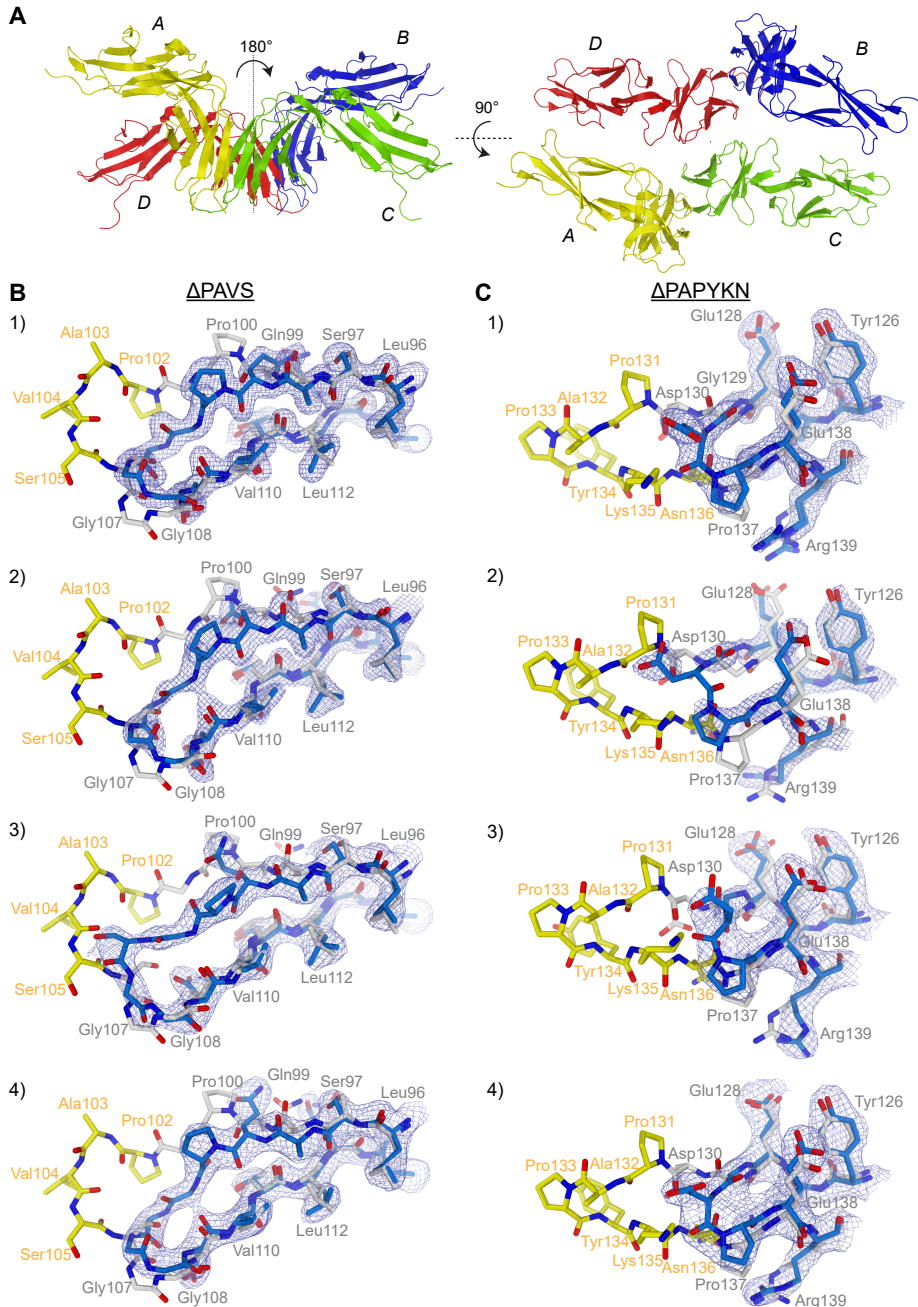


Figure S3. Crystal structure of the GPVI Δ PAVS- Δ PAPYKN mutant and conformations of the truncated loops. (A) 1.9 Å-resolution structure of the GPVI Δ PAVS- Δ PAPYKN mutant showing a dimer-of-dimer arrangement with two GPVI-dimers (yellow-green and blue-red) that are related by a non-crystallographic 2-fold rotation. (B-C) Conformation of the truncated Δ PAVS- (B) and Δ PAPYKN-loops (C) as observed in GPVI-monomers A (panel 1) and C (panel 2) of the unbound structure, as well as GPVI-monomers A (panel 3) and B (panel 4) of the GPVI-(GPO)₅ complex. Truncated loops are shown in blue sticks within the respective $2F_o-F_c$ electron density map contoured at 1.2σ . Also shown in every panel is the conformation of the native loops from the structure of wild type GPVI (PDB-ID: 2GI7²⁸; yellow and gray sticks) for comparison, which highlights the regions that are truncated in the mutant.

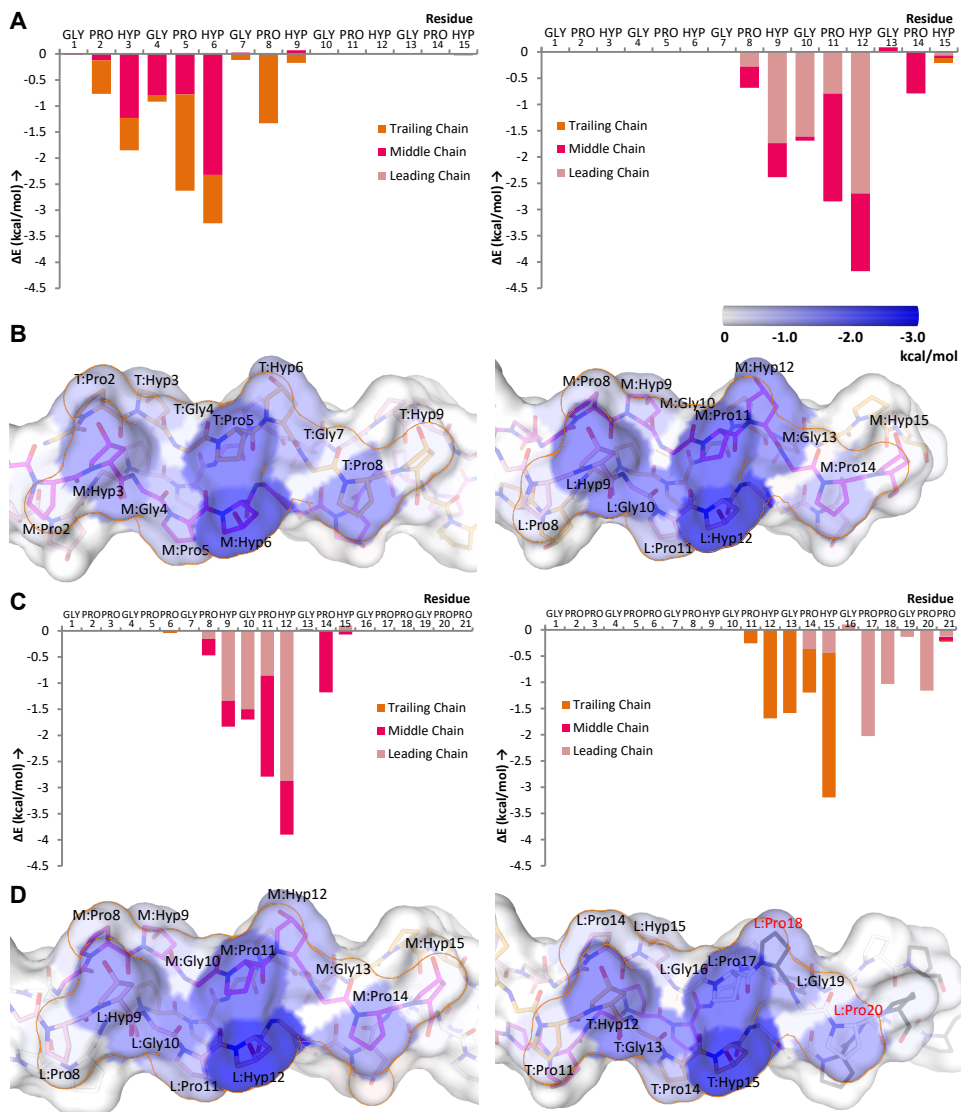


Figure S4. Analysis of the interaction energies in the GPVI-collagen complexes using FoldX. Relative contributions to the binding energy for formation of the GPVI-collagen peptide complexes, calculated per collagen residue using the FoldX *SequenceDetail* function³² (see *supplementary methods*). **(A-B)** Contributions for binding of the first (left) and second (right) GPVI-molecule in the GPVI-(GPO)₅-complex specified per collagen residue and chain, and shown in a graph **(A)** and on a (GPO)₅-surface rendering **(B)**. The surface is colored on a blue-white scale to indicate regions that have high (blue) and low (white) contributions. The surface region that is located within 4.5 Å of GPVI is surrounded by an orange line. **(C-D) *id.*** For the GPVI-(GPO)₃-complex. Two prolines of the (GPP)₂-terminus of (GPO)₃ that are situated within the region interacting with the second GPVI-molecule, but present as hydroxyproline in all the other interfaces are indicated with a red label. In all cases, the FoldX-analysis indicates the strongest contributions for a central Pro-Hyp-patch (T:Pro5-M:Hyp6 in (GPO)₅), followed by somewhat smaller involvement of surrounding residues (M:Hyp3, T:Hyp6 and T:Pro8 in (GPO)₅). Residues at the extremes of the binding site (M:Pro2 and T:L:Hyp9 in (GPO)₅) are very minimally contributing.

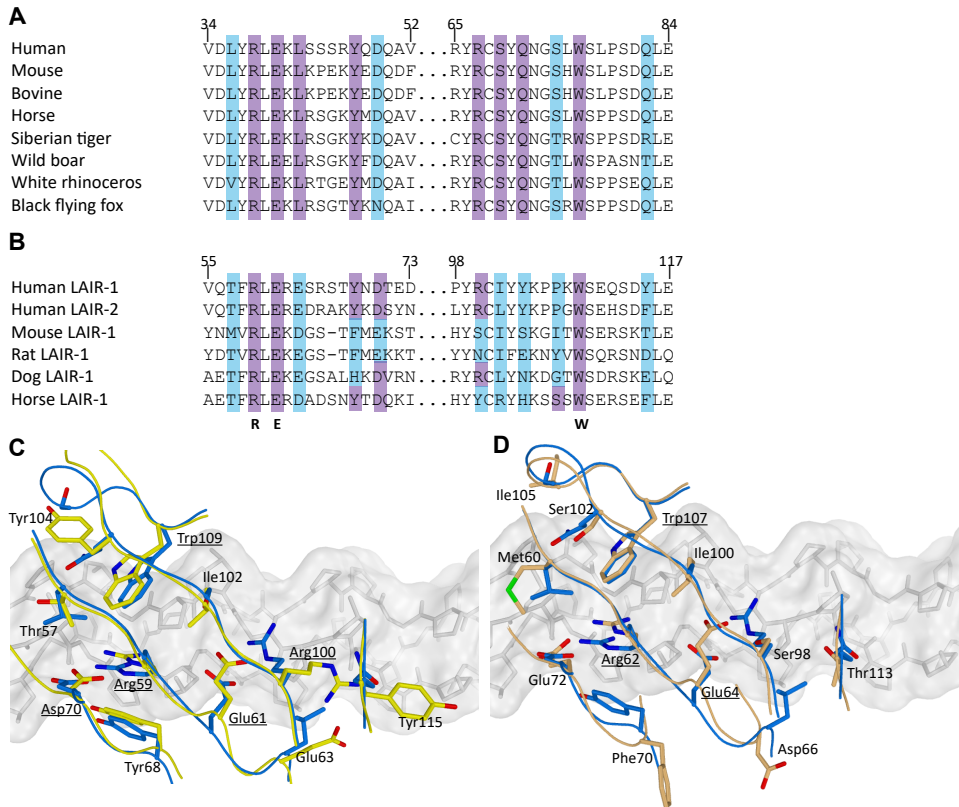


Figure S5. Alignment of GPVI and LAIR-orthologues. (A) Sequence alignment of the GPVI-collagen binding region of different vertebrate species. Conserved and non-conserved residues involved in collagen-binding are highlighted in purple and blue, respectively. (B) Sequence alignment between human GPVI and LAIR-1 orthologues highlighting conserved (green) and different (yellow) residues observed in the collagen-binding site of GPVI. Only the three GPVI-residues with the strongest contribution to collagen-binding, Arg38, Glu40, and Trp76, are conserved within all LAIR-1 orthologues (Arg59, Glu61, and Trp109 in human LAIR-1). (C-D) Superposition of the structure of human (C) and mouse (D) LAIR-1 (PDB-IDs: 3KGR and 4ESK)³⁰ onto the GPVI-(GPO)₅ complex emphasizing the structural difference within the collagen binding site. Shown in sticks representation are the GPVI-residues (blue) situated in the collagen binding site, and residues at equivalent positions in hLAIR-1 (yellow) and mLAIR-1 (salmon). Labels refer to the residues and numbering in LAIR-1. Also shown in gray are the (GPO)₅-residues and surface.

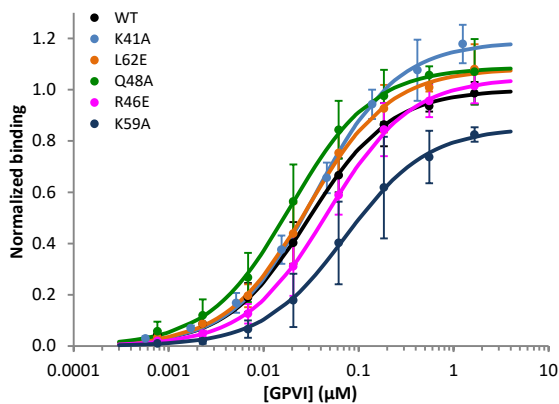
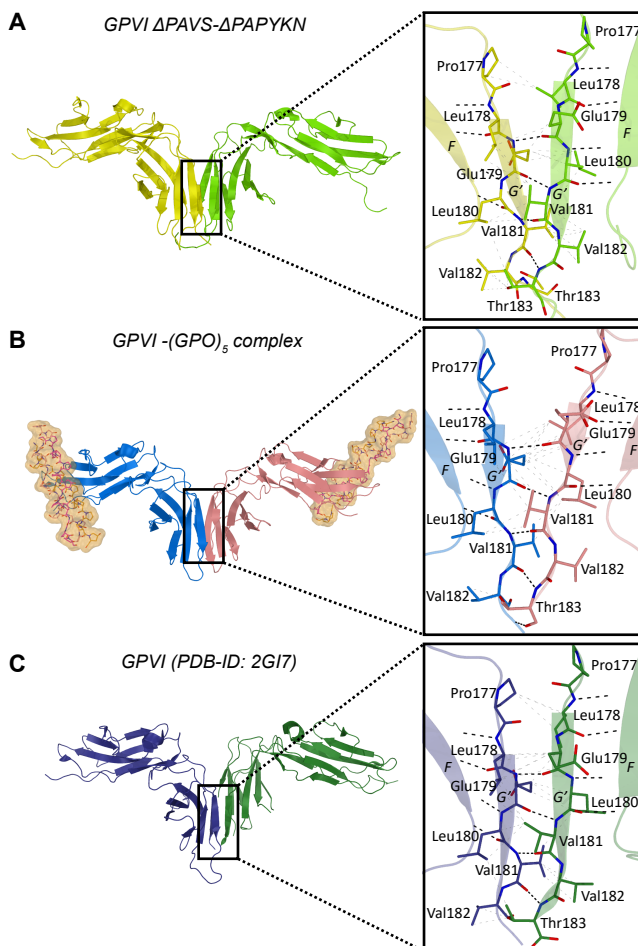
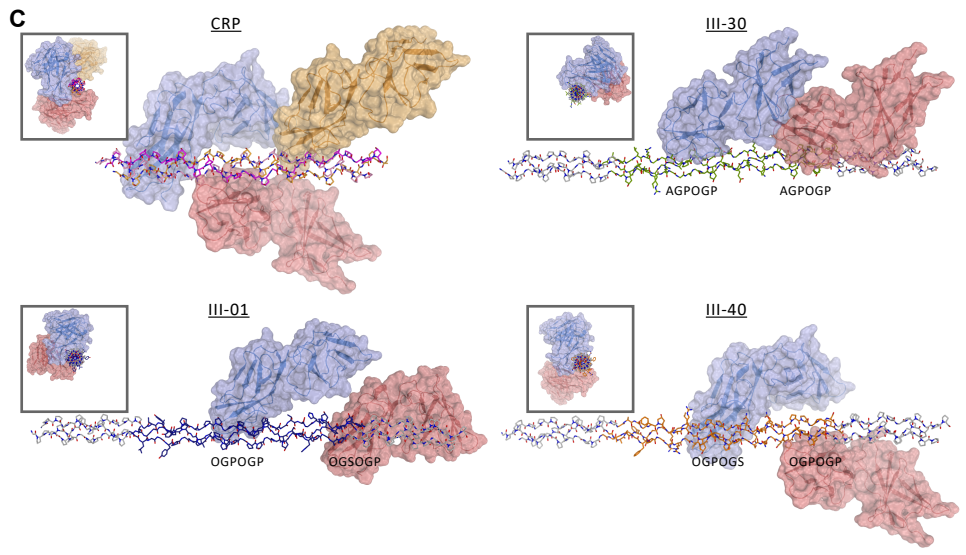
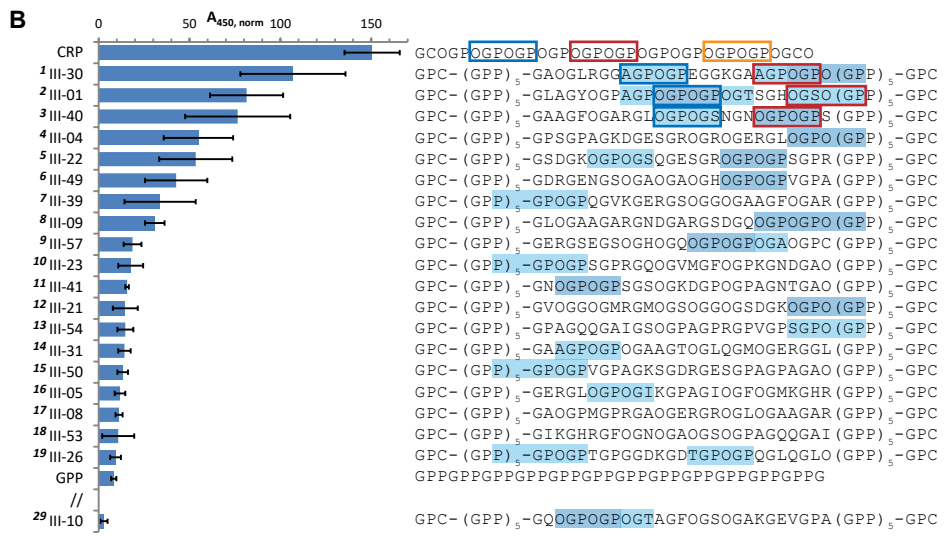
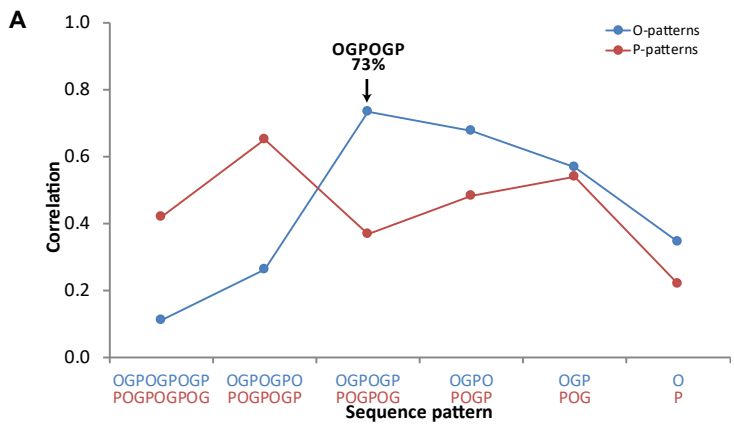


Figure S6. Mutational analysis of GPVI-residues outside the primary collagen-binding site. Collagen I-binding of wild type GPVI (black) and point mutants located in the groove between β -strands C'-E (colored series) as measured in a solid-state assay. Data measured as A490 is normalized to the binding of wild type protein. All data points represent the mean \pm SD of at least three independent experiments. Binding curves are fitted to the equation $Abs = (B_{max} \times c) / (K_{D,app} + c)$ using non-linear regression in SigmaPlot, where c is the GPVI concentration (μ M).

Figure S7. Crystallographic GPVI-dimer formation by back-to-back interaction of D2-domains. Back-to-back dimerization of GPVI D2-domains as observed in the structures of the GPVI Δ PAVS- Δ PAPYKN mutant (left), the GPVI-(GPO)₅ complex (middle) and the previously solved structure of GPVI PDB ID code: 2GI7²⁸. The lower panels show the extended β -sheet formed by two D2-domains. Labeled are β -strands F and G', as well as residues 177-183 that are involved in formation of the dimer. Inter- and intramolecular hydrogen bonds between backbone atoms are indicated with the dashed lines.





D

```

5      10     15     20     25     30     35     40     45     50
1  GLAGYOGPAGFOGPOGPOGTSGHOGSOGSOGYQGPQGEQAGPSPGPGP
51  OGAIGPSGPAGKDGESGROGROGERGLOGPOGIKGPAGIOGFOGMKGRHG
101 FDGRNGEKGETGAOGLKGENGLOEENGAOGPMGPRGAOGERGROGLOGAA
151 GARGNDGARGSDGQOGPOGPOGTAGFPFSOGAKGEVGPAGSOGSNGAOGQ
201 RGEQFPQGHAGAQQGPOGPOGINGSOGGKGMGPAGIOGAOGLMGARGPOG
251 PAGANGAOGLRGGAGEOGKNGAKGEOGPRGERGEAGIOGVOGAKGEDGKD
301 GSOGEOGANGLOGAAGERGAOFRGPAGPNGIOGEKGPAGERGAOGPAGP
351 RGAAGEOGRDGVVOGGOMRGMOGSOGGOGSDGKOGPOGSQGESGROGPOG
401 PSGPRGQOGVMGFQPKGNDGAOGKNGERGGOGGOGPOGPOGKNGETGPO
451 GPOGPTGPGGDKGDTGPOGPQQLQGLQGTGGPOGENGKOGEOGPKGDAGA
501 OGAOGGKGDAGAOGERGPOGLAGAOLRGGAGPOGPEGGKGAAGPOGPOG
551 AAGTOGLQGMOGERGGLGSGOPKEDKKEOGGOGADGVOGKDGPRGPTGPI
601 GPOGPAGQOGDKGEGGAOGLQGIAGPRGSOGERGETGPOGPAGFOGAOGQ
651 NGEQGGKGERGAOGEKKEGGPOGVAGPOGGSGPAGPOGPQGVKGERGSOG
701 GOGAAGFOGARGLOGPOGSNGNOGPOGPSSOGKDGPOGPAGNTGAOGSO
751 GVSQPKGDAGQOGEKGSOGAQQGPOGAOGLGIAGITGARGLAGPOGMOGP
801 RGSOGPQGVKGESGKOGANLSGERGPOGPQGLQGLAGTAGEOGRDNQO
851 SDGLOGRDGSOGGKDRGENSGOGAOGAOGHOGPOGPVGPAGKSGDRGES
901 GPAGPAGAOGPAGSRGAOGPQGRGDKGETGERGAAGIKGHRGFQNOGA
951 OGSOGPAGQQGAIGSOGPAGPRGVPVPSGPOGKDGTSGHOGPIGPOGPRG
1001 NRGERGSESGOHOGOGPOGPOGAOGPCCG

```

E

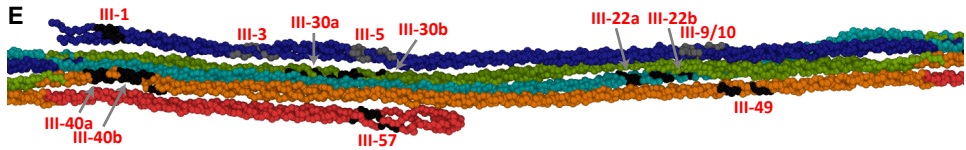
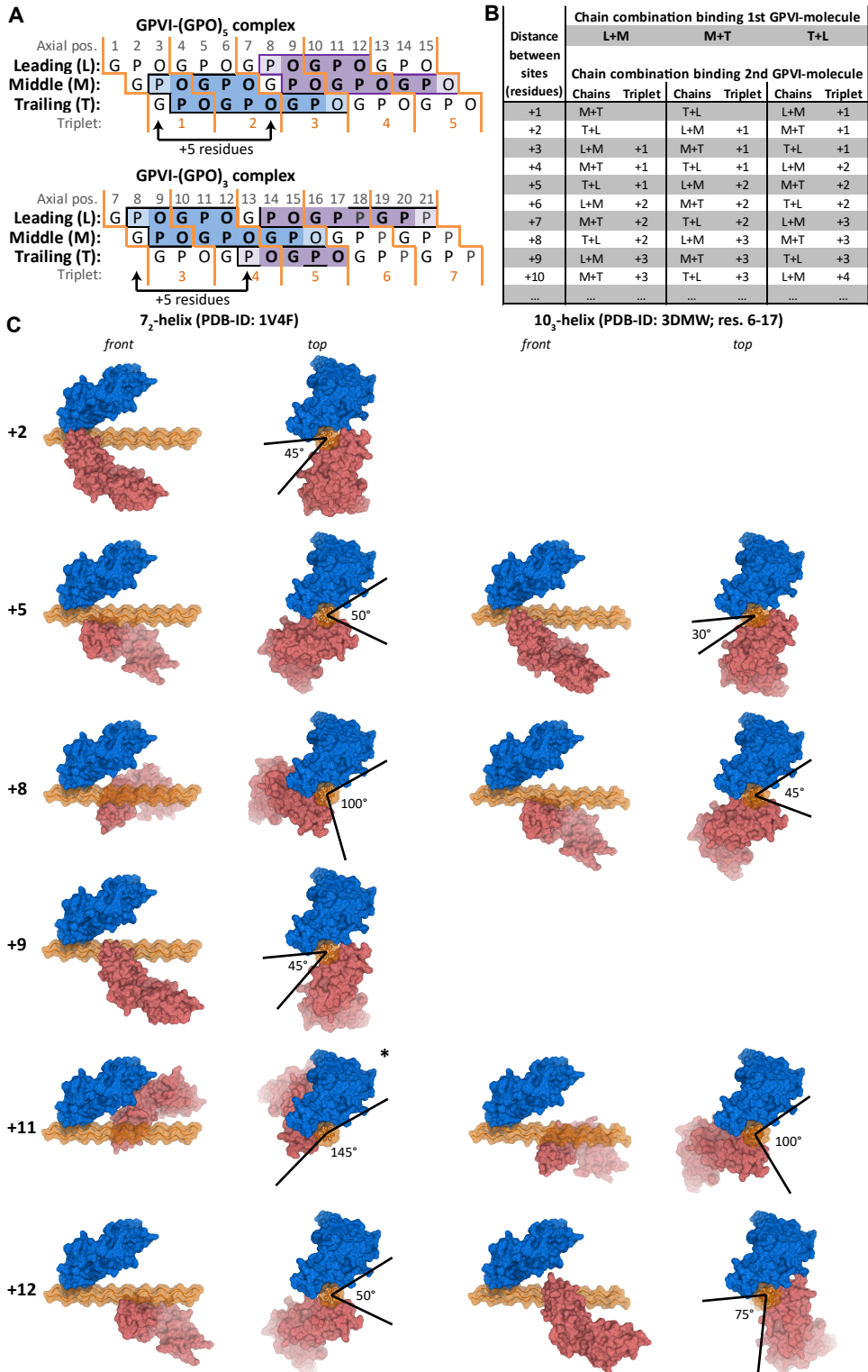


Figure S8. Putative positions of GPVI-binding sites in Toolkit III-peptides and collagen III. (A) Correlation graph showing the correlation between previously obtained data screening for collagen Toolkit III peptides that bind GPVI dimer (data adapted from Jung et al.)¹¹ and the existence of non-overlapping sequence patterns of various length starting with hydroxyproline (O; blue) or proline (P; red) in these peptides. The highest correlation (73%) was found between the measured data and the presence of an OGPOGP-pattern, suggesting this pattern to be the primary recognition sequence in Toolkit-III peptides. (B) Previously obtained data of GPVI-binding to CRP, GPP, and Toolkit III peptides (left) and the respective sequences (right) in order of decreasing potential for binding to GPVI. OGPOGP-, XGPOGP-, and OGPOGX-sequences in Toolkit-peptides are highlighted in blue. The outlined sequences in glue, red, and yellow refer to superposition of GPVI-molecules onto helix models in subsection C. (C) Binding poses of GPVI onto a 7_2 helix model of CRP (PDB-ID: 1V4F; top left panel)⁴⁸ and 10_3 helix (PDB-ID: 3DMW; residues 6-17³⁷) models of peptide III-30 (top right), III-01 (bottom left), and III-40 (bottom right). GPVI-molecules were positioned by superimposition onto sites constituted by the outlined sequences in subsection B and colored accordingly. The insets show the projection along the helix axis, indicating that one face of the peptide would still be accessible and that these binding modes are likely possible in solid-state experiments. (D) Amino acid sequence of the full collagen III triple helical part, in which potential OGPOGP-, XGPOGP-, and OGPOGX-motifs that appear in Toolkit-III peptides are emphasized and colored according to strong (black) or weak (gray) GPVI-binding of the peptide in previous assay. (E) One D-segment of a collagen microfibril³⁶ indicating the positions of potential strong (black) and weak (dark gray) GPVI-binding sites in the context of a fibril. D-segments are colored as follows: blue (D1), light blue (D2), green (D3), orange (D4), and red (D5), according to the color scheme in Herr & Farndale.



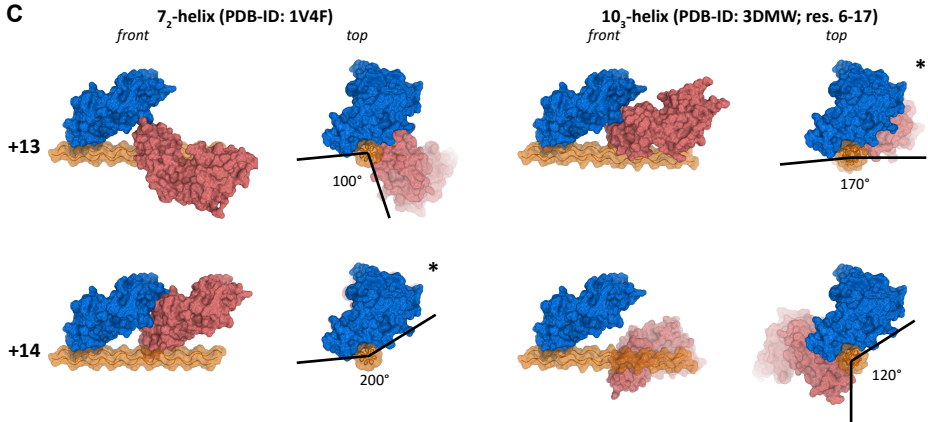


Figure S9. Dual binding modes of GPVI-molecules to a single 7₂- or 10₃-helix. (A) Schematic representation of the (GPO)₅ (top) and (GPO)₃ (bottom) peptides similar as main Figure 1b, but here showing the relative positioning between two GPVI-binding sites on the same collagen peptide: 5 residues apart on both (GPO)₅ and (GPO)₃. (B) Table showing the relative positioning between GPVI-binding sites on a helix and the possible chain combinations involved in binding to the first and second GPVI-molecule. Also indicated is the relative position in the collagen sequence, specified as the number of triplets with which the second site is shifted compared to the first site. (C) Hypothetical models of two GPVI-molecules binding to a 7₂- (left) or 10₃-type (right) collagen-helix, as obtained by superimposition of GPVI on existing models of collagen peptides having PDB-IDs 1V4F⁴⁸ (7₂-helix) and 3DMW³⁷ (10₃-helix; residues 6-17). Only presented are models that are sterically feasible, with the GPVI-molecules shown in blue (static molecule) and salmon (mobile molecule) surface representation, and the collagen-helix in orange C α -traces and surfaces. Each model is displayed in both a side-projection and a projection along the triple helix, of which the latter also indicates which fraction (°) of its circumference is accessible. Models marked with * have an open fraction of 145° or more, which suggests that a particular binding mode also may occur in the context of a collagen fibril.

N-GDA**GAOGER**GPOGLAGAOLRGG**AGPOGP**EGGKGA**AGPOGP**OGAAGTOGLQ**GMOGER**GGLGSOGPK-C
 III-29 III-30 III-31 III-32

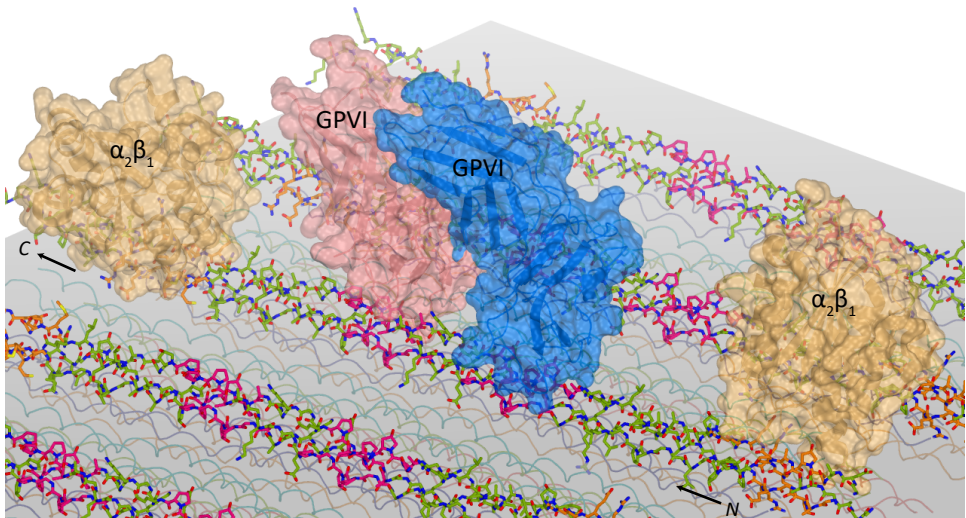
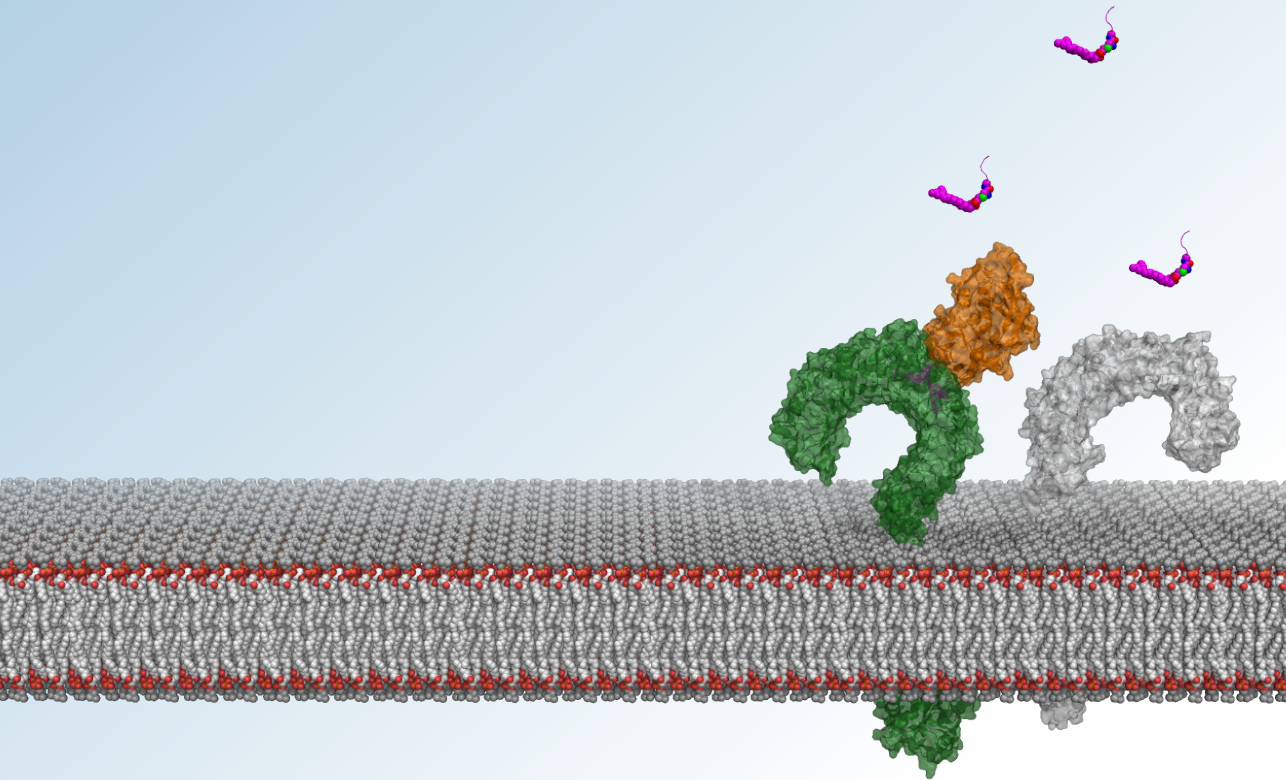


Figure S10. Hypothetical binding modes of GPVI and integrin $\alpha_2\beta_1$ to collagen fibrils. Fibril surface model of collagen type III with parallel collagen helices around the D3-segment covered by Toolkit-III peptides 29-32. Superimposition of GPVI shows that two molecules (blue and salmon) can bind without sterical hindrance to the adjacent AGPOGP-sites, also in addition to two $\alpha_2\beta_1$ -molecules of integrin $\alpha_2\beta_1$ to nearby GAOGER- and GMOGER-sites.

3



Structural basis for inhibition of TLR2 by staphylococcal superantigen-like protein 3 (SSL3)

Louris J. Feitsma^{a,1}, Kirsten J. Koymans^{b,1}, T. Harma C. Brondijk^a, Piet C. Aerts^b, Eddie Lukkien^a, Philip Lössl^c, Kok P.M. van Kessel^b, Carla J.C. de Haas^b, Jos A.G. van Strijp^{b,2}, Eric G. Huizinga^{a,2}

^a Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Department of Chemistry, Faculty of Science, Utrecht University, Utrecht, The Netherlands

^b Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands

^c Biomolecular Mass Spectrometry and Proteomics, Bijvoet Centre for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, and Netherlands Proteomics Center, Utrecht University, Utrecht, The Netherlands

Significance statement

Staphylococcus aureus secretes a range of virulence factors to evade immune recognition. One of these, Staphylococcal superantigen-like protein 3 (SSL3), disrupts an important component of our innate immune system: activation of Toll-like receptor 2 (TLR2) by bacterial lipopeptides. The crystal structure of the SSL3-TLR2 complex now provides the structural basis for a unique mechanism of full TLR2 antagonism in which SSL3 interferes with both ligand binding and receptor dimerization. Our novel insights on the host-pathogen interaction may contribute to vaccine development and form a starting point for the design of structure-based mimics to inhibit aberrant TLR2 activation in several inflammatory diseases and disease states.

Keywords: *S. aureus*; SSL3; Toll-like receptor; lipopeptide; immune evasion; innate immunity

Author contribution: K.J.K., L.J.F., T.H.C.B., P.L., K.P.M.v.K., C.J.C.d.H., J.A.G.v.S., and E.G.H. designed research; K.J.K., L.J.F., T.H.C.B., P.C.A., E.L., P.L., and C.J.C.d.H. performed research; K.J.K., L.J.F., P.L., K.P.M.v.K., C.J.C.d.H., J.A.G.v.S., and E.G.H. analyzed data; J.A.G.v.S. and E.G.H. performed supervision; and K.J.K., L.J.F., J.A.G.v.S., and E.G.H. wrote the paper.

Data deposition: The atomic coordinates and structure factors for SSL3ΔN and for the SSL3ΔN-mTLR2 complex have been deposited in the Protein Data Bank, www.rcsb.org (PDB ID codes 5D3D and 5D3I).

¹ L.J.F. and K.J.K. contributed equally to this work.

² J.A.G.v.S. and E.G.H. contributed equally to this work.

Toll-like receptors (TLRs) are crucial in innate recognition of invading micro-organisms and their subsequent clearance. Bacteria are not passive bystanders and have evolved complex evasion mechanisms. *S. aureus* secretes a potent TLR2 antagonist, Staphylococcal superantigen-like protein 3 (SSL3), which prevents receptor stimulation by pathogen-associated lipopeptides. Here, we present crystal structures of SSL3 and its complex with TLR2. The structure reveals that formation of the specific inhibitory complex is predominantly mediated by hydrophobic contacts between SSL3 and TLR2 and does not involve interaction of TLR2-glycans with the conserved Lewis^x binding site of SSL3. In the complex SSL3 partially covers the entrance to the lipopeptide binding pocket in TLR2 reducing its size by about 50%. We show that this is sufficient to inhibit binding of agonist Pam₂CSK₄ effectively, yet allows SSL3 to bind to an already formed TLR2-Pam₂CSK₄ complex. The binding site of SSL3 overlaps those of TLR2-dimerization partners TLR1 and TLR6 extensively. Combined, our data reveal a robust dual mechanism in which SSL3 interferes with TLR2 activation at two stages: by binding to TLR2, it blocks ligand binding and thus inhibits activation. Secondly, by interacting with an already formed TLR2-lipopeptide complex, it prevents TLR-heterodimerization and downstream signalling.

Introduction

In recent years, *S. aureus* has become a major health threat to both humans and domestic animals. It is found as a commensal bacterium in approximately 30% of the human population, but when it becomes infectious it can cause a wide diversity of diseases, ranging from mild skin infections to life-threatening invasive conditions such as pneumonia and sepsis¹. Increased antibiotic resistance and a high amount of virulence factors secreted by *S. aureus* contribute to its emergence as a pathogen. Amongst these secreted virulence factors are the staphylococcal superantigen-like proteins (SSLs), a family of 14 proteins located on two genomic clusters²⁻⁴. Recently, we and others identified SSL3 as a potent inhibitor of Toll-like receptor 2 (TLR2)^{5,6}, an innate immunity receptor which is a dominant factor in immune recognition of *S. aureus*⁷⁻¹⁰.

TLR2 belongs to a family of ten homologous innate immunity receptors that are activated by pathogen-associated molecular patterns (PAMPs)¹¹. TLR2 binds bacterial lipopeptides and lipoproteins. Subsequent formation of heterodimers with TLR1 or TLR6 leads to MYD88-dependent activation of the NF-κB pathway¹². TLR2 has dual ligand specificity that is determined by its dimerization partner; stimulation by diacyl lipopeptides from Gram-positive bacteria including *S. aureus* induces the formation of heterodimers with TLR6¹³, whereas triacyl lipopeptides from Gram-negative bacteria initiate formation of TLR2-TLR1 dimers¹⁴. The structural basis for lipopeptide specificity was revealed by crystal structures of TLR2-TLR1 and TLR2-TLR6 complexes with their respective lipopeptide analogs Pam₃CSK₄ and Pam₂CSK₄: TLR2 binds two lipid tails in a large hydrophobic pocket, whereas the third lipid tail of triacyl lipopeptides is accommodated by a smaller pocket present in TLR1, but not in TLR6^{15,16}.

The family of SSL proteins, including SSL3, share structural similarities to superantigens, but lack superantigenic activity. Interestingly, the functions that have been

discovered for SSLs so far have all been linked to immune evasion. SSL5 inhibits neutrophil extravasation^{17,18} and phagocyte function^{19,20}, SSL7 binds IgA and inhibits complement²¹, and SSL10 inhibits IgG1-mediated phagocytosis^{22,23}, blood coagulation²⁴ and the chemokine receptor CXCR4²⁵. In addition to SSL3, also weak TLR2 inhibitory activity was observed for SSL4⁵, but it remains unknown whether that is its dominant function. This variety of immunomodulatory molecules and functions reflects the importance of the different components of our innate immune system in the defense against *S. aureus*²⁶.

In this study we determined the crystal structures of SSL3 and the SSL3-TLR2 complex. In combination with mutagenesis and binding studies our data provide a novel working mechanism of a functional TLR2 antagonist.

Results

Structure of SSL3ΔN

To study the structural basis for inhibition of TLR2 activation by virulence factor SSL3, we expressed and purified SSL3ΔN, which lacks 133 N-terminal residues. Deletion of the N-terminal region proved essential to obtain crystals, but does not affect its activity towards TLR2 (Figure S1a). The crystal structure of SSL3ΔN, with two molecules in the asymmetric unit, was solved at 1.95 Å resolution (Figure S2a) by molecular replacement. SSL3 exhibits the characteristic two-domain fold of superantigens and other SSLs^{27,28}. The C-terminal β-grasp domain (residues 228-326) contains a V-shaped binding site for sialyl Lewis^x, which is conserved in SSL2-6, and -11 (Figure S3a-b)²⁸. The N-terminal OB-domain (residues 134-227) displays well-defined but markedly different conformations for loops β1-β2 and α3-β4 (Figure S2b-c). These conformational differences likely arise from crystal contacts, and suggest considerable flexibility of these loops in solution.

Structure of the SSL3ΔN-mTLR2 complex

To facilitate expression and crystallization of TLR2, previous structural studies utilized constructs in which the C-terminal cap domain (LRRCT) together with one leucine rich repeat (LRR) had been replaced by a fragment of a hagfish variable lymphocyte receptor (VLR)^{15,29}. We successfully produced a mouse TLR2 (mTLR2) construct covering the entire extracellular region of the protein and crystallized it in a 1:1 complex with SSL3ΔN. The structure was solved to 3.2 Å resolution (Figure 1) using molecular replacement with the structures of SSL3ΔN and the mTLR2-VLR fusion (PDB-ID: 2Z81)¹⁵.

Overall, the structures of TLR2 and SSL3 are well-defined (Figure S4a-b); the N- and C-terminal regions of TLR2, however, display increased average temperature factors. The LRRCT domain of TLR2 is structurally similar to that of TLR3 (Figure S4c-f), although 22 C-terminal residues appear disordered and could not be modeled satisfactorily. The observed flexibility of this region might, at least in part, account for the success of the VLR-fusion approach.

After refinement of the TLR2 and SSL3 structures, residual electron density in the lipid-binding pocket located between LRR 11 and 12 suggested the presence of a phospholipid (Figure S5a). Subsequent native mass spectrometry analysis of TLR2 detected a mixture of phosphatidylcholine (PC) lipids with acyl chain lengths varying between 12 and 20 (Figure S5b-h). Apparently, PC binds sufficiently tightly as to remain associated with TLR2 during

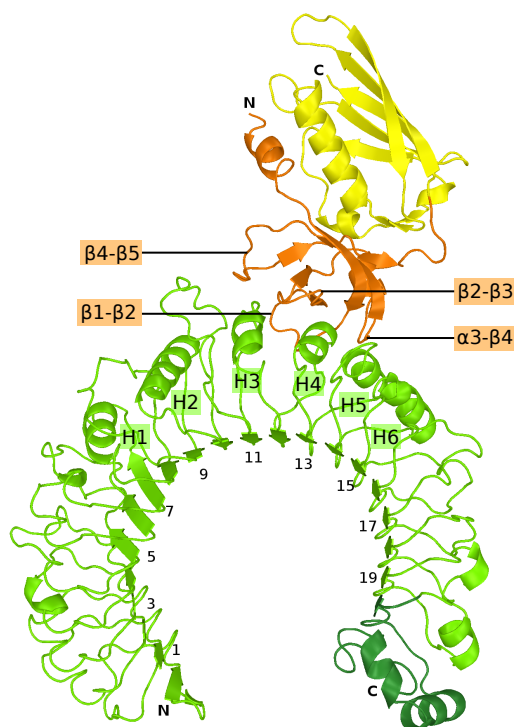


Figure 1. Crystal structure of SSL3 Δ N-mTLR2. Structure of the SSL3 Δ N-mTLR2 complex, with SSL3 Δ N OB- and β -grasp domains shown in respectively orange and yellow, mTLR2 in green, and the mTLR2 LRRCT-domain in a darker shade of green. Odd-numbered LRRs, helices H1-H6 of TLR2, and SSL3 loops that contact TLR2 are labeled.

The TLR2-SSL3 binding interface

The interface between SSL3 and TLR2 buries 1640 Å² of solvent accessible surface and is predominantly hydrophobic in nature. It consists of TLR2 residues located in LRR11 to LRR13, including helices H2 to H4, and SSL3 residues in four loops of the OB-domain as indicated in Figure 1. Three of these SSL3 loops differ in conformation compared to the structure of SSL3 alone (Figure S2d) suggesting that TLR2 binding is accompanied by considerable conformational changes in SSL3 (Figure S2e).

The SSL3 footprint on TLR2 is arc-shaped and surrounds three sides of the entrance to the lipopeptide binding pocket (Figure 2a). At one end of the arc, near helix H5, a continuous hydrophobic patch comprising SSL3 residues Phe156, Phe158, Leu160, and Pro194 interacts with TLR2 residues Phe349, Leu350, Gln375, Tyr376, and Asn379. In the center of the arc, a stretch of residues from the β 2- β 3 loop is positioned on top of TLR2 helices H3 and H4. Besides many hydrophobic interactions, this region contains the only hydrophilic interactions observed in the interface: Arg175 forms a salt bridge with Asp327, whereas hydrogen bonds are present between Arg175 and Ser329, and between Asn174 and His358. At the other end of the arc Trp163 stacks on Tyr323 in TLR2, whereas Leu211 and Lys213 have interactions with TLR2 residues Leu324 and Tyr326, respectively.

the purification process. The residual density in the lipid-binding pocket was subsequently modeled as PC, with its phosphoglycerol moiety positioned just inside, and its choline head group outside the pocket.

In the crystal structure of the SSL3-TLR2 complex, SSL3 binds with its OB-domain on the convex face of the characteristic horseshoe-like structure of TLR2 and partially covers the entrance of the lipopeptide binding pocket. Quantitative assessment of the SSL3-TLR2 interaction using the AlphaScreen assay³⁰ yields a binding affinity of 0.6 ± 0.4 nM (Figure S3c). The β -grasp domain of SSL3 does not contact TLR2. Its Lewis^x binding site is located more than 50 Å away from the nearest N-glycosylated asparagine in TLR2, a distance that cannot be bridged by a glycan antenna (Figure S3d). Formation of the TLR2-SSL3 complex does therefore not involve binding of TLR2 glycans to the Lewis^x binding site of SSL3, but is mediated by protein-protein interactions only.

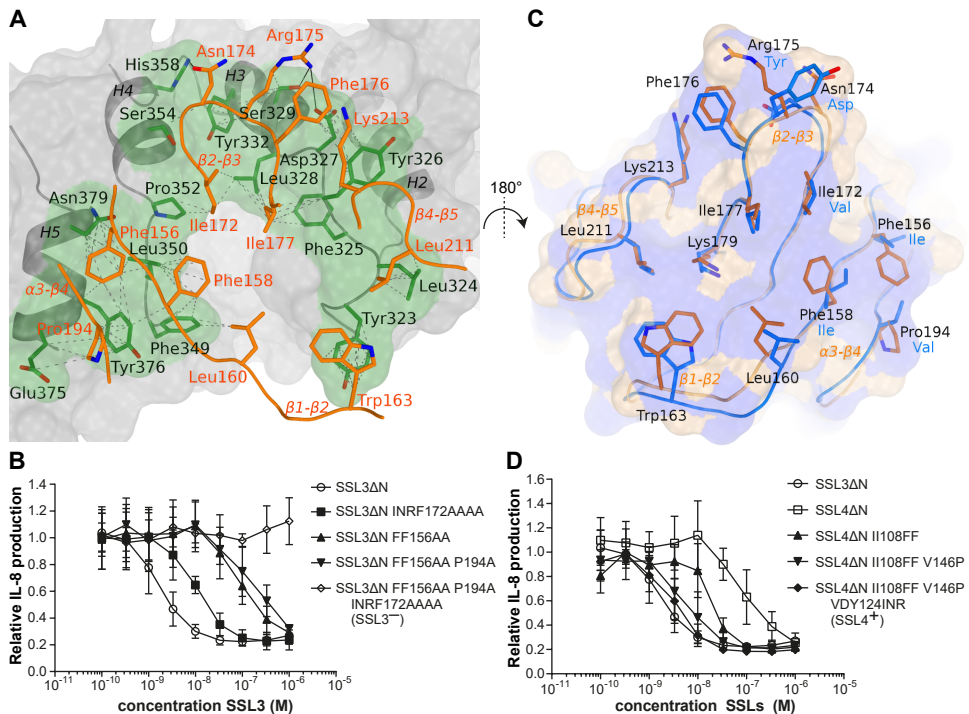


Figure 2. The SSL3-TLR2 interface and characterization of the TLR2-binding sites in SSL3 and SSL4. (A) Footprint of SSL3 (green) on the van der Waals surface of TLR2 (grey). Residues of SSL3 and TLR2 that are within 5 Å of its binding partner are shown in orange and green sticks representation, respectively. Van der Waals interactions are shown as dashed lines; hydrogen bonds and salt bridges as solid lines. (B) TLR2 inhibitory activity of SSL3 mutants. IL-8 production was measured after 6h of MALP2 (3 ng/ml) stimulation of HEK TLR2/6 cells and is expressed relative to cells not treated with SSL3. Data points represent the mean \pm SD of at least three independent experiments. (C) Comparison of the TLR2-binding site of SSL3 and the corresponding region of SSL4. Residues of SSL3 and SSL4 are shown in orange and blue sticks representation, respectively. Black labels refer to the SSL3 sequence; substitutions in SSL4 are labeled in blue. Also shown is the van der Waals surface of SSL3 with hydrophobic regions colored purple and hydrophilic regions colored wheat, emphasizing the hydrophobic nature of the TLR2 binding site. (D) TLR2 inhibitory activity of SSL4 mutants. Indicated amino acids of SSL4 were replaced by amino acids of SSL3. Data points represent the mean \pm SD of at least three independent experiments.

TLR2 residues that contact SSL3 in the crystal structure are conserved between mouse and human TLR2 (hTLR2), except for a single Ser354Leu substitution at the periphery of the binding site. Therefore, the structures of the human and mouse SSL3-TLR2 complexes are likely very similar.

Mutagenesis of SSL3 and SSL4

To confirm the binding site observed in the crystal structure we mutated SSL3 residues located in the interface to alanines (Figure 2a). The effect of mutation on inhibitory capacity was measured through IL-8 production after MALP-2 stimulation of HEK cells stably expressing human TLR2-TLR6. Single mutants showed no or only minor effects, with at most a 2-fold decrease in SSL3 activity (Figure S1b-d). Mutation of both Phe156 and Phe158 gave a 100-fold reduction of (Figure 2b). If in addition to Phe156 and Phe158 also nearby residue Pro194 was mutated, a further small decrease in activity was observed. Mutating a stretch of residues in loop β 2- β 3, Ile172, Asn174, Arg175, and Phe176,

3

resulted in a moderate 10-fold decrease in activity. Complete loss of SSL3 function could be achieved by combining mutations of the Phe156/Phe158/Pro194 patch and the β 2- β 3 stretch (SSL3⁻ in Figure 2b). Mutation of Trp163 and nearby residue Leu211 had no effect on SSL3 activity (Figure S1d), suggesting that this region of the interaction surface does not contribute significantly to TLR2 binding. It appears that strong SSL3-TLR2 binding is the sum of many, mainly hydrophobic, interactions in which residues Phe156 and Phe158 play a prominent role.

SSL3 and SSL4 show high sequential and structural homology, but substantially differ in their capacity to inhibit TLR2⁵. SSL3 residues important for TLR2 binding are poorly conserved in SSL4 (Figure 2c), which may explain the 100-fold less potency of SSL4 as a TLR2 inhibitor. The equivalent SSL4 residues in these OB-domain loops, however, are also predominantly hydrophobic, and suggest that TLR2 binding involves the same site in SSL4. Additionally, the main chain conformation of these loops in SSL4 is more similar to TLR2-bound SSL3 than free SSL3 itself (Figure S2d-e). To investigate the difference in inhibitory capacity between the two proteins, we replaced amino acids in SSL4 by their counterpart in SSL3. Replacement of both Ile108 and Ile110 by phenylalanine results in a 5-fold increased TLR2 inhibition (Figure 2d). Additional replacement of Val146 by proline enhances its function 20-fold as compared to SSL4. Replacement of the β 2- β 3 stretch (Val124Ile, Asp125Asn, Tyr126Arg) on top of this has a minor additional effect, and generates an SSL4 mutant with the potency of SSL3 (SSL4⁺ in Figure 2d). The observed gradual increase of SSL4 potency upon progressive introduction of SSL3 residues confirms that the TLR2 binding sites of SSL4 and SSL3 are located at equivalent sites.

SSL3 inhibits TLR dimerization and lipopeptide binding

TLR2 activation vitally depends on the binding of bacterial lipopeptides and subsequent formation of TLR2-TLR1 or TLR2-TLR6 heterodimers. The mechanism of TLR2 inhibition by SSL3 could involve interference in either or both of these steps. From our structural data presented here, it is directly evident that SSL3 blocks productive dimerization; SSL3 binding extensively overlaps with the region of TLR2 that is involved in dimerization with TLR6 (Figure 3a) as well as TLR1 (Figure S6a). Since dimerization is crucial for signaling, the functional consequence of SSL3 binding is that TLR2 stimulation by diacyl as well as triacyl lipopeptides is inhibited.

The structure of the SSL3-TLR2 complex furthermore suggests that binding of lipopeptides itself is inhibited, because SSL3 docks over the entrance to the ligand-binding pocket. However, an opening of about 5 by 9 Å remains in the SSL3-TLR2 interface (Figure 3b), which is about half of the original entrance size. In our AlphaScreen assay we observed concentration-dependent inhibition of Pam₂CSK₄-TLR2 binding by SSL3, whereas the loss of function mutant SSL3⁻ had no effect (Figure 3c). These data show that the observed size reduction of the pocket entrance upon binding of SSL3 effectively inhibits lipopeptide binding to TLR2.

SSL3 binds to the TLR2-Pam₂CSK₄ complex

Our observation of a PC molecule in the lipid-binding pocket of the SSL3-TLR2 complex shows that PC does not block SSL3-binding. The conformation of bound PC is noticeably similar to the previously observed binding modes for the synthetic phosphatidylethanolamine

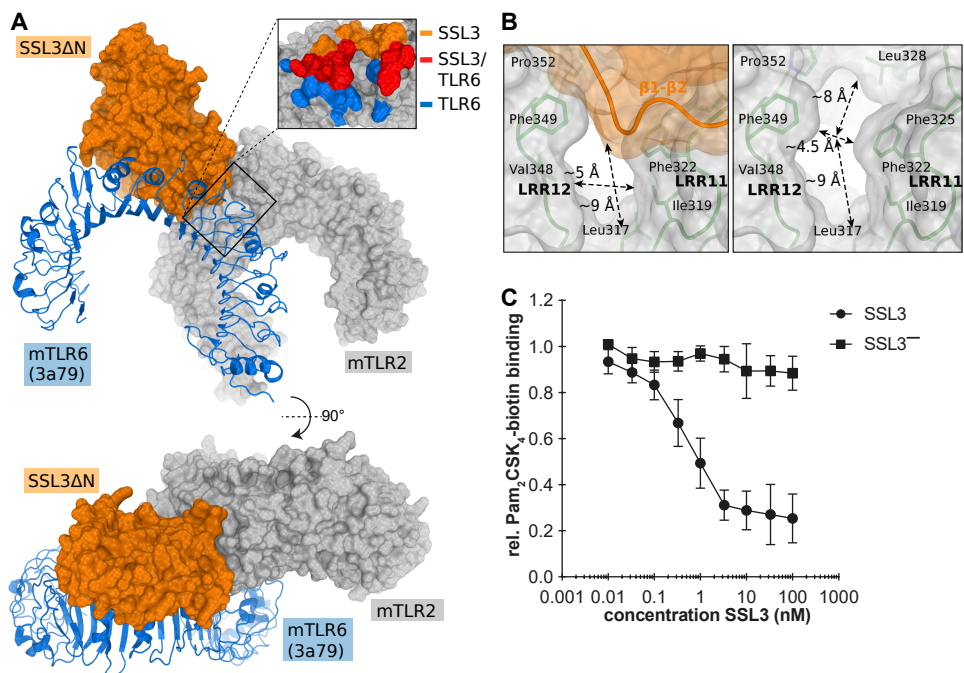


Figure 3. Inhibition mechanism of SSL3. (A) Hypothetical complex of TLR2 (gray surface), SSL3 (orange surface), and TLR6 (blue cartoon) as obtained by superposing SSL3-TLR2 and TLR2-TLR6 (PDB-code: 3A79)¹⁶. The inset shows TLR2 residues involved in binding to SSL3 (orange), TLR6 (blue), or both (red). **(B)** Dimensions of the entrance to the TLR2 lipopeptide binding pocket in the SSL3-TLR2 complex, measured in the presence (left) and absence (right) of SSL3. **(C)** AlphaScreen assay measuring the binding of Pam₂CSK₄-biotin to mTLR2-Fc fusion protein pre-incubated with different concentrations of SSL3 or SSL3⁻. Data is expressed relative to binding in absence of SSL3, and data points represent the mean ± SD of at least three independent experiments.

derivative PE-DTPA (Figure 4a and b) and saccharolipid lipoteichoic acid (*pn*LTA) from *S. pneumonia*¹⁶; ligands that have little or no ability to activate TLR2^{16,31,32}. In these complexes and our structure (ignoring the presence of SSL3), the lipopeptide binding pockets display similar open conformations and the conformations of *pn*LTA and PE-DTPA appear to be compatible with binding in the TLR2-SSL3 complex (Figure S6b-c).

These observations raise the question whether SSL3 can also bind if an activating ligand like Pam₂CSK₄ is present; a scenario that would enable SSL3 to block TLR2 signalling even after a bacterial ligand is engaged. The binding modes of non-activating ligands in TLR2 and Pam₂CSK₄ in the TLR2-TLR6 complex are, however, completely different. In the latter complex the TLR2- pocket is nearly closed due to a conformational change of LRR 10 and 11, and the glycerol moiety of the ligand is oriented differently with the head group cysteine bound in the so-called “sulfur site”¹⁶; a conformation that would not be compatible with SSL3 binding (Figure 4a and b).

To establish experimentally whether SSL3 is capable of binding a preformed TLR2-Pam₂CSK₄ complex, we used native PAGE and visualized the presence of bound lipopeptide with fluorescent Pam₂CSK₄-rhodamine. Addition of Pam₂CSK₄-rhodamine to TLR2 generates a fluorescent band at the same height as TLR2 alone (Figure 4c, panels 1 and 2). Incubation of TLR2 with SSL3 followed by the addition of Pam₂CSK₄-rhodamine results

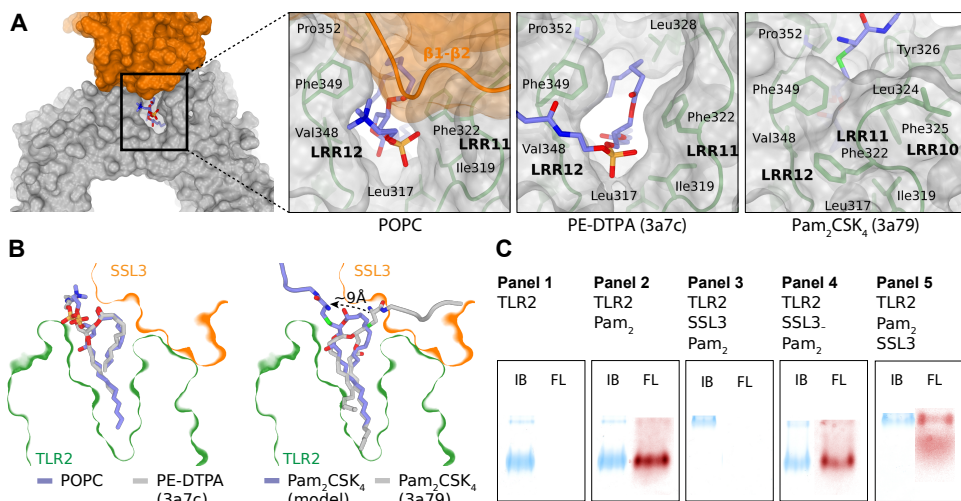


Figure 4. Binding of SSL3 to TLR2-lipid complexes. (A) Positioning of lipid head groups in the entrance to the TLR2 binding pocket: PC in the SSL3-TLR2 complex (left panel), PE-DTPA in TLR2 (middle panel; PDB-code 3A7C), Pam₂CSK₄ in TLR2-TLR6 complex (right panel, TLR6 not shown, PDB-code 3A79)¹⁶. (B) Cross-sections of the SSL3-TLR2 surface near the lipopeptide pocket with ligands from panel A in sticks representation: PC (blue, left), PE-DTPA (gray, left), and Pam₂CSK₄ (gray, right). Binding of SSL3 in the presence of Pam₂CSK₄ would require a substantial conformational change of its head group as shown in the modeled Pam₂CSK₄ (blue, right). (C) Native PAGE analysis of hTLR2 (panel 1) and hTLR2-complexes formed after incubation of hTLR2 (7 μ M) with Pam₂CSK₄Rhodamine (20 μ M; 18h at 37°C) and/or SSL3 (40 μ M; 30 min at 20°C) in the designated order (panel 2-5). Bands were visualized by Rhodamine fluorescence (FL, red) and subsequent staining with Instant Blue (IB, blue).

in the appearance of a more slowly migrating, non-fluorescent band containing the SSL3-TLR2 complex (panel 3) as was confirmed by in-gel digestion mass spectrometry, whereas no complex is formed with the loss of function mutant SSL3⁻ (panel 4). If, however, Pam₂CSK₄-rhodamine is allowed to bind TLR2 prior to addition of SSL3, we observe that the band corresponding to the SSL3-TLR2 complex is fluorescent (panel 5), implying the formation of a SSL3-TLR2-Pam₂CSK₄ triple complex. The existence of this triple complex was confirmed by native mass spectrometry (Figure S7a-d). Furthermore, binding of Pam₂CSK₄ to TLR2 does not affect association with SSL3 (Figure S7e). Therefore, SSL3 is indeed able to block TLR2 signalling after a bacterial ligand is engaged.

In view of the structural data presented above, TLR2 and Pam₂CSK₄ within the triple complex must adopt a conformation typically observed for TLR2 bound to non-activating ligands. Modelling shows that it is indeed possible to accommodate Pam₂CSK₄ in the SSL3-TLR2 complex (Figure 4b and S7f). Combined, our data show that SSL3 is able to interfere with TLR2 activation at two stages. First, its binding to TLR2 prevents lipopeptide binding; secondly, its binding to an already formed TLR2-lipopeptide complex prevents dimerization.

Discussion

Recognition of bacterial lipopeptides by TLR2 is critical for the defense against *S. aureus*. From the opposite perspective, inhibition of TLR2 by SSL3 is a powerful mechanism of *S. aureus* to survive inside its host. The crystal structure of the SSL3-TLR2 complex presented here shows that the highly hydrophobic binding interface is critically dependent on a set

of seven SSL3 residues with prominent roles for Phe156 and Phe158. This set of seven residues appears to be highly conserved among SSL3s from different *S. aureus* strains, but is absent in SSL4, the closest SSL3-relative within the SSL family and itself a weak TLR2 inhibitor. Introduction of these residues in SSL4 enhances its capacity to inhibit TLR2 to a similar level as SSL3 (Figure 2d). Interestingly, in strain MRSA252(SAR0425) these residues are present in SSL4, whereas they are not conserved in SSL3 (Figure S8), and, accordingly, SSL4 is the stronger TLR2 inhibitor⁵. Possibly, this strain underwent a genetic recombination event in which its overall capacity to evade TLR2 activation has been preserved, underlining the importance of TLR2 evasion.

Sialyl Lewis^x-dependent mechanisms have been described for functional activity of multiple SSL proteins, including SSL5 and SSL11^{27,28}. The sialyl Lewis^x binding site is fully conserved in SSL3, but its role in TLR2 inhibition has been unclear. SSL3 residue Arg308, previously described to be crucial for sialic acid binding, was found to be involved in, yet not crucial for, binding and activity of SSL3⁵. Yokoyama *et al.* reported that mutation of Phe297-Glu298, residues also involved in Lewis^x binding, results in decreased binding to cells, but has no effect on binding to TLR2 itself⁶. Our crystallographic data show that the distance from the Lewis^x binding site of SSL3 to the nearest N-linked glycosylation site in both mouse and human TLR2 is too large for interaction to occur (Figure S3d). Thus, glycan binding does not contribute directly to formation of the specific inhibitory complex, which is therefore exclusively mediated by protein-protein interactions. We hypothesize that the actual functional role of glycan binding is to increase the local SSL3 concentration on the immune cell surface, which is known to be rich in sialyl Lewis^x sugars³³; a pre-concentration step that would lead to more efficient TLR2 inhibition.

In this study we show that SSL3 interferes in TLR2 activation at two stages: first, SSL3 inhibits binding of bacterial lipopeptides, and, second, if a lipopeptide has already been engaged by TLR2, SSL3 prevents the formation of TLR2-TLR1 and TLR2-TLR6 heterodimers. A critical aspect of the SSL3-TLR2 complex that enables this dual mechanism is the opening to the lipopeptide binding pocket that remains after SSL3 binding. SSL3 only blocks about half of the pocket entrance and our experiments show that this is sufficient to inhibit lipid entry, but does allow for the accommodation of the head group of a lipopeptide that is already bound to TLR2 prior to SSL3 binding. Whereas this provides a functional role for the opening, it remains to be seen whether binding of SSL3 to a TLR2-lipopeptide complex is a prevalent pathway *in vivo*. Alternatively, the opening may also serve a different purpose, namely enabling the binding of SSL3 to TLR2-phospholipid complexes. It has not been established that TLR2 associates with phospholipids *in vivo*, however, the presence of copurified PC in our TLR2 preparation suggests that this may well be the case. In this scenario an opening to the binding pocket of TLR2 is required to prevent steric hindrance of non-activating phospholipids upon binding of SSL3.

Unraveling the mechanism of TLR2 inhibition by SSL3 gives new insights in the host-pathogen interaction and provides new tools to study TLR2 receptor biology. Aberrant TLR2 activation is linked to several diseases, including acute and chronic inflammatory conditions³⁴, making it an interesting therapeutic target. Our structural data provide a starting point for the development of SSL3 derivatives that could be used to block TLR2 activation in a therapeutic setting.

Materials and Methods

Expression and purification of SSL3 and SSL4 mutants

The *SSL3* and *SSL4* genes of *S. aureus* strain NCTC 8325 (SAOUHSC_00386 and SAOUHSC_00389) were used for construction of truncated proteins SSL3 Δ N comprising residues 134-326, SSL4 Δ N (residues 79-278), and mutants of SSL3 Δ N and SSL4 Δ N listed in Table S2. All variants were expressed with a non-cleavable N-terminal His₆-tag in *E. coli* Rosetta-gami(DE3)*pLysS*, refolded from insoluble fractions and purified as described⁵. Proteins were stored in PBS and protein purity was determined as >95% by SDS-PAGE.

For crystallization purposes SSL3 Δ N was expressed with a TEV cleavable N-terminal His₆-tag and isolated following the same procedure. TEV cleavage was performed overnight in 25 mM Tris-Cl buffer pH 8.2 and 150 mM NaCl. After addition of imidazole to a final concentration of 10 mM, TEV protease and any residual undigested SSL3 were removed by filtration through a HiTrap chelating HP column. SSL3 Δ N was ultimately purified by size exclusion chromatography over a Superdex75 column (GE-Healthcare) equilibrated in 10 mM Tris-Cl buffer pH 8.2 and 150 mM NaCl, and concentrated to 12 mg/mL.

Expression and purification of TLR2 ectodomains

Ectodomains of mouse (Gln25-Ala588, NM_011905) and human (Lys19-Ala589, NM_003264) TLR2 were transiently expressed with the N-terminal His₆-StrepII₃-TEV tag in HEK293-EBNA1-S and HEK293-EBNA-1 cells, respectively (U-protein Express BV, Utrecht, the Netherlands) as described⁵. Protein yields were optimized by plasmid titration³⁵, which indicated that transfections with 10-fold dilutions of expression plasmid in non-expressing dummy plasmid improved TLR2 production about two to three-fold. Further improvement of protein yield was achieved by co-transfecting a PRAT4A (NM_006586) expression plasmid at a ratio of 1:40. Crystallization experiments with mTLR2 were preceded by removal of the purification tag with TEV protease as described for SSL3 Δ N and gelfiltration on a pre-equilibrated Superdex200 column (GE-Healthcare) with 10 mM Tris-Cl buffer pH 8.2 and 150 mM NaCl.

Crystallization and data collection of SSL3 Δ N and the SSL3 Δ N-mTLR2 complex

SSL3 Δ N crystals were grown at 292 K using sitting drop vapor diffusion against a well solution containing 0.2 M potassium thiocyanate and 20% (w/v) PEG3350. Crystals were cryoprotected in well condition containing 20% (v/v) glycerol before flash-freezing in liquid nitrogen. Diffraction data to 1.94 Å resolution were collected at the Swiss Light Source on the PX beamline. For crystallization of the SSL3-TLR2 complex the individual proteins were mixed in a 1.1:1 molar ratio with final concentrations of 1.4 mg/mL and 3.8 mg/mL, respectively. Crystals were obtained through sitting drop vapor diffusion against a well solution containing 0.1 M PCB-buffer pH 5.0 (sodium propionate, sodium cacodylate, and Bis-Tris propane; Qiagen) and 25% (w/v) PEG1500. For data collection crystals were cryoprotected in well solution containing 20% glycerol (v/v) before flash-freezing in liquid nitrogen. X-ray diffraction data to 3.2 Å resolution were collected at the PETRAIII beamline (DESY, Hamburg). Details about structure determination and refinement procedures are included in *SI Materials and Methods*. Statistics of data processing and refinement are listed in Table S1.

Cell lines and transfections

Human embryonic kidney (HEK) cells expressing TLR2 and TLR6 were obtained from Invivogen and cultured in Dulbecco's Modified Eagle Medium (DMEM) in the presence of 10 µg/ml blasticidin, 100 units/ml penicillin, 100 µg/ml streptomycin, and 10% fetal calf serum.

Ligand-induced cytokine production

HEK-TLR2/6 cells were seeded in 96-wells culture plates. After reaching confluency, cells were incubated with the SSLs or SSL mutants for 30 min at 37°C. MALP-2 (Santa Cruz) was then added to a final concentration of 3 ng/ml. After 6 h culture supernatants were collected and tested for IL-8 production using specific enzyme-linked immunosorbent assay (ELISA), following manufacturer's instructions (Sanquin).

Binding studies

The AlphaScreen assay (Perkin Elmer Life Sciences)³⁰ was used to determine TLR2-ligand interactions. Murine TLR2-Fc (R&D systems), final concentration 9 nM, was mixed with a concentration range (0.01-100 nM) of SSL3ΔN or SSL3ΔN⁻ (FF156AA P194A INRF172AAAA) in PBS containing 0.05% human serum albumin. After 45 min, Pam₂CSK₄-biotin (Tocris) was added to a final concentration of 9 nM and incubation was continued for another 45 min. Next, 20 µg/ml of Streptavidin donor beads and 20 µg/ml Protein-G acceptor beads were added and incubated for 45 min. Samples were measured at 680 nm in a CLARIOstar microplate reader (BMG Labtech).

Native PAGE experiments were performed to study TLR2-ligand interactions as described by Jiménez-Dalmaroni *et al.*³⁶. Purified human TLR2 (7 µM, final concentration), in some cases pre-incubated with SSL3ΔN or SSL3ΔN⁻ (40 µM) for 30 min at room temperature, was mixed with Pam₂CSK₄Rhodamine (20 µM, Invivogen) and incubated for 18 h at 37°C. To examine whether ligand binding and SSL3 binding can simultaneously occur, first TLR2 and Pam₂CSK₄Rhodamine were allowed to bind for 18 h at 37°C after which SSL3 was added. Samples were loaded on 12.5% native glycine gels and run for 3 h at 200V. Rhodamine fluorescence was detected using a LAS 4010-imaging system (GE Healthcare) equipped with a 520 nm excitation LED and a 575-20BP emission filter. Subsequently, the gel was stained with Instant Blue protein stain (Expedeon).

Acknowledgements

This work was supported in part by the Dutch Top Institute Pharma Project D1-101, by ECHO grant 700.58.006 from the Council of Chemical Sciences of the Netherlands Organization for Scientific Research to E.G.H., by ManiFold project 317371 from the 7th Framework Programme of the European Union to the Bijvoet center, and by ZonMw grant 205200004 from the Netherlands Organization for Health Research and Development to J.A.G.S. We thank the Swiss Light Source (SLS), Villigen (CH) and DESY, Hamburg (DL) for providing data collection facilities and the beamline scientists for their help with data collection.

References

1. Lowy, F. D. Staphylococcus aureus infections. *N. Engl. J. Med.* **339**, 520–32 (1998).
2. Williams, R. J. *et al.* Identification of a Novel Gene Cluster Encoding Staphylococcal Exotoxin-Like Proteins: Characterization of the Prototypic Gene and Its Protein Product, SET1. *Infect. Immun.* **68**, 4407–4415 (2000).
3. Lina, G. *et al.* Standard nomenclature for the superantigens expressed by Staphylococcus. *J. Infect. Dis.* **189**, 2334–6 (2004).
4. Jongerius, I. *et al.* Staphylococcal complement evasion by various convertase-blocking molecules. *J. Exp. Med.* **204**, 2461–71 (2007).
5. Bardoel, B. W. *et al.* Evasion of Toll-like receptor 2 activation by staphylococcal superantigen-like protein 3. *J. Mol. Med.* **90**, 1109–1120 (2012).
6. Yokoyama, R. *et al.* Staphylococcal superantigen-like protein 3 binds to the Toll-like receptor 2 extracellular domain and inhibits cytokine production induced by Staphylococcus aureus, cell wall component, or lipopeptides in murine macrophages. *Infect. Immun.* **80**, 2816–25 (2012).
7. Bubeck Wardenburg, J., Williams, W. A. & Missiakas, D. Host defenses against Staphylococcus aureus infection require recognition of bacterial lipoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 13831–6 (2006).
8. Hashimoto, M. *et al.* Lipoprotein is a predominant Toll-like receptor 2 ligand in Staphylococcus aureus cell wall components. *Int. Immunol.* **18**, 355–62 (2006).
9. Yimin *et al.* Contribution of toll-like receptor 2 to the innate response against Staphylococcus aureus infection in mice. *PLoS One* **8**, e74287 (2013).
10. Takeuchi, O., Hoshino, K. & Akira, S. Cutting edge: TLR2-deficient and MyD88-deficient mice are highly susceptible to Staphylococcus aureus infection. *J. Immunol.* **165**, 5392–5396 (2000).
11. Gay, N. J., Symmons, M. F., Gangloff, M. & Bryant, C. E. Assembly and localization of Toll-like receptor signalling complexes. *Nat. Rev. Immunol.* **14**, 546–558 (2014).
12. O’Neill, L. a J. & Bowie, A. G. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat. Rev. Immunol.* **7**, 353–64 (2007).
13. Takeuchi, O. *et al.* Discrimination of bacterial lipoproteins by Toll-like receptor 6. *Int. Immunol.* **13**, 933–940 (2001).
14. Takeuchi, O. *et al.* Cutting Edge: Role of Toll-Like Receptor 1 in Mediating Immune Response to Microbial Lipoproteins. *J. Immunol.* **169**, 10–14 (2002).
15. Jin, M. S. *et al.* Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **130**, 1071–82 (2007).
16. Kang, J. Y. *et al.* Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity* **31**, 873–84 (2009).
17. Bestebroer, J. *et al.* Staphylococcal superantigen-like 5 binds PSGL-1 and inhibits P-selectin-mediated neutrophil rolling. *Blood* **109**, 2936–43 (2007).
18. Walenkamp, A. M. E. *et al.* Staphylococcal SSL5 binding to human leukemia cells inhibits cell adhesion to endothelial cells and platelets. *Cell. Oncol.* **32**, 1–10 (2010).
19. de Haas, C. J. C. *et al.* Staphylococcal superantigen-like 5 activates platelets and supports platelet adhesion under flow conditions, which involves glycoprotein Ibalpha and alpha IIb beta 3. *J. Thromb. Haemost.* **7**, 1867–74 (2009).
20. Bestebroer, J. *et al.* Staphylococcal SSL5 inhibits leukocyte activation by chemokines and anaphylatoxins. *Blood* **113**, 328–337 (2009).
21. Bestebroer, J. *et al.* Functional basis for complement evasion by staphylococcal superantigen-like 7. *Cell. Microbiol.* **12**, 1506–1516 (2010).
22. Patel, D., Wines, B. D., Langley, R. J. & Fraser, J. D. Specificity of staphylococcal superantigen-like protein 10 toward the human IgG1 Fc domain. *J. Immunol.* **184**, 6283–92 (2010).
23. Itoh, S. *et al.* Staphylococcal superantigen-like protein 10 (SSL10) binds to human immunoglobulin G (IgG) and inhibits complement activation via the classical pathway. *Mol. Immunol.* **47**, 932–938 (2010).

24. Itoh, S. *et al.* Staphylococcal superantigen-like protein 10 (SSL10) inhibits blood coagulation by binding to prothrombin and factor Xa via their γ -carboxyglutamic acid (Gla) domain. *J. Biol. Chem.* **288**, 21569–80 (2013).
25. Walenkamp, A. M. E. *et al.* Staphylococcal superantigen-like 10 inhibits CXCL12-induced human tumor cell migration. *Neoplasia* **11**, 333–44 (2009).
26. Bardoel, B. W. & Strijp, J. A. G. Molecular battle between host and bacterium: recognition in innate immunity. *J. Mol. Recognit.* **24**, 1077–86 (2013).
27. Chung, M. C. *et al.* The crystal structure of staphylococcal superantigen-like protein 11 in complex with sialyl Lewis X reveals the mechanism for cell binding and immune inhibition. *Mol. Microbiol.* **66**, 1342–55 (2007).
28. Baker, H. M. *et al.* Crystal structures of the staphylococcal toxin SSL5 in complex with sialyl Lewis X reveal a conserved binding site that shares common features with viral and bacterial sialic acid binding proteins. *J. Mol. Biol.* **374**, 1298–308 (2007).
29. Kim, H. M. *et al.* Structural diversity of the hagfish variable lymphocyte receptors. *J. Biol. Chem.* **282**, 6726–32 (2007).
30. Bosse, R., Illy, C., Chelsky, D. & Sciences, P. L. *Application Note - Principles of AlphaScreen.* (2002).
31. Han, S. H. *et al.* Pneumococcal Lipoteichoic Acid (LTA) Is Not as Potent as Staphylococcal LTA in Stimulating Toll-Like Receptor 2. *Infect. Immun.* **71**, 5541–5548 (2003).
32. Zähringer, U., Lindner, B., Inamura, S., Heine, H. & Alexander, C. TLR2 - promiscuous or specific? A critical re-evaluation of a receptor expressing apparent broad specificity. *Immunobiology* **213**, 205–24 (2008).
33. Munro, J. M. *et al.* Expression of sialyl-Lewis X, an E-selectin ligand, in inflammation, immune processes, and lymphoid tissues. *Am. J. Pathol.* **141**, 1397–1408 (1992).
34. Liu, Y., Yin, H., Zhao, M. & Lu, Q. TLR2 and TLR4 in Autoimmune Diseases: a Comprehensive Review. *Clin. Rev. Allergy Immunol.* **47**, 136–147 (2014).
35. Half, E. F., Versteeg, M., Brondijk, T. H. C. & Huizinga, E. G. When less becomes more: Optimization of protein expression in HEK293-EBNA1 cells using plasmid titration - A case study for NLRs. *Protein Expr. Purif.* **99**, 27–34 (2014).
36. Jiménez-Dalmaroni, M. J. *et al.* Soluble human TLR2 ectodomain binds diacylglycerol from microbial lipopeptides and glycolipids. *Innate Immun.* (2014). doi:10.1177/1753425914524077
37. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235–42 (2011).
38. Hermans, S. J. *et al.* Structural and functional properties of staphylococcal superantigen-like protein 4. *Infect. Immun.* **80**, 4004–4013 (2012).
39. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–21 (2010).
40. Bell, J. K. *et al.* The molecular structure of the Toll-like receptor 3 ligand-binding domain. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10976–80 (2005).
41. Van Den Heuvel, R. H. H. *et al.* Improving the performance of a quadrupole time-of-flight instrument for macromolecular mass spectrometry. *Anal. Chem.* **78**, 7473–7483 (2006).

Supplementary data

Materials and Methods

Data collection and refinement of the SSL3ΔN and SSL3ΔN-mTLR2 structures

Diffraction data to 1.94 Å obtained for crystals of SSL3ΔN was integrated with MOSFLM and further processed using CCP4 software³⁷. Integrated data were first scaled with Aimless. Molecular replacement was then performed with PHASER and the structure of SSL4 as initial search model (PDB-code: 4DXF)³⁸. The structure was refined with REFMAC alternated with manual model improvement using COOT. R_{work} and R_{free} had final values of 0.177 and 0.224. Statistics of data processing and refinement are listed in Table S1.

Diffraction data to 3.2 Å obtained for crystals of the SSL3ΔN-mTLR2 complex was integrated with XDS. The structure was solved by molecular replacement in PHASER using chain A of the SSL3ΔN structure and residues 27-506 of mTLR2 (PDB-code: 2Z81)¹⁵ as initial search models. Refinement of the model was performed with Phenix³⁹. The initial molecular replacement solution that includes the structures of SSL3ΔN and mTLR2₂₇₋₅₀₆ was further optimized by performing a rigid body refinement using separate rigid bodies comprising SSL3 residues 136-224 and 225-326 and TLR2 residues 27-250 and 251-506. Further refinement included TLS refinement using automatic TLS group identification. In view of the limited resolution of the diffraction data, external dihedral angle restraints derived from the structures of SSL3ΔN and the 1.8 Å resolution structure of mTLR2₂₇₋₅₀₆ were applied, except for regions where the electron density convincingly indicated a different conformation and partially modelled N-linked glycans connected to Asn414 and Asn442. One of these regions concerns loops in LRR10 and 11, which determine the shape and size of the lipopeptide pocket entrance. LRR11 contacts SSL3 and its conformation is well-defined. The long LRR10 loop is not involved in SSL3 binding and although it displays increased B-factors (Figure S4a), the electron density allowed building of the entire loop.

In the mTLR2-VLR6 fusion protein, LRR19 and the LRRCT domain of mTLR2 have been replaced by a fragment of hagfish VLR. Sequence identity between this VLR fragment and mTLR2 is only 11%. Therefore, we used the corresponding region of hTLR3 (PDB-code 2A0Z)⁴⁰, which is 25% identical (Figure S4f), to guide building of LRR19 and the LRRCT domain. The LRRCT domain harbors a CXC motive typical for C-cap domains of LRR-structures. Both cysteines are involved in disulfide bridge formation. Whereas the Cys537-Cys564 bridge could be modeled with confidence (Figure S4e), no electron density was visible for the Cys539-Cys585 bridge. Possibly, radiation damage suffered by this bond explains the poor ordering of this region, although it cannot be excluded that it is inherently flexible. TLR2 residues on the C-terminal side of Cys564 either were modeled as poly-alanine (residues 569-575), or are absent from the final model (residues 576-587).

Continuous residual density in the lipid binding pocket of TLR2 suggested the presence of a co-purified di-acetylated lipid (Figure S5a). Native mass spectrometry confirmed the presence of a lipid and identified it as phosphatidylcholine (PC) with acyl chain lengths varying from C12 to C20 (see Figure S5b-h). Accordingly, we modelled PC in the binding pocket with its phosphoglycerol moiety positioned in the pocket-opening and the choline headgroup located just outside, interacting with a chloride ion. The available electron density allowed modeling of C13 and C12 acyl chains at the *sn1* and *sn2* hydroxyl groups, respectively. The size of the TLR2 pocket, however, does not exclude binding of longer acyl chains. Refinement of the complex after inclusion of PC had a negligible effect on R_{work} and R_{free} that have final values of 0.239 and 0.287, respectively. Statistics of data processing and the final model are listed in Table S1.

Mass spectrometry

hTLR2 and Pam₂CSK₄-biotin were allowed to bind overnight at RT, before addition of SSL3. The final complex with equimolar concentrations of 12 μM was incubated at RT for 1 h. mTLR2 and the SSL3-hTLR2-Pam₂CSK₄-biotin complex were buffer exchanged to 150 mM ammonium acetate, pH 7.0 (for native MS analysis) or 0.1% formic acid solution (for MS under denaturing conditions). Buffer exchange was performed in 8 steps at 13,000 x g and 4°C using centrifugal filter units with a 30 kDa molecular weight cut-off (Millipore, UK). Subsequently, samples were diluted to a protein concentration of 3–5 μM and introduced into the mass spectrometer with gold-coated borosilicate needles to facilitate electrospray ionization. Denatured samples were analyzed on a modified LCT time-of-flight mass spectrometer (Waters, UK) with the instrument settings optimized as follows: capillary voltage 1100–1300 V, sample cone voltage 40–100 V, extraction cone voltage 0–20 V, source region pressure 4–6 mbar. Native MS was performed using a modified quadrupole time-of-flight instrument (Waters, UK)⁴¹, which was operated with the following parameters: capillary voltage 1300–1400 V, sample cone voltage 20–30 V, extraction cone voltage 0–10 V, collision cell voltage 10 V, source region pressure 8 mbar, collision gas pressure 0.01 mbar. For native tandem MS experiments, in-source activation was optimized by increasing sample and extraction cone voltages to 30–100V and 10–20 V, respectively. The collision cell voltage, which is applied after mass selection of the precursor ion, was varied to monitor the respective dissociation of phospholipids,

SSL3, or Pam₂CSK₄-biotin. All mass spectra were acquired in positive ion mode and calibrated using Csl clusters. Data analysis was performed with MassLynx v4.1 (Waters, UK).

Construction of BAP-tagged SSL3ΔN and *in vitro* biotinylation

A biotin acceptor peptide (BAP-tag; GLNDIFEAQKIEWHE) was incorporated by PCR in between the His-tag and the coding sequence of SSL3ΔN. Expression and purification was performed as described in the main materials and methods. SSL3-BAP was dialyzed to 50 mM Tris, 25 mM NaCl, pH 8.0. BAP-tagged SSL3 was biotinylated *in vitro* at 37°C for 4 h using a mixture with final concentrations of 22 μM protein, 10 mM ATP, 50 μM D-biotin (Sigma-Aldrich), 10 mM MgCl₂, and 0.44 μM BirA biotin-protein ligase (Avidity). Biotinylation was confirmed with Western Blot analysis using Streptactin-HRP (1:10.000, IBA).

Binding studies

AlphaScreen assay was performed for K_d determinations of the SSL3-TLR2 complex using a competitive binding format (30). mTLR2-Fc (0.1 nM, final concentration) was simultaneously incubated with biotin-labeled BAP-SSL3 (1 nM) and different concentrations of unlabeled SSL3 (ranging from 0.01 nM to 100 nM) for 45 min. Binding was determined by addition of 20 μg/ml Streptavidin donor beads and 20 μg/ml Protein-G acceptor beads. To examine the effect of TLR2-ligand binding, Pam₂CSK₄ (1 nM) was pre-incubated with TLR2 for 45 min before the competitive binding assay was performed.

Tables & Figures

Table S1. Data collection and refinement statistics

	SSL3	TLR2-SSL3
Data Collection		
Wavelength (Å)	0.9999	1.0000
Space group	P2 ₁ 2 ₁ 2 ₁	P6 ₁
Cell dimensions		
a, b, c (Å)	58.00, 79.22, 96.38	104.20, 104.20, 153.63
α, β, γ (°)	90, 90, 90	90, 90, 120
Resolution range (Å)*	61.2 - 1.94 (1.99 - 1.94)	52.1 - 3.2 (3.31 - 3.2)
Total no. reflections	223,907 (16,027)	36,297 (7,265)
No. unique reflections	33,556 (2,419)	14,687 (3,088)
R _{merge}	0.093 (0.73)	0.225 (1.08)
I/σI	11.1 (2.7)	5.6 (2.0)
Redundancy	6.7 (6.6)	2.5 (2.4)
Completeness (%)	99.8 (99.9)	94.1 (96.7)
CC(1/2)	0.99 (0.63)	0.92 (0.41)
Refinement		
R _{work} / R _{free}	0.177 / 0.224	0.231 / 0.271
No. atoms	3,521	5,976
Protein	3,266	5,890
Water / other ligands	214 / 9	0 / 86
Average B / Wilson B (Å ²)	37.5 / 32.2	57.4 / 44.6
RMS deviations		
Bond lengths (Å)	0.0079	0.0070
Bond angles (°)	1.21	1.23
Ramachandran Plot		
Favored (%)	97.0	92.8
Allowed (%)	3.0	6.7
Outliers (%)	0	0.5

*Numbers between brackets refer to the outer resolution shell.

Table S2. List of all SSL3 and SSL4 mutants and their IC50 values

	Mutations	IC50 values (M)
SSL3 mutants		
SSL3	-	$2.2 \pm 0.7 \times 10^{-9}$
SSL3ΔN	Δ1-133	$1.9 \pm 0.8 \times 10^{-9}$
SSL3ΔN N174A	N174A	$3.1 \pm 1.1 \times 10^{-9}$
SSL3ΔN R175A	R175A	$1.5 \pm 0.7 \times 10^{-9}$
SSL3ΔN NR174AA	N174A, R175A	$3.3 \pm 2.1 \times 10^{-9}$
SSL3ΔN INRF172AAAA	I172A, N174A, R175A, F176A	$1.2 \pm 0.3 \times 10^{-8}$
SSL3ΔN F156A	F156A	$4.4 \pm 2.2 \times 10^{-9}$
SSL3ΔN F158A	F158A	$2.6 \pm 0.6 \times 10^{-9}$
SSL3ΔN P194A	P194A	$2.0 \pm 0.5 \times 10^{-9}$
SSL3ΔN W163A	W163A	$2.4 \pm 0.5 \times 10^{-9}$
SSL3ΔN W163A L211A	W163A, L211A	$5.8 \pm 3.4 \times 10^{-9}$
SSL3ΔN FF156AA	F156A, F158A	$1.3 \pm 0.5 \times 10^{-7}$
SSL3ΔN FF156AA P194A	F156A, F158A, P194A	$7.3 \pm 10.7 \times 10^{-6}$
SSL3ΔN FF156AA P194A INRF172AAAA	F156A, F158A, P194A, I172A, N174A, R175A, F176A	mutant does not reach IC50
SSL4 mutants		
SSL4ΔN	Δ1-78	$1.7 \pm 1.3 \times 10^{-7}$
SSL4 II108FF	I108F, I110F	$2.3 \pm 1.2 \times 10^{-8}$
SSL4 II108FF V146P	I108F, I110F, V146P	$7.6 \pm 5.9 \times 10^{-9}$
SSL4 II108FF V146P VDY124INR	I108F, I110F, V146P, V124I, D126N, Y127R	$3.6 \pm 1.7 \times 10^{-9}$

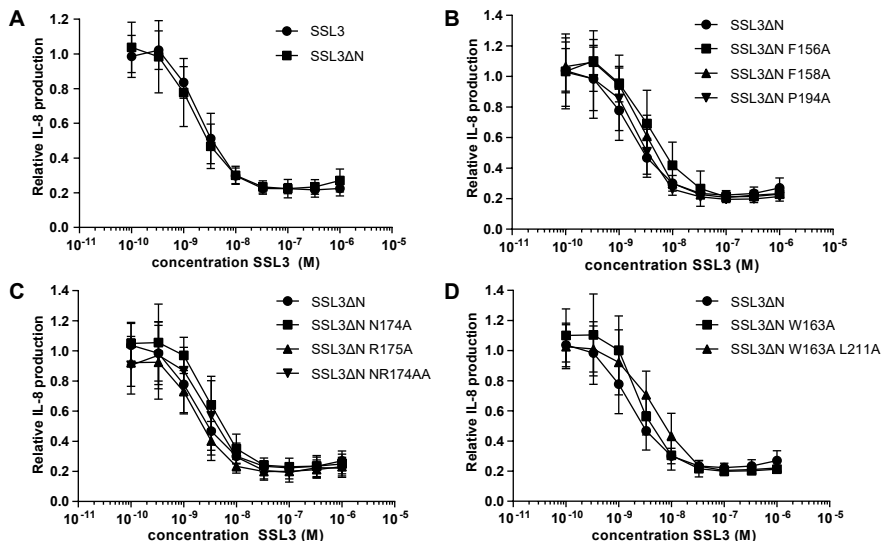


Figure S1: TLR2 inhibition by SSL3 WT, SSL3ΔN, and SSL3ΔN point mutants. HEK TLR2/6 cells were stimulated with MALP-2 and IL-8 production was measured in the presence of increasing concentrations of SSL3 or SSL3ΔN. IL-8 production is relative to cells not treated with SSL3. Data points represent at least three independent experiments and error bars show mean \pm standard deviation (SD). **(A)** SSL3 and SSL3ΔN, which lacks amino-terminal residues 1-133, have equal TLR2 inhibitory capacities. **(B-D)** IL-8 production in the presence of SSL3ΔN mutants that did not have a significant effect on TLR2 inhibitory capacity.

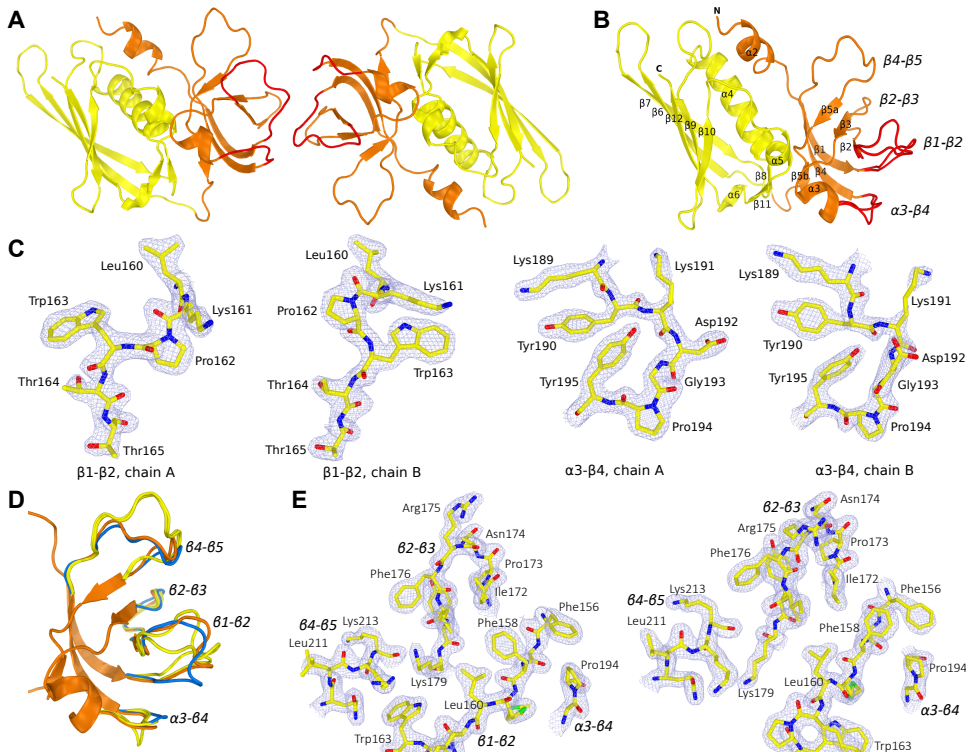


Figure S2: Structure of SSL3ΔN shows structural variation in loops of the β-grasp domain. (A) Cartoon visualization of the SSL3ΔN asymmetric unit that contains two molecules SSL3, both comprising residues 134-326. Two SSL3 molecules contact via three loops (β1-β2, β2-β3, and α3-β4) in the OB-domain (orange); an arrangement that is clearly different from the previously observed SSL dimers formed by antiparallel packing of two β-grasp domains (yellow)²⁸. (B) Cartoon visualization of the SSL3ΔN structure showing the N-terminal OB-domain (residues 134-227, orange) and the C-terminal β-grasp domain (residues 228-326, yellow). The OB-domain comprises a five-stranded β-barrel and two short α-helices, and two conformations observed for the β1-β2 and α3-β4 loops are indicated in red. The β-grasp domain contains a central α-helix packed against a mixed five-stranded β-sheet. Residues in strand β10 and helix α5 contribute to a V-shaped binding site for sialyl Lewis^x containing glycans. (C) Two conformations of loop β1-β2 (panel 1 and 2) and loop α3-β4 (panel 3 and 4) that are observed in the crystal structure of SSL3 shown as sticks representation in their corresponding 2Fo-Fc maps contoured at 1.2 σ above the average density. Except for the Lys191 side chain, these loops are well-defined in the electron density maps, and differences likely arise from their involvement in crystal packing. (D) Structural variation of the β1-β2, α3-β4, and β4-β5 loops in superimposed structures of 'free' SSL3 (SSL3A and B, yellow), SSL4 (blue; PDB-code: 4DXF³⁸), and the SSL3-TLR2 complex (orange). Whereas free SSL3 differs substantially from TLR2-bound SSL3, the structure of SSL4 shows much smaller differences. (E) Structural variation of the TLR2-binding site as observed in the SSL3-TLR2 complex (orange; bottom left) and the corresponding region in free SSL3A (yellow, top left), SSL3B (yellow, top right), and SSL4 (blue, bottom right), all shown as sticks representation in their corresponding 2Fo-Fc maps contoured at 1.2 σ above the average density. The orientation of the displayed regions is equal to Figure 2C.

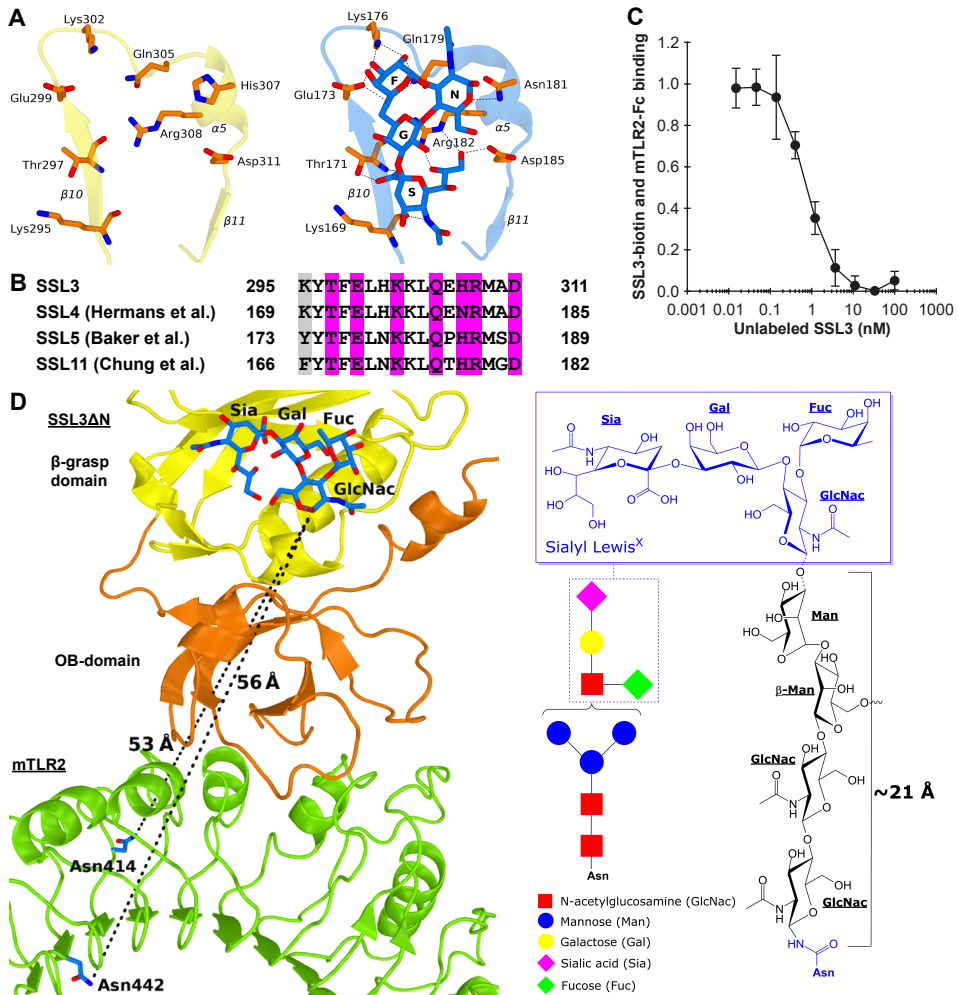


Figure S3: Nanomolar binding of SSL3 to TLR2 does not involve sialyl Lewis^x interactions. (A) Sticks representation of residues that are observed to bind sialyl Lewis^x in SSL4 (right; PDB-code: 4DXG³⁸) and the corresponding region of SSL3 (left). Sialyl Lewis^x sugar moieties are labeled: sialic acid (S), galactose (G), N-acetylglucosamine (N), and fucose (F). Hydrogen bonds between SSL4 residues and Sialyl Lewis^x are indicated with dashed lines. (B) Sequence alignment of SSL3, SSL4, SSL5, and SSL11 shows the conserved sialyl Lewis^x binding pattern (TxExxKxxQx[N/H]RxxD; purple)^{27,28,38}. Residues highlighted in gray only contribute to the hydrogen bonding network through backbone atoms, and therefore these interactions are sequence independent. (C) SSL3 and TLR2 have nanomolar affinity. In an AlphaScreen competitive binding assay unlabeled SSL3 was titrated in and competes with biotin-labeled SSL3 for TLR2-Fc binding. The IC₅₀ value corresponds to a K_d of 0.6 ± 0.4 nM. Data is expressed relative to binding of SSL3-biotin to TLR2 with no unlabeled SSL3 present. Data points represent the mean ± SD of at least three independent experiments. (D) Geometric analysis of the feasibility of sialyl Lewis^x interactions in the SSL3-TLR2 complex. (Left) Sialyl Lewis^x was positioned in the SSL3-TLR2 complex by superimposing the Lewis^x binding sites of SSL3 and the SSL4-sialyl Lewis^x complex (PDB-code: 4DXG³⁸). Indicated are the shortest distance between the Lewis^x glycan and the two nearest glycosylated asparagine residues in TLR2 (shown in blue sticks representation). (Right) Chemical structure of a N-linked glycan including a mannosylated sialyl Lewis^x (GlcNac(-Fuc)-Gal-Sia). The linker between the asparagine and Lewis^x consists of two N-acetylglucosamine and two mannose units that can span a distance of at most 21 Å in a fully extended conformation, as was estimated on the basis of crystal structures in the protein data bank that contain this glycan motif. This distance is two-fold smaller than the distances to the nearest N-linked asparagines in TLR2, excluding the involvement of sialyl Lewis^x binding in the observed SSL3-TLR2 complex.

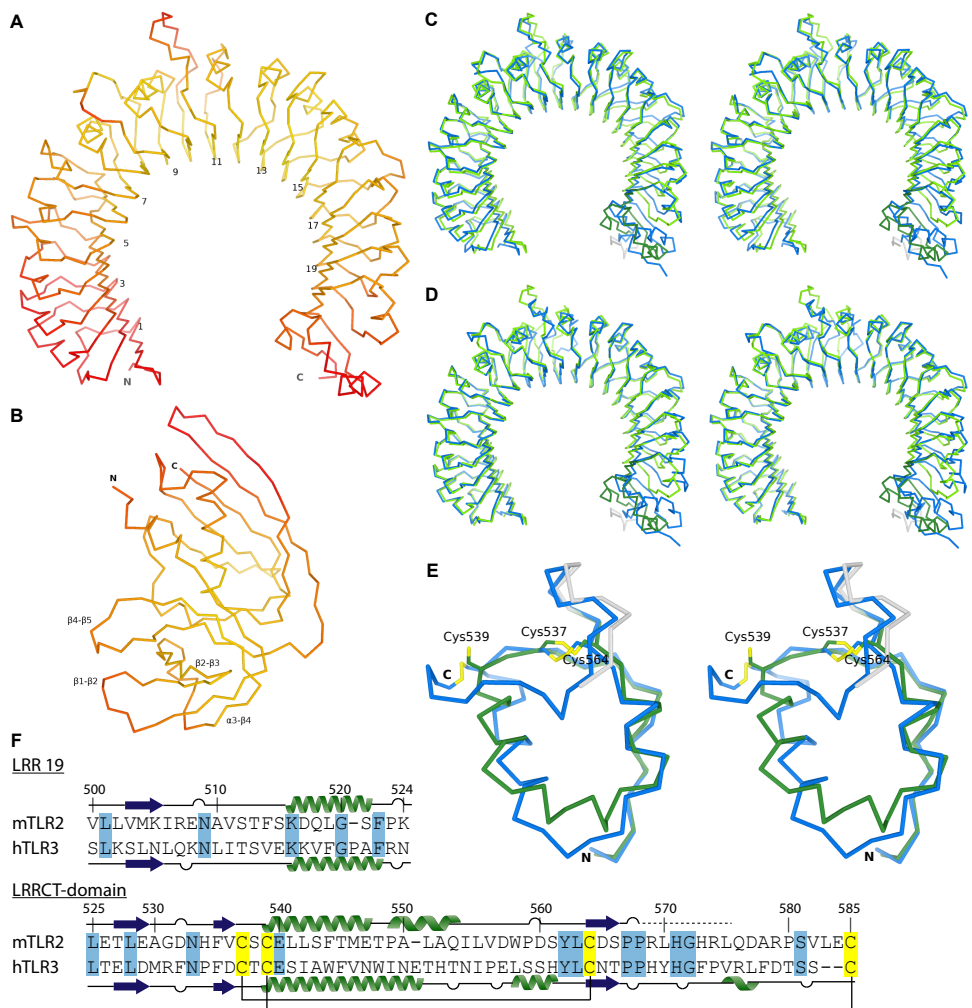


Figure S4: The SSL3-TLR2 complex. (A-B) α -traces of TLR2 (A) and SSL3 (B) in the SSL3-TLR2 complex colored according to their atomic B-factor on a scale between 20 \AA^2 (yellow) and 100 \AA^2 (red). (C-D) Superpositions of mouse TLR2 from the SSL3-TLR2 complex (green) on the TLR2-VLR6 fusion proteins used for crystallization of the TLR2-PE-DTPA complex (blue, C) and the TLR2-TLR6-Pam₂CSK₄ complex (blue, D)¹⁶. The structures are displayed as α -traces in stereo view. Overall the structures are very similar. The native mTLR2 LRRCT-domain, however, deviates significantly from the chimeric VLR6 LRRCT-domain, which share 11% sequence identity. The TLR2 structures also show conformational variations in LRR10, and the extended loop of LRR 11; two regions that determine the size and shape of the lipopeptide pocket entrance. The conformation of these loops in the SSL3-TLR2 complex is very similar to those observed for TLR2 in complex with non-activating ligand PE-DTPA (C), and clearly different from the TLR2-TLR6-Pam₂CSK₄ complex (D). (E) Overlay of the LRRCT-domains of mTLR2 (green and gray) and hTLR3 (blue)⁴⁰, shown in stereo view with the conserved cysteine-pattern highlighted in yellow. The TLR2 region that was modelled as poly-alanine is shown in gray. The overall RMSD difference between the two LRRCT-domains, which have 25% sequence identity, is 1.1 \AA . The largest observed difference comprises the 13 residue α -helix in TLR3 (res. 651-663), which is split in a shorter 9-residue α -helix (res. 539-547) and a 5-residue 3_{10} helix (res. 549-553) in TLR2. Residues 576-587 are absent in the TLR2 model, including the disulfide bridge between Cys539 and Cys585, which suggest that this region suffered from radiation damage or is intrinsically flexible. (F) Sequence alignment of LRR19 and the LRRCT-domain from mTLR2 and hTLR3. Also shown is the secondary structure as observed in the SSL3-TLR2 complex and hTLR3 (PDB-code: 2A0Z⁴⁰). Residues in the SSL3-TLR2 complex that were modelled as poly-alanine are indicated by a dashed line. Identical residues between TLR2 and TLR3 are highlighted in blue, and the conserved cysteines in yellow.

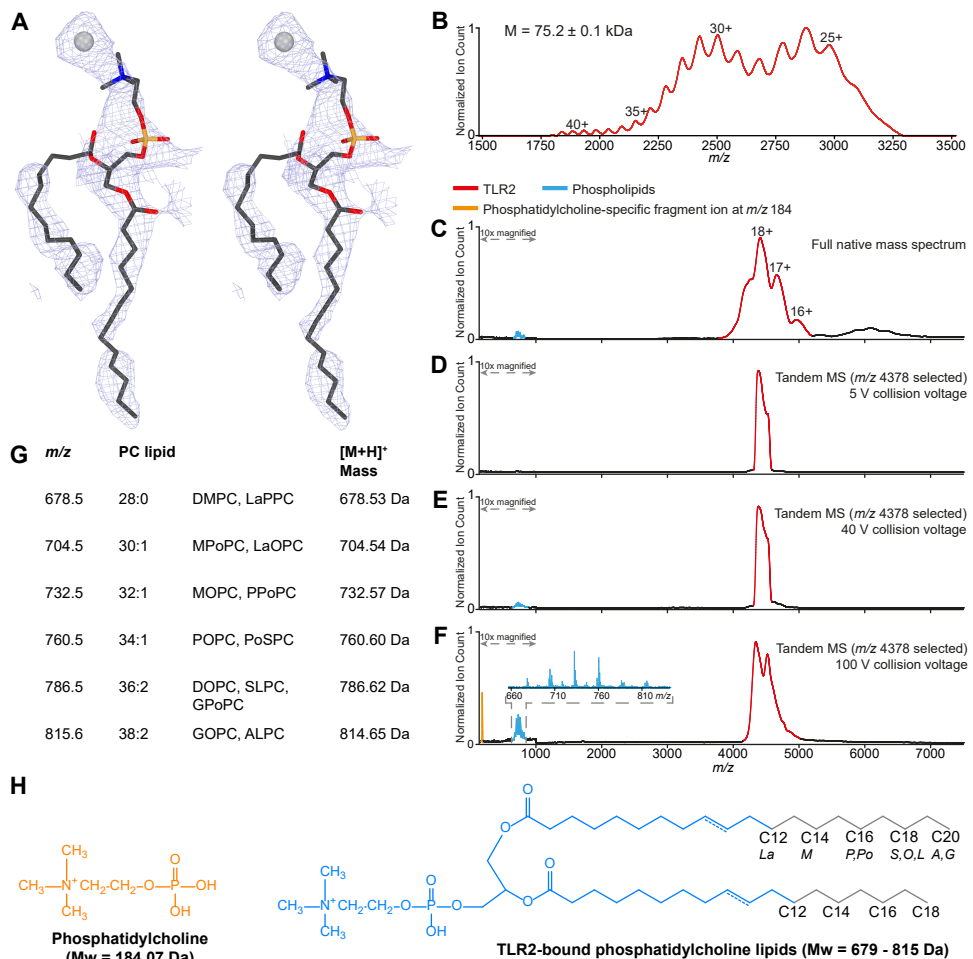


Figure S5: Phosphatidylcholine lipids in the TLR2 lipopeptide binding pocket. (A) Stereo view of the 2Fo-Fc map calculated using phases derived from the SSL3-TLR2 structure without ligand, and contoured at 0.7σ . Also shown are sticks representations of the phosphatidylcholine (PC) lipid ligand with C12 and C13 acyl chains and a chloride ion that were modeled in the electron density. Carbon, oxygen, nitrogen, and phosphorus atoms are respectively colored gray, red, blue, and orange; the chloride ion is shown as a gray sphere. (B) Baseline-corrected mass spectrum of denatured TLR2 with native glycans. TLR2 forms a broad charge state envelope, based on which an average molecular weight of 75.2 ± 0.1 kDa was calculated. Charge states are indicated above the peaks. (C) Native MS analysis of TLR2. Broad-range mass spectrum of native TLR2 (red) with charge states indicated above the peaks. Due to the existence of different TLR2 glycoforms and/or binding of different phospholipid species, the m/z signals are relatively broad, allowing only an approximate molecular weight calculation (79.1 ± 0.1 kDa). Free phospholipids (cyan) are visible in the low m/z region. (D-F) Native tandem MS analysis of TLR2. The region around m/z 4378 was mass selected in the quadrupole mass analyzer and subsequently fragmented by collision induced dissociation in the collision cell. Shown are the resulting tandem mass spectra. (D) At low collisional activation only phospholipid-bound TLR2 is detected. (E) Increasing the collision to 40 V leads to dissociation of phospholipids from TLR2. (F) Applying a collision voltage of 100 V results in enhanced phospholipid dissociation and fragmentation. The fragment ion at m/z 184 (orange) is characteristic for fragmentation of the phosphatidylcholine head group. The inset shows a zoom-in of the m/z region where phospholipids are detected. (G) List of the observed m/z peaks (660-830 m/z) measured in the tandem MS analysis showing a range of TLR2-bound PC lipid species with two acyl chains that have a total of

28-38 carbon atoms, 0, 1, or 2 unsaturations, and a $[M+H]^+$ mass for the single charged ions as depicted in the last column. Lipid chain abbreviations used: *La*, Lauroyl (12:0); *M*, Myristoyl (14:0); *P*, Palmitoyl (16:0); *Po*, Palmitoleoyl (16:1); *S*, Stearoyl (18:0); *O*, Oleoyl (18:1); *L*, Linoleoyl (18:2); *A*, Arachidoyl (20:0); *G*, Gadoleoyl (20:1). **(H)** Chemical structures of the phosphatidylcholine fragment ion (orange) and the TLR2-bound PC lipids with the consensus part (blue) and variable lipid tail length (gray). For ready comparison, all phosphatidylcholine structures and masses depicted here correspond to single charge (+1) ions.

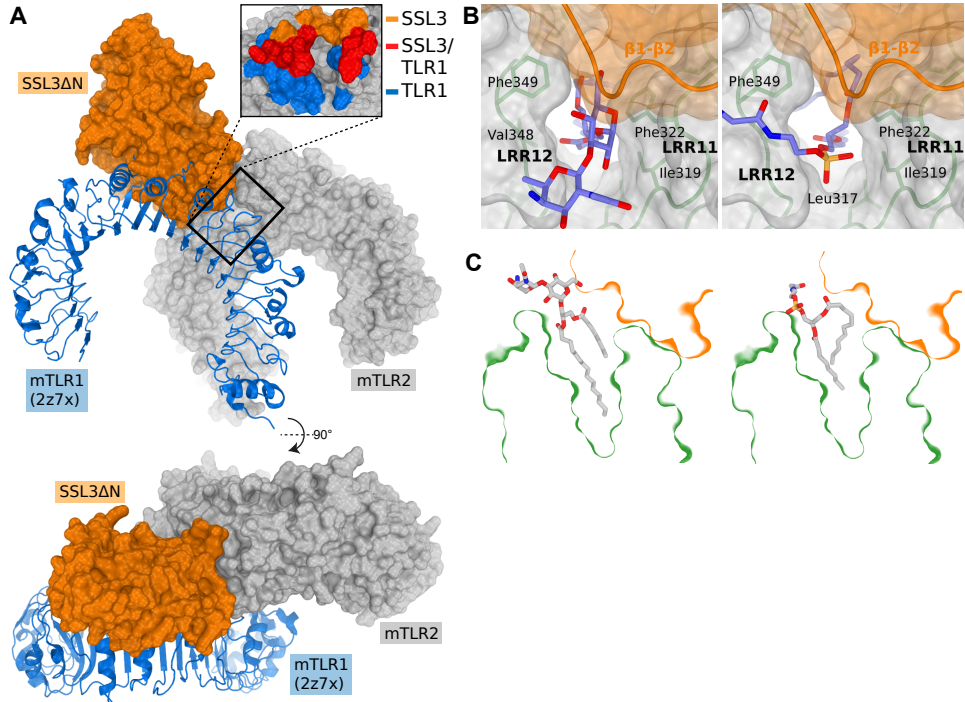


Figure S6: Binding of *pnLTA* and PE-DTPA, but not TLR1 is compatible with binding to the SSL3-TLR2 complex. **(A)** Surface representation of the SSL3-TLR2 complex (orange and gray) and a cartoon view of TLR1 (blue), as obtained by superposing the SSL3-TLR2 and TLR2-TLR1 (PDB-code: 2Z7X¹⁵) crystal structures. The inset shows TLR2 residues involved in binding to SSL3 (orange), TLR1 (blue), or both (red). Overlapping binding sites demonstrate that simultaneous binding of SSL3 and TLR1 to TLR2 is excluded. **(B)** Superposition of the TLR2-*pnLTA* (left) and TLR2-PE-DTPA (right) structures on the SSL3-TLR2 complex places the head groups of both lipids in the opening to the lipopeptide binding pocket. TLR2 and SSL3 are respectively shown in gray and orange surface representation, the lipids in blue sticks representation. The TLR2-*pnLTA* (PDB-code: 3A7B) and TLR2-PE-DTPA (PDB-code: 3A7C) structures have been obtained from Kang et al.¹⁶. Superposition of the observed conformations of the *pnLTA* and PE-DTPA ligands in the SSL3-TLR2 structure is possible without severe clashes, indicating that both ligands can be accommodated in a SSL3-TLR2 complex. **(C)** Cross-section of the SSL3-TLR2 complex, including the TLR2-bound conformations of *pnLTA* (left) and PE-DTPA (right), showing that both lipids can be placed in the SSL3-TLR2 complex without severe clashes.

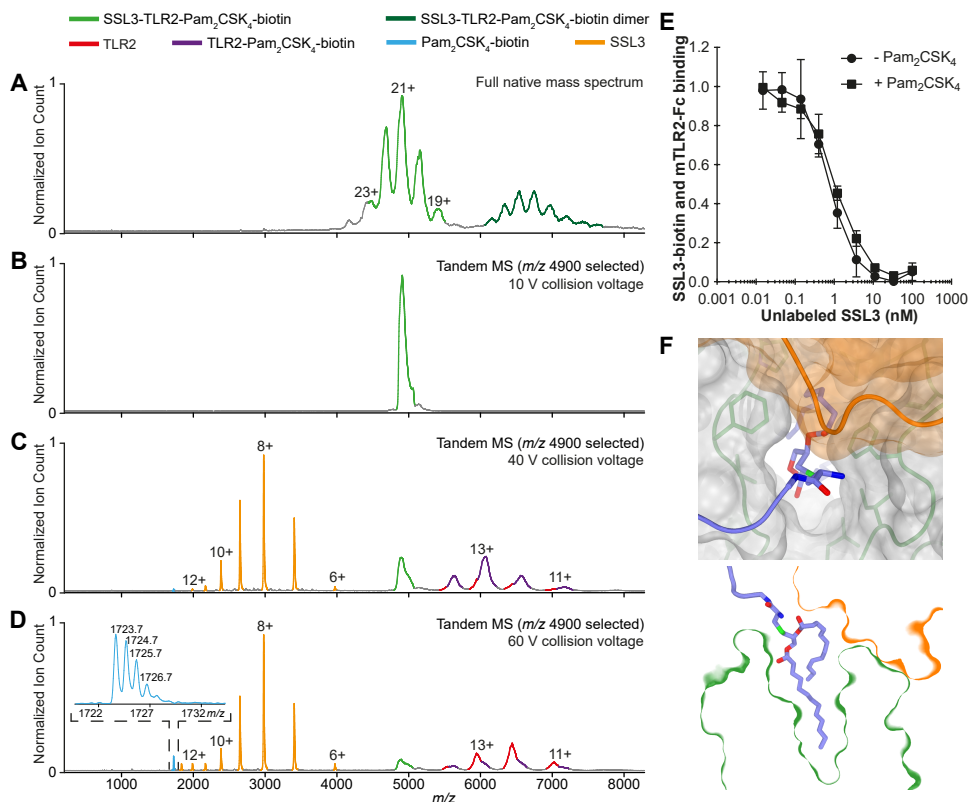
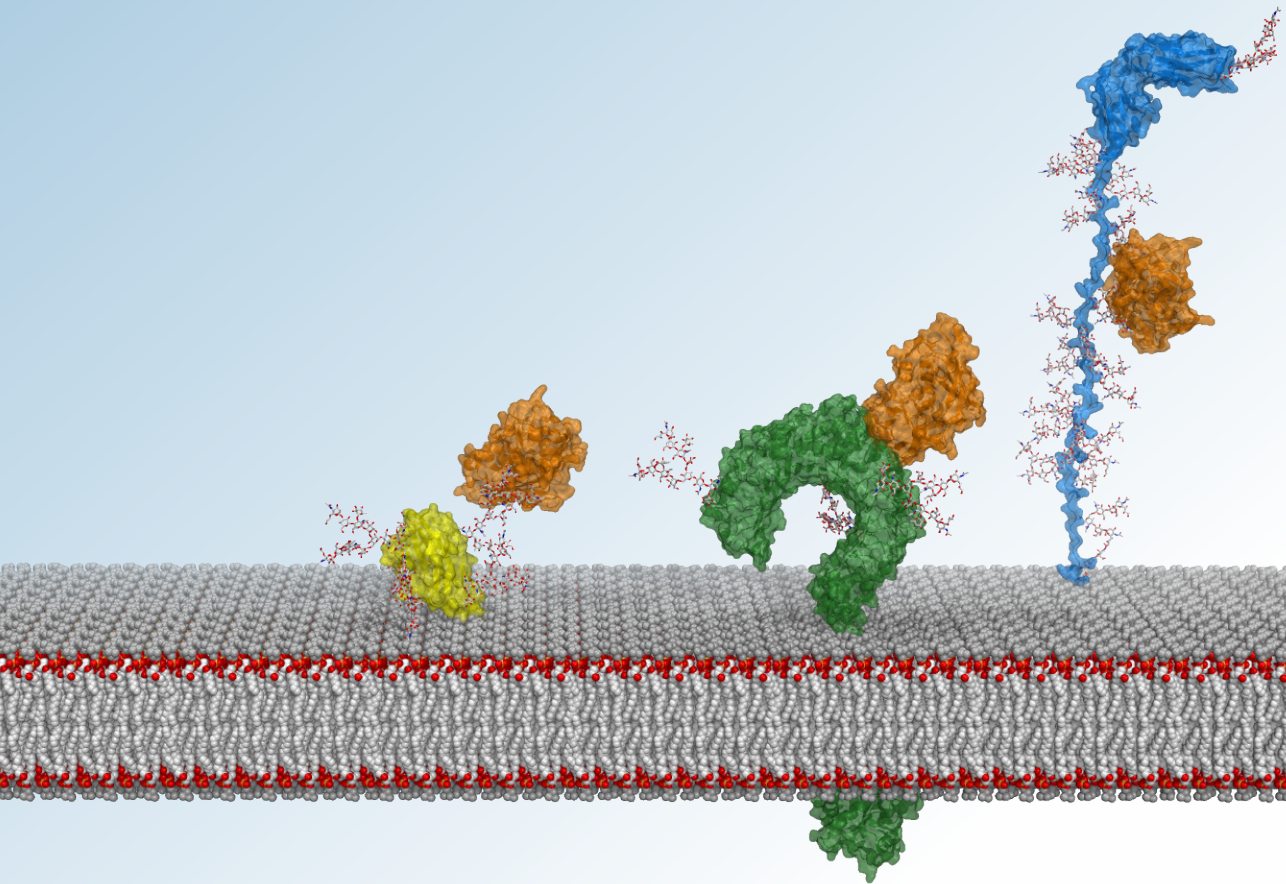


Figure S7: SSL3 binding to a TLR2-Pam₂CSK₄ complex. (A-D) Native MS analysis of the SSL3-TLR2-Pam₂CSK₄-biotin complex: (A) Broad-range native mass spectrum showing the ternary complex (light green, 103.1 ± 0.1 kDa) and a second species (dark green), which likely represents a dimer form of the complex (estimated molecular weight 200-209 kDa). (B-D) Tandem MS analysis of the SSL3-TLR2-Pam₂CSK₄-biotin complex. The region around *m/z* 4900 was mass selected and the sequential disassembly of the complex was monitored by increasing the collision voltage to induce dissociation. Shown are the resulting tandem mass spectra. (B) The protein complex remains intact at 10 V collisional activation. (C) At a collision voltage of 40 V, SSL3 (orange, 23.814 ± 0.001 kDa) is expelled from the complex and Pam₂CSK₄-biotin (cyan) starts to dissociate. Consequently, both free TLR2 (red, 77.27 ± 0.01 kDa) and TLR2-Pam₂CSK₄-biotin (purple, 79.03 ± 0.05 kDa) are visible. (D) Increasing the collision voltage further to 60 V leads to almost complete dissociation of Pam₂CSK₄-biotin, the isotope pattern of which is depicted in the inset. The average molecular weight of the ligand was determined to 1724.7 Da (expected: 1724.4 Da). (E) Binding of Pam₂CSK₄ to TLR2 does not affect the binding of SSL3. Pam₂CSK₄ was allowed to bind mTLR2-Fc before a competitive AlphaScreen binding assay with biotinylated SSL3 and unlabeled SSL3 was performed. In the presence of Pam₂CSK₄ a *K_d* of 0.8 ± 0.2 nM was obtained, compared to the reference of 0.6 ± 0.4 nM in absence of ligand (same reference as shown in Figure S3c). Data points represent the mean ± SD of at least three independent experiments. (F) Top view and cross-section of the SSL3-TLR2 surface near the lipopeptide pocket. Modelling of Pam₂CSK₄ in the SSL3-TLR2 complex in a similar conformation as was observed for PC shows that the peptide moiety of Pam₂CSK₄ can protrude into the solvent region through the observed pocket opening. Pam₂CSK₄ must adopt this conformation characteristic for non-activating ligands (see Figure 4a-b), because its conformation as observed in the Pam₂CSK₄-TLR2-TLR6 complex is not compatible with the formation of a SSL3-TLR2-Pam₂CSK₄ triple complex.

SSL3 NCTC 8325	134	MTPKYEDLRAYYTKPSFEFEKQ GF MLKPWTTVRFMNV IPNRF IYKIALVGKDEKKYK DF YDNID
SSL3 MRSA252	119	INPKFKGLRSYYTKSSLEFKNELGI II KKWTTIRFMN IV PDY F IYKIALVGKDDKKYGE GV HRHVD
SSL4 NCTC 8325	86	INPKFKDLRAYYTKPSLEFKNE IGI ILKKWTTIRFMN V PDY F IYKIALVGKDDKKYGE GV HRHVD
SSL4 MRSA252	91	INPKYKDLRTYYTKPSLEFEKQ GF MLKPWTTVRFMNV IPDWF IYKIALVGKDDKKYK DF YDNID
SSL3 NCTC 8325	200	VFIVLEDNKYQLKKYSVGGITK TNS KKVN H KVELSIT KK DNQGMISRDVSEYMITKEE IS LKELDF
SSL3 MRSA252	185	VFIVLEQNKYVDKYSVGGITKANR KK V DY KTGISIT K EDK GT ISHDVSEYKITKEE IS LKELDF
SSL4 NCTC 8325	152	VFVVLEENNYNLEKYSVGGITKS NS KKV DH KAGVRIT K EDN KG TISHDVSEFKITKE Q ISLKELDF
SSL4 MRSA252	157	VFIVLEDNKYQLKKYSVGGITK TNS KKV DH KAELSIT KK DEK GK ISHDDSEYKITKEE IS LKELDF
SSL3 NCTC 8325	266	KLR Q LIEKH N LYGNMGSGTIVIK M KN G GKY T FEL H KK L Q E HRMADVIDGT N ID N IEV N IK--
SSL3 MRSA252	251	KLR Q LIE Q H N LYGNIGSGTIVIK M KN G GKY T FEL H KK L Q Q HRMADVIDGT N ID R IEV N LK S S
SSL4 NCTC 8325	218	KLR Q LIEKH N LYGN V SGKIVIK M KN G GKY T FEL H KK L Q E NRMADVIDGT N ID N IEV N IK--
SSL4 MRSA252	223	KLR Q LIE Q H N LYGNIGSGTIVIK T KN G GKY T FEL H KK L Q E HRMADVIDGT S IERIEV N LK S S

Figure S8: Alignment of SSL3 and SSL4. Sequence alignment highlighting residues found in the TLR2-binding interface of SSL3 from *S. aureus* strain NCTC 8325 and their conservation in SSL4 from the same strain and SSL3 and SSL4 from strain MRSA252. Nearly all of the SSL3 amino acids involved in TLR2 interactions are present in SSL4 MRSA252, explaining its enhanced TLR2 inhibitory activity with respect to SSL4 from strain NCTC 8325⁹. On the other hand, SSL3 from MRSA252 contains only one of the seven conserved amino acids, indicating that genetic recombination may have occurred in this strain that switched the activity between its SSL3 and SSL4. The length of the proteins, however, corresponds more closely to their original family members, with both SSL3s having the longer N-terminal domain. SSL3 and SSL4 have been assigned based on their gene order in the genomes of NCTC 8325 (SAOUHSC_00386 and SAOUHSC_00389) and MRSA252 (SAR_RS02110 and SAR_RS02115).

4



Analysis of SSL-specific interactions with Fc-tagged receptor ectodomains using multichannel SPR

Louris J. Feitsma^a, Kirsten J. Koymans^b, Piet C. Aerts^b, Carla J.C. de Haas^b, Jos A.G. van Strijp^b, Eric G. Huizinga^a

^a Crystal and Structural Chemistry, Bijvoet Center for Biomolecular Research, Department of Chemistry, Faculty of Science, Utrecht University, Utrecht, The Netherlands

^b Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, The Netherlands

Significance statement

Staphylococcus aureus virulence factors modulate vital parts of our immune system to evade killing of the bacterium. A subset of the Staphylococcal superantigen-like virulence factors (SSLs) bind sialyl Lewis^x (sLe^x) carbohydrates thereby supporting enrichment on the immune cell surface and inhibition of the immune response. Here we use a novel approach providing an experimental basis to discriminate between generic lectin and specific host-pathogen interactions. Binding experiments reveal glycan-enhanced binding of SSL6 to Cluster of Differentiation 47 (CD47) and glycan-independent binding of SSL3 to Toll-like receptor 2 (TLR2). Binding of SSL5 to glycoprotein VI (GPVI) is mediated by carbohydrate interactions only, showing that GPVI is not a specific target of SSL5. These findings can contribute to a better understanding of immune evasion mechanisms.

Keywords: *S. aureus*; SSL; SPR; lectin; innate immunity; immune evasion.

Author contribution: L.J.F. and E.G.H. designed research; L.J.F., K.J.K., P.C.A., and C.J.C.d.H. performed research; L.J.F., K.J.K., C.J.C.d.H., J.A.G.v.S., and E.G.H. analyzed data; J.A.G.v.S. and E.G.H. performed supervision; and L.J.F., K.J.K., J.A.G.v.S., and E.G.H. wrote the manuscript.

Staphylococcus aureus secretes a repertoire of virulence factors that regulate immune reactivity of the host towards the bacterium. An important set of virulence factors comprises the Staphylococcal superantigen-like proteins (SSLs), which interfere in pathogen-recognition, neutrophil extravasation, complement, and phagocytosis. The capacity to associate with sialyl-Lewis^x (sLe^x) containing glycans by a subset of these proteins appears an important determinant for realization of a gradient in the direction of the immune cell and establishing local immune inhibition. Alternatively, or simultaneously, glycan-binding may contribute to specific receptor binding; however, the impact of receptor-specific glycan-binding for most SSLs remains unclear. In this manuscript we present a novel experimental setup to analyze the role of glycan-interactions in receptor binding and use it to study three distinct SSL-target interactions: SSL3-TLR2, SSL5-GPVI, and SSL6-CD47. SPR binding experiments using glycosylated (*g*) and mannosylated (*m*) receptor forms revealed that binding of SSL3 to Toll-like receptor 2 (TLR2) does not involve TLR2-glycans. In contrast, high-affinity CD47-binding by SSL6 is mediated by both protein-protein and lectin-glycan interactions; substantially weaker binding of CD47 by SSL5 is mediated by lectin-glycan interactions only. Binding of SSL5 and SSL6 to glycoprotein VI (GPVI) relies exclusively on interactions with *N*- or *O*-linked GPVI-glycans, demonstrating that previous claims of specific SSL5-GPVI interaction solely supported by limited binding data have been premature. Combined, our data show glycan-enhanced SSL6-CD47 binding and glycan-independent SSL3-TLR2 binding; the latter implying a different role of sLe^x-binding for SSL3.

Introduction

Staphylococcus aureus is a Gram-positive bacterium, primarily found as a commensal skin bacterium in ~30% of the human population. Nonetheless, it is a successful human pathogen upon host entrance and responsible for a great diversity of diseases including skin and respiratory infections, food poisoning, sepsis and pneumonia¹. Emergence of this bacterium as a pathogen and lack of sufficient clinical remedy is caused by 1) increased antibiotic resistance, such as various methicillin-resistant *S. aureus* strains (MRSA)^{2,3} and 2) a repertoire of secreted toxins that interfere with vital host processes involved in recognition and clearance of staphylococci⁴.

Among these toxins are the staphylococcal superantigen-like proteins (SSLs), which target various components of the host-immune system, including complement (SSL7 and SSL10)⁵⁻⁸, chemokine receptors (SSL5 and SSL10)^{9,10}, P-selectin glycoprotein ligand-1 (SSL5 and SSL11)^{11,12}, Toll-like receptor 2 (SSL3 and SSL4)¹³, 'marker-of-self' receptor CD47 (SSL6)¹⁴, Immunoglobulins (SSL7 and -10)^{5,15-17}, matrix metalloproteinases (SSL1 and SSL5)^{18,19} and FcαRI (SSL11)¹². In addition, an activating role in hemostasis has been described for SSL5, which can induce platelet aggregation via platelet receptor glycoprotein GPIbα, integrin α_{IIb}β₃²⁰, and putatively GPVI²¹. This plurality of SSL-targets in immunity and hemostasis emphasizes the diverse mechanisms that *S. aureus* uses to evade immune activation and to emerge as a pathogen.

A shared aspect of the SSL-subset consisting of SSLs 2-6 and -11 is the capacity to bind sialyl-Lewis^x (sLe^x): a GlcNAc(-Fuc)-Gal-Sia⁵ tetrasaccharide frequently present in complex

§ Abbreviations used: *GlcNAc*, N-acetylglucosamine; *Gal*, Galactose; *Fuc*, Fucose; *Sia*, Sialic acid; *GalNAc*, N-acetylgalactosamine.

N- and *O*-linked glycans at the immune cell surface (Figure 1a-b)²². Crystal structures of SSL4, SSL5, and SSL11 in complex with sLe^x have been solved and show a conserved TxExxKxxQx[D/H/N]RxxD-motif in the β -grasp domain involved in sLe^x-binding^{12,23,24}. Scarcely available binding data indicate substantially varying affinities for sLe^x-binding, ranging from 87 nM (SSL4, strain Newman)²⁴ to 2.3 μ M (SSL11, strain US6610)¹². In the case of SSL11, affinities for sLe^x and derivatives like sLacNac (GlcNac-Gal-Sia) even appear to deviate between different *S. aureus* subspecies. These differences possibly arise from

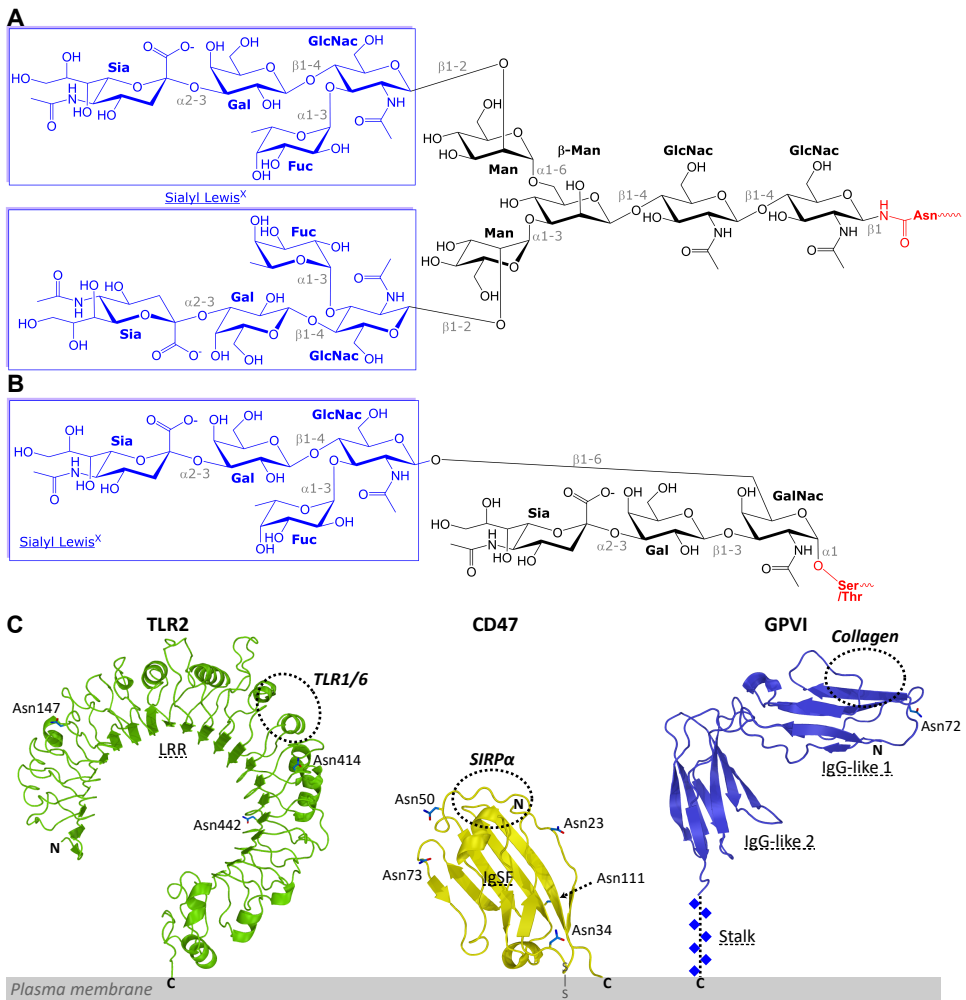


Figure 1: Structures of *N*- and *O*-linked glycans and receptor ectodomains used in this study. (A) Structure of a complex *N*-linked glycan antenna, coupled to asparagine-residues. The tetrasaccharide sialyl-Lewis^x (sLe^x; GlcNac(-Fuc)-Gal-Sia; outlined in blue) can exist at the end of two branches, coupled via a β 1-2 glycosylic linkage between the GlcNac-moiety and a mannose. (B) Structure of the type 2 *O*-linked glycan antenna, coupled to serine or threonine-residues. The sLe^x tetrasaccharide is coupled via a β 1-6 glycosylic linkage of the GlcNac-moiety to the first GalNac-moiety. (C) Ectodomain structures (not to scale) of three SSL-targets: TLR2 (left; target of SSL3), CD47 (middle; target of SSL6), and GPVI (right; target of SSL5). Potential *N*-glycosylation sites are indicated by labeled Asn residues; blue squares indicate *O*-glycans in GPVI. Regions involved in ligand binding are indicated by ellipses.

a distinct GlcNAc and Fuc hydrogen bonding pattern depending on the presence of Asp, His, or Asn in the binding motif¹². The availability and accessibility of sLe^x on SSL-specific receptors is in most cases not known; their presence depends on the expression of sufficient glycosyltransferases in the specific cell type, in particular galactosyl- and sialyl-transferases in the trans-Golgi²⁵.

The contribution of sLe^x-binding to the host-specific SSL-interaction is often not fully understood, but can have marked impact on target-binding. Sialomucin P-selectin glycoprotein ligand-1 (PSGL-1) is expressed by most leukocytes and contains numerous sLe^x-containing *O*-glycans that are essential for P-selectin binding²⁶. SSL5 competes with P-selectin for sialylated glycans on PSGL-1, and is a potent inhibitor of P-selectin-mediated neutrophil rolling¹¹. SSL11 has equal affinity for the Fc receptor for IgA (FcαRI) and a BSA-sLe^x conjugate. It has been suggested that glycan-binding may be used for rapid internalization of this virulence factor into the host cell *in vivo*¹², meanwhile blocking the IgA-FcαRI interaction. The complex between SSL3 and Toll-like receptor 2 (TLR2) as observed in the crystal structure (PDB-ID: 5D3I)²⁷ is solely constituted by protein-protein specific interactions, without involvement of TLR2-glycans and the sLe^x-binding site of SSL3. Whether or not SSL3 can separately associate with TLR2-glycans is not known. Overall, the presence of sLe^x-glycans on the immune cell surface is an important determinant for SSL-targeting in the direction of the immune cell, but the contribution of SSL-sLe^x binding to specific immune modulating processes appears to vary considerably.

Here, we present a novel experimental approach of receptor ectodomain expression and multi-channel SPR binding experiments to analyze contributions of target-specific glycan binding by lectin-SSLs to previously described host-microbe interactions. We examined three previously observed interactions, SSL3-TLR2, SSL6-CD47, and SSL5-GPVI (Figure 1c), by expression and purification of glycosylated (prefix *g*) and mannosylated (prefix *m*) Fc-tagged receptor variants, and their immobilization by Fc-ProteinA affinity. For each receptor we analyzed binding to the receptor-specific SSL previously described in literature, and one non-specific SSL for which binding is expected to be purely mediated by sLe^x-glycans.

Results

Expression of putative SSL target-receptors as chimeric Fc-dimers and immobilization by ProteinA-affinity

We expressed receptor ectodomains of TLR2, CD47 and GPVI fused to the dimeric IgG1 Fc-region in two glycoforms: glycosylated (*g*) and mannosylated (*m*). Proteins in the latter form were expressed in HEK-cells deficient in GNTI. Expression was tested using a full-length C-terminal TEV-Fc-tag – a TEV-cleavable linker followed by the IgG1 hinge-CH2-CH3 domains – and a short Fc-tag in which a four-residue AAAG-linker replaces the TEV-cleavable linker and hinge-region. *A priori*, this shorter tag forces receptor domains to be considerably closer together, for which only sufficient expression is expected if this relative conformation is sterically feasible. Expressions tests using the short tag resulted in equal or higher yields compared to TEV-Fc variants (see Figures S1-4). Analysis of the expressed proteins by SDS-PAGE indicated high degrees of purity (>95%) after purification

with ProteinA. Analytical gel filtration furthermore showed about equal retention times for both Fc-forms, indicating that the short Fc-tagged receptors form non-covalently linked dimers *in vitro*. In the cases tested here, the four-residue linker is apparently of sufficient length.

Previous attempts to immobilize TLR2 via amine-carboxylate coupling and biotin-streptavidin coupling resulted in substantial on-chip aggregation as was verified by the resonance-per-pixel option of the IBIS-MX96 (Figure S5A-B). The use of low pH coupling conditions, or the intermediate *in vitro* biotinylation step in these experiments may have damaged the receptor. Expression of Fc-tagged receptors enabled generic coupling by Fc-proteinA/G affinity for SPR binding experiments. Initial tests indicated that immobilization required only 75-300 ng of purified protein at concentrations between 0.5 and 2.0 $\mu\text{g}/\text{mL}$ per region of interest (ROI). This compares favorably to the amount of protein required for amine-carboxylate coupling (15-30 μg per ROI) and biotin-streptavidin coupling (1.5-4 μg per ROI) of TLR2, in order to obtain homogeneous spot-densities with appropriate receptor capture (Figure S5C). Also CD47 and GPVI were effectively coupled using this method (Figure S5D-E), although the dynamic range of receptor immobilization is limited in all cases. Additionally, baseline drift after receptor coupling indicated slow dissociation of weakly bound protein and demanded that binding experiments were preceded by repetitive blank cycles to obtain a stable baseline. Although the immobilization procedure could be further optimized, including a feasibility test of on-chip purification of crude expression medium, the results already show that the method in its current form provides a sound basis to study host-pathogen interactions.

Binding of mTLR2 by SSL3 does not involve interaction with TLR2-glycans

SSL3 and -4 were shown to inhibit Toll-like receptor 2 (TLR2)¹³; the receptor involved in sensing of bacterial lipopeptides^{28,29}. Inhibition of TLR2 was demonstrated to be specific for SSL3 and -4²⁷, and is not mediated by other SSLs, including lectin SSLs¹³. The SSL3-TLR2 complex suggest that SSL3, while inhibiting TLR2, cannot simultaneously bind *N*-linked glycans of TLR2, since they are located too far away from the sLe^x-binding site of SSL3. The exact role of the sLe^x-binding site in SSL3 for TLR2-binding remains therefore elusive.

We analyzed SSL3-binding to immobilized TLR2 in an SPR titration experiment, by consecutive injections of SSL3 with concentrations starting from 1.5 nM and followed by a stepwise two-fold increase to a final concentration of 3 μM . SSL3 bound all TLR2-constructs, independent of the TLR2 glycoform, with an apparent K_D of approximately 0.5 μM (Figure 2a-c and Table 1). Binding is therefore substantially weaker than previously obtained from inhibition experiments using the alpha-assay²⁷. Also the total ratio of SSL3-TLR2 binding in our titration experiments is different from equimolar: SSL3 binds TLR2 in a molar ratio between 3:1 and 4:1 (Figure 2e). The observation that multiple SSL3-molecules bind TLR2, however, is independent of the TLR2 glycoform, and therefore probably not related to SSL3-glycan binding. Possibly, the predominantly hydrophobic nature of the SSL3 and TLR2 molecules may cause increasing formation of on-chip aggregates upon increase of the SSL3-concentration, which could also explain the much weaker affinity compared to previous experiments.

We also measured binding of SSL6 to TLR2 (Figure 2b and d). SSL6 has not been

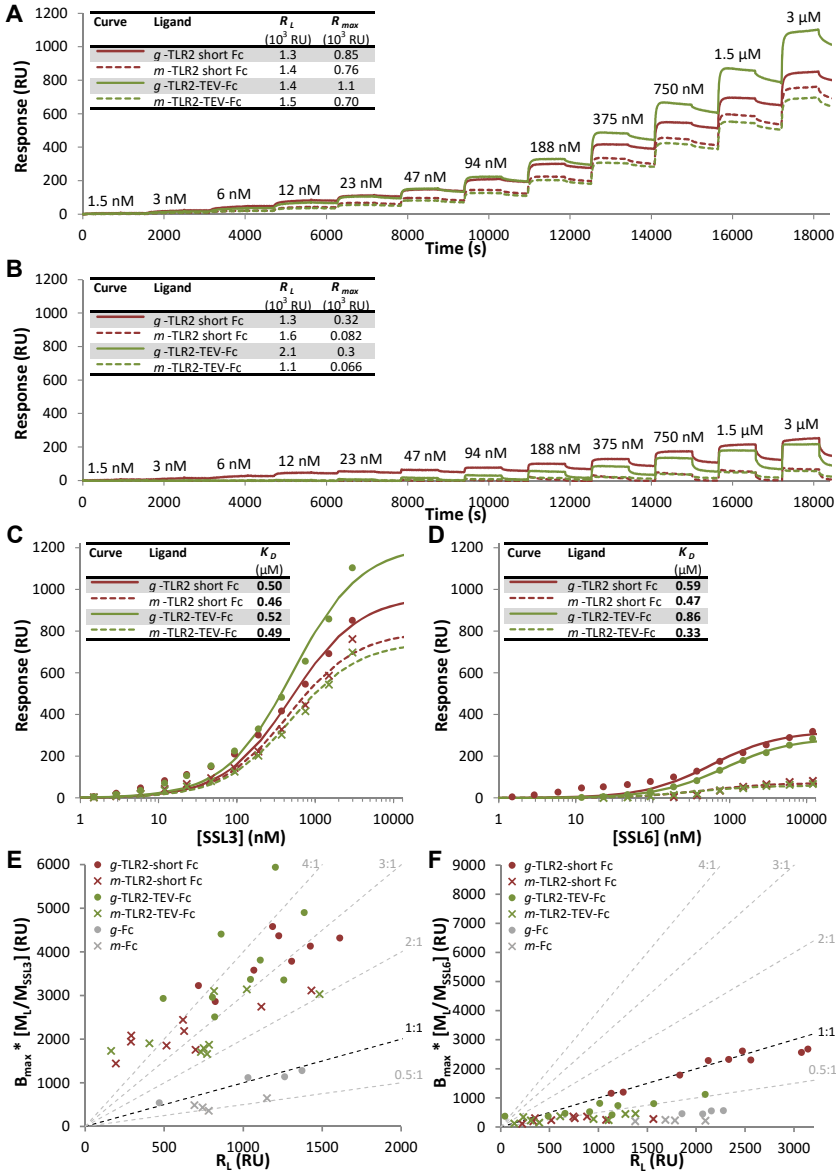


Figure 2: SSL3- and SSL6 binding to TLR2-Fc. (A-B) Kinetic titration experiment using successive injections of stepwise increased concentrations (3 nM - 6 μ M) of SSL3 (A) or SSL6 (B) to TLR2-short Fc (red lines) and TLR2-TEV-Fc (green lines). Responses (response units, RU) of SSL-binding were measured to both glycosylated (solid lines) and mannosylated (dashed lines) forms of TLR2. SSL-binding to each TLR2-variant was measured in nine different regions. A binding curve for one representative region is shown. Titration cycles consisted of a 900-sec. association phase directly followed by a 750-sec. dissociation phase. The inset shows the initial ligand capture responses (R_L , RU) and maximal SSL-binding (R_{max}) reached for the depicted curves. (C-D) Non-linear regression fits of a 1:1 binding model to the end-of-injection points measured for SSL3- (c) and SSL6-binding (d) to TLR2. Confidence intervals of K_D values are presented in Table 1. (E-F) Relation between the fitted maximum binding (B_{max}) of SSL3- (E) and SSL6- (F) binding (B_{max} , RU) and the initial ligand capture responses (R_L) of immobilized TLR2-short Fc (red), TLR2-TEV-Fc (green), or Fc (gray). For an equimolar comparison of the initial ligand and analyte responses, B_{max} -values have been corrected by the molecular mass-ratio $[M_L/M_{SSL}]$ between the corresponding ligands and the SSL-analyte.

postulated to target TLR2 *in vivo*, and, as expected, substantially lower responses were observed. SSL6 bound TLR2 in a glycan-dependent manner with a molar ratio lower than 1:1, showing that SSL6 interacts with, at best, one TLR2-glycan. These results are in line with the results of SSL3-TLR2 binding, indicating that interactions of SSL3 with TLR2-glycans do not noticeably contribute to complex formation. The minimal ability of both SSL3 and SSL6 to interact with TLR2-glycans suggest that these glycans may contain no or very few functional groups such as sLe^x or sLacNac, to which SSL3 and SSL6 could bind. Combined, formation of the SSL3-TLR2 complex is mediated by protein-protein interactions only, and the lectin capacities of SSL3 must therefore target other, non-TLR2 glycans, possibly for pre-concentration on the immune cell surface.

Table 1: Kinetic parameters of SSL3- and SSL6-binding to TLR2 homodimers

Protein	SSL3-TLR2			SSL6-TLR2		
	K_D (μ M)	95% interval (μ M)	SSL3:TLR2	K_D (μ M)	95% interval (μ M)	SSL6:TLR2
<i>g</i> -TLR2-short Fc	0.50	0.44 - 0.56	3.4 \pm 0.6	0.59	0.50 - 0.68	1.0 \pm 0.1
<i>g</i> -TLR2-TEV-Fc	0.52	0.47 - 0.57	3.9 \pm 1.1	0.86	0.80 - 0.92	0.6 \pm 0.1
<i>g</i> -Fc	0.32	0.28 - 0.36	1.0 \pm 0.1	0.50	0.34 - 0.66	0.2 \pm 0.1
<i>m</i> -TLR2-short Fc	0.46	0.41 - 0.51	4.4 \pm 2.1	0.47	0.23 - 0.70	0.4 \pm 0.2
<i>m</i> -TLR2-TEV-Fc	0.49	0.43 - 0.54	2.9 \pm 0.9	0.33	0.19 - 0.47	0.6 \pm 0.4
<i>m</i> -Fc	0.31	0.28 - 0.34	0.6 \pm 0.1	0.43	0.38 - 0.48	0.1 \pm 0.1

High-affinity SSL6-CD47 binding depends on the CD47 glycoform

Antibody-competition studies recently identified SSL6 as a target of cell-surface receptor CD47¹⁴: a widely distributed receptor regulating several cellular functions including phagocytosis, T-cell activation, regulated cell death, and the marker of the body's own red blood cells^{30,31}. Moreover, CD47 is upregulated on tumor cells^{32,33} as well as on haematopoietic stem and progenitor cells upon inflammatory stimuli³⁴, making it an interesting target for the development of anti-cancer and anti-inflammatory therapies. It has been shown that low concentrations of SSL6 induce phagocytosis by neutrophils, which suggests interference of SSL6 in the variety of cellular processes that are regulated by CD47. The 14-kDa extracellular IgSF-domain of CD47 is involved in SIRP α -binding and has 5 predicted *N*-linked glycosylation sites³⁵, which may be involved in SSL6-binding, since neuraminidase-treatment of CD47 in previous experiments abolished the interaction with SSL6¹⁴.

We measured SSL6-binding to both glycosylated (*g*) and mannosylated (*m*) forms of CD47-Fc in kinetic titration experiments. SSL6 binds both forms, but with a large difference in affinity (Figure 3A and Table 2). Binding of SSL6 to *g*-CD47 is biphasic: a saturable high-affinity binding step with an apparent $K_{D,obs}$ of 14 nM is well-separated from a second low affinity step with a K_D of about 2.4 μ M (Figure 3B). On the other hand, SSL6 binds *m*-CD47 in a single phase with a K_D of 0.31 μ M (Figure 3C), which is about 22-fold weaker than the 14 nM affinity for *g*-CD47. The Hill-plots corresponding to the nanomolar binding steps to *g*-CD47 and *m*-CD47 show Hill-coefficients (n_H) of 2.9 and 1.1, respectively. Apparently, multiple SSL6-molecules bind cooperatively to glycosylated CD47-dimers in a glycan-dependent mode, but not to CD47-dimers that lack sLe^x-glycans. The considerable

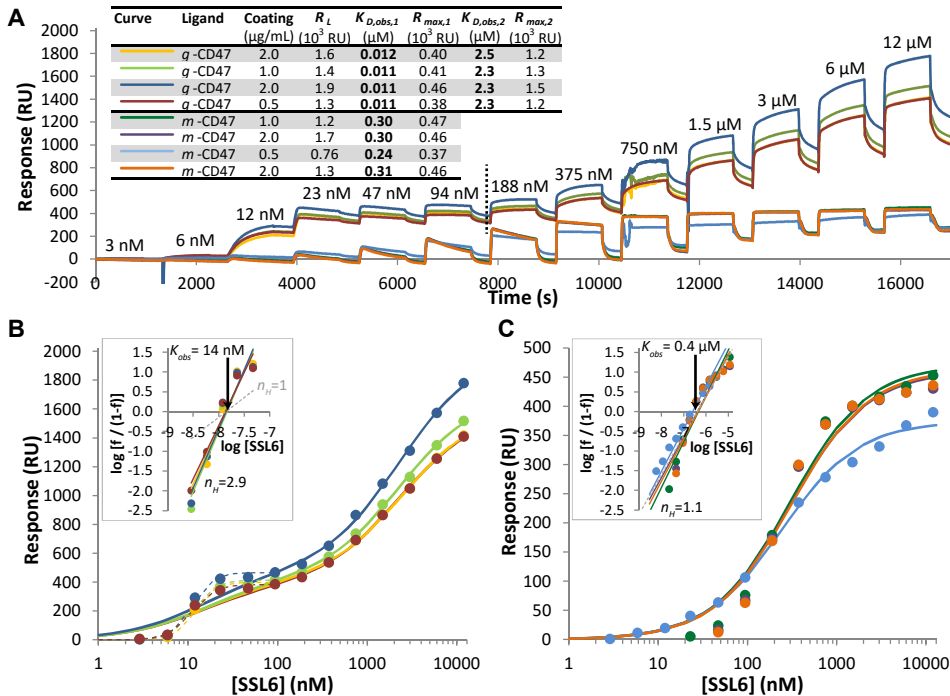


Figure 3: SSL6-binding to wild type *g*-CD47 and *m*-CD47 (A) Kinetic titration experiment using successive injections of stepwise increased SSL6-concentrations (3 nM - 12 μM) to immobilized wild type CD47-TEV-Fc. Responses (response units, RU) of SSL6-binding were measured to four independent *g*-CD47 (light orange, green, blue, and Bordeaux-red curves) and *m*-CD47 regions (dark green, purple, light blue, and dark orange curves). Titration cycles consisted of a 900-sec. association phase directly followed by a 600-sec. dissociation phase per injection. The inset shows the initial ligand capture responses (R_L , RU) and fitted kinetic parameters $K_{D,obs,1}$, $R_{max,1}$, $K_{D,obs,2}$ and $R_{max,2}$ (*g*-CD47) or $K_{D,obs}$ and R_{max} (*m*-CD47) for SSL6-binding per CD47-region. (B-C) Non-linear regression of the end-of-injection responses measured for biphasic SSL6-*g*-CD47 (B) and monophasic SSL6-*m*-CD47 (C) binding (solid lines). The insets show the Hill-plots corresponding to the high-affinity binding phases, containing the end-of-injection responses measured in the 3-94 nM range (*g*-CD47) or the full concentration range (3 nM-12 μM; *m*-CD47). Linear fitting indicates Hill-coefficients (n_H) of 2.9 and 1.1 for SSL6-*g*-CD47 and SSL6-*m*-CD47 binding, respectively. Data for SSL6-*g*-CD47 binding including cooperativity are added to the plots as dashed lines.

differences in affinities for *g*-CD47 and *m*-CD47 confirm that SSL6-binding is strongly dependent on the CD47-glycans, yet, not restricted to the glycosylated form only, which suggests that protein-protein specific components and SSL-glycan interactions both contribute to SSL6-CD47 binding.

The total capacity of both receptor forms to bind SSL6 deviates about 4-fold (Figure 3A). Interestingly, the maximum responses observed for the 14 nM-binding to *g*-CD47 ($R_{max,1}$) and overall maximum binding to *m*-CD47 are in the same order of magnitude, about 450 RU. Given the very similar receptor densities in both cases as estimated by the initial ligand responses, this is in accordance with approximately equal SSL6-binding to both receptor forms: most likely two SSL6-molecules per CD47-dimer, i.e. one per CD47-molecule. Only *g*-CD47 and not *m*-CD47 associates with additional SSL6-molecules in a second binding phase, up to a maximum response that is 2.6 ± 0.6 times higher than observed for the high-affinity phase. It is therefore expected that this additional binding

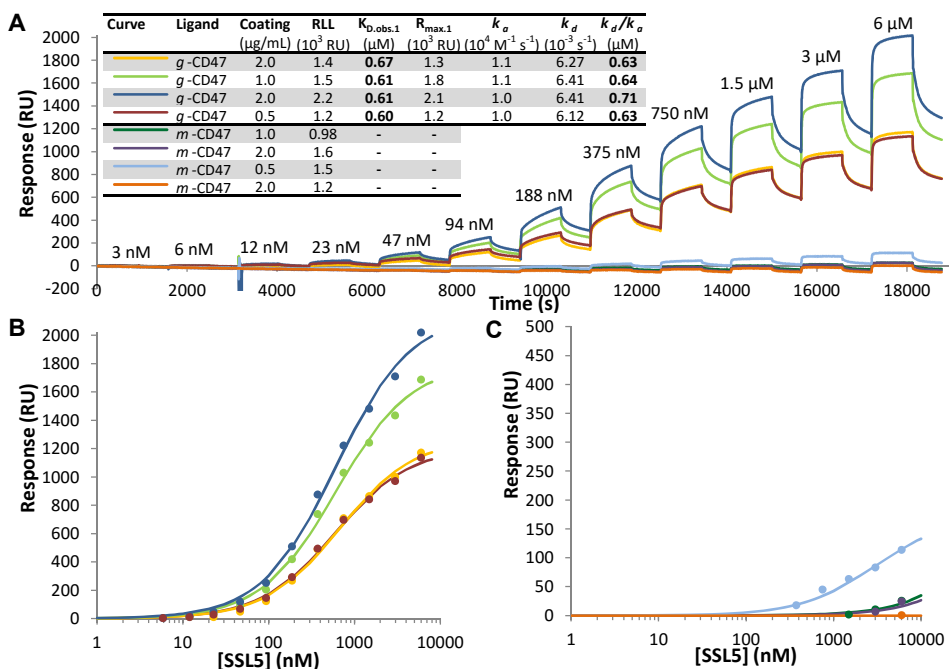


Figure 4: SSL5-binding to wild type *g*-CD47 and *m*-CD47. (A) Kinetic titration experiment using successive injections of stepwise increasing SSL5-concentrations (3 nM - 6 μM) to immobilized wild type CD47-Fc. Responses (response units, RU) of SSL5-binding were measured to four independent *g*-CD47 (light orange, green, blue, and Bordeaux-red curves) and *m*-CD47 regions (dark green, purple, light blue, and dark orange curves). Titration cycles consisted of a 900-sec. association phase directly followed by a 750-sec. dissociation phase per injection. The inset shows the initial ligand capture responses (R_L , RU) and fitted kinetic parameters K_D and R_{max} for SSL5-binding per CD47-region. (B-C) Non-linear regression of the end-of-injection points measured for *g*-CD47 (B) and *m*-CD47 (C).

purely relies on SSL6-binding to 2 or 3 out of the 4 remaining CD47-glycans that are not involved in formation of the high-affinity SSL6-CD47 complex.

We also measured binding of SSL5 to both CD47 glycoforms. Like SSL6, SSL5 exhibits the ability to bind glycans through its sLe^x-binding site, but unlike SSL6 lacks the capacity for SSL-protein interactions with CD47; a trend that is also observed in our binding experiments (Figure 4A). No detectable SSL5-binding was obtained for three out of four *m*-CD47-regions within the tested concentration-range here (3 nM-6 μM; Figure 4C). For one spot, small responses were detected that correspond to less than one SSL5-molecule per CD47-dimer at the highest SSL5-concentrations. Consistent with its capacity for SSL-glycan interactions, SSL5-binding to *g*-CD47 is monophasic (Figure 4B) and comparable with the low affinity SSL6-*g*-CD47 binding step. SSL5 binds *g*-CD47 with a K_D of approximately 0.94 μM, which is interestingly close to the K_D previously measured for the specific SSL5-PSGL-1 interaction (0.82 μM)¹¹. Absence of SSL5-CD47 specific interactions dictate that the observed *g*-CD47-binding by SSL5 must be explained purely as interaction between the sLe^x-binding site of SSL5 and CD47-glycans. Combined, SSL6 and not SSL5 specifically targets the CD47-receptor, and the SSL6-CD47 interaction is mediated by both protein-protein and SSL-glycan interactions.

High-affinity SSL6-CD47 binding is independent of glycosylation of Asn23, Asn73, and Asn111

To analyze contribution of individual CD47-glycans to SSL6-binding, we also evaluated binding to *N*-glycosylation mutants of CD47. We expressed single, double, triple, and quadruple mutants, knocking-out the N23, N34, N50, N73, and N111 glycosylation sites, and combinations between them. Expression yields of these mutants and protein purity as assessed by SDS-PAGE were generally lower than that of wild type protein (Figure S3), emphasizing the role of the glycans for a correct fold of the CD47-domain itself. Sufficient amounts of folded protein could be expressed using the short-Fc fusion.

Both glycosylated and mannosylated forms of four single mutants, N23Q, N50Q, N73Q, and N111Q, bind SSL6 equally well as wild type protein (Figure S6A and Table 2). Glycosylated mutants only vary in their capacity to bind additional SSL6-molecules in the second, glycan-mediated binding phase. Based on the responses, SSL6 could bind between 1 and 4 glycans per receptor molecule in addition to the high-affinity binding step. Apparently, elimination of these glycans individually does not affect the SSL6-CD47 specific interactions. Only the T36N mutation, which eliminates the glycosylation motif of N34, fully impaired both the 14 nM and 0.3 μM binding steps observed for the wild-type *g*- and *m*-forms respectively. The remaining binding to *g*-T36N has comparable affinity and capacity for SSL6 (1.2 μM) and SSL5 (0.62 μM; Figure S7A and Table 2). In view of the location of T36 proximate to the cell membrane *in vivo*, however, its involvement in SSL6-binding is unlikely. In line with the previously solved crystal structure of CD47³⁵, indicating that first GlcNac-moiety of N34 is moiety buried between β-strands A and B, the drastic effect of this mutation suggests partial unfolding of the N-terminal region of CD47 and disruption of the site involved for SSL6-binding.

The N23Q-N50Q-N73Q-N111Q quadruple mutant, which lacks all glycans except for the structurally important N34, binds SSL6 with a K_D of 0.35 μM in both glycoforms and

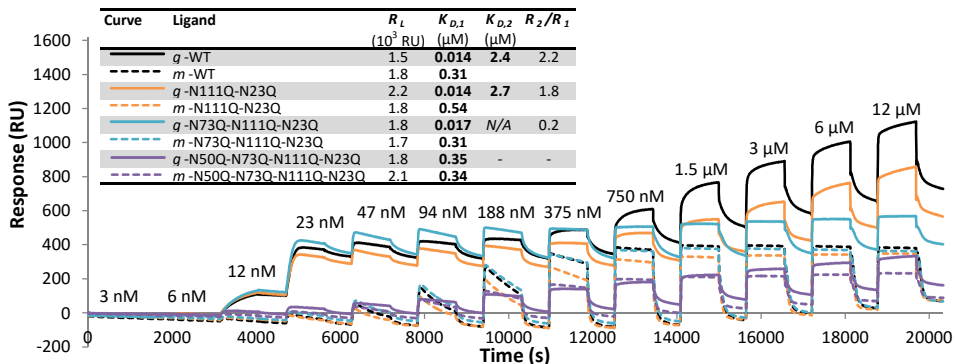


Figure 5: SSL6-binding to wild type CD47 and a representative double, triple, and quadruple glycan mutants. SSL6-binding to a selection of CD47 glycan mutants that exhibit highly similar patterns for the high-affinity binding phase as wild type *g*-CD47 (black), N111Q-N23Q (orange) and N73Q-N111Q-N23Q (light blue), or as *m*-CD47 (black), N50Q-N73Q-N111Q-N23Q (purple). Responses (response units, RU) of SSL6-binding were measured for regions containing *g*-CD47 (solid lines) or *m*-CD47 (dashed lines). SSL6-binding was measured in at least three independent ROIs per mutant, of which one representative curve per mutant is shown here. The inset shows the initial ligand capture responses (R_i), the kinetic parameters $K_{D,1}$ and $K_{D,2}$, and the ratio R_2/R_1 for the relative binding capacity of the second, low-affinity step with respect to the first phase.

with about equal capacity (Figure 5 and Table 2). The observed glycoforms-independence of this interaction suggests that, although the N34 glycan is required for proper CD47 folding, it does not contribute to the high affinity binding step. The other quadruple mutants show some ability to bind SSL6 in mannosylated form, albeit with much lower affinity (Figure S6D), but none of these mutants exhibits the high-affinity phase for *g*-CD47. To reach the 14nM SSL6-*g*-CD47 affinity, at least one other glycan is necessary in addition to the N34-glycan.

The highest order mutant in our series that binds SSL6 similar to wild type protein in both glycoforms is triple mutant N23Q-N73Q-N111Q (Figure 5). This would suggest that both the N34-glycan (structural integrity) and the N50-glycan (involved in lectin-binding) are required for high affinity SSL6-*g*-CD47 binding. However, the N50Q-mutation itself did not show an effect on SSL6-binding (Figure S6A), indicating that complex formation is not limited to the presence of the N50 glycan. This more complex situation could be explained if multiple sLe^x-moieties can reach the sLe^x-binding site in a SSL6-CD47 complex and the presence of at least one of them would lead to high-affinity SSL6-binding. SSL6 binds only marginally to remaining glycans of the N23Q-N73Q-N111Q mutant upon saturation of the 14 nM binding phase, suggesting that no other glycan is present or accessible, making this mutant an interesting target for studies to reveal the structure of the SSL6-CD47 complex.

Table 2: Kinetic parameters of SSL6- and SSL5-binding to wild type CD47 and glycan point mutants

Mutant	<i>g</i> -CD47-SSL6					<i>m</i> -CD47-SSL6		<i>g</i> -CD47-SSL5	
	$K_{d,1}$ (μ M)	95% int. (μ M)	$K_{d,2}$ (μ M)	95% int. (μ M)	$R_{max,2}/R_{max,1}$	K_D (μ M)	95% int. (μ M)	K_D (μ M)	95% int. (μ M)
WT	0.014	0.011 - 0.017	2.4	2.1 - 2.8	2.6 ± 0.6	0.31	0.27 - 0.36	0.94	0.84 - 1.0
N23Q	0.011	0.007 - 0.015	3.3	2.0 - 4.6	2.2 ± 0.2	0.28	0.21 - 0.34	0.57	0.51 - 0.63
T36N	-	-	1.2	1.0 - 1.3	-	-	-	0.62	0.56 - 0.67
N50Q	0.015	0.010 - 0.020	2.1	1.0 - 3.1	1.3 ± 0.1	0.22	0.17 - 0.26	0.74	0.65 - 0.83
N73Q	0.015	0.010 - 0.019	1.9	1.5 - 2.2	3.8 ± 0.1	0.25	0.20 - 0.30	0.52	0.47 - 0.57
N111Q	0.014	0.009 - 0.019	1.8	1.5 - 2.1	4.3 ± 0.3	0.30	0.24 - 0.36	0.64	0.57 - 0.70
N23Q-T36N	-	-	1.4	1.2 - 1.5	-	0.88	0.55 - 1.2	1.2	1.0 - 1.3
T36N-N50Q	-	-	1.4	1.3 - 1.5	-	0.62	0.51 - 0.72	1.4	1.3 - 1.5
N50Q-N73Q	0.018	0.011 - 0.024	1.6	1.4 - 1.8	7.9 ± 1.2	0.35	0.21 - 0.50	1.4	1.2 - 1.5
N73Q-N111Q	-	-	1.1	1.0 - 1.2	-	0.90	0.40 - 1.4	1.0	0.91 - 1.1
N111Q-N23Q	0.014	0.007 - 0.020	2.7	1.2 - 4.0	1.8 ± 0.1	0.36	0.26 - 0.81	1.3	1.1 - 1.5
N23Q-T36N-N50Q	0.025	0.015 - 0.035	0.8	0.4 - 1.1	1.6 ± 0.2	0.59	0.22 - 0.96	2.6	2.1 - 3.1
T36N-N50Q-N73Q	-	-	1.1	0.9 - 1.2	-	0.44	0.29 - 0.59	1.1	1.0 - 1.3
N50Q-N73Q-N111Q	-	-	1.0	0.9 - 1.1	-	0.41	0.19 - 0.62	1.4	1.3 - 1.5
N73Q-N111Q-N23Q	0.017	0.007 - 0.027	N/A*	-	0.22 ± 0.02	0.31	0.16 - 0.45	1.5	0.5 - 2.5
N111Q-N23Q-T36N	0.024	0.015 - 0.033	1.1	0.9 - 1.3	4.6 ± 0.9	0.92	0.32 - 1.5	1.1	1.0 - 1.2
N23Q-T36N-N50Q-N73Q	-	-	1.5	1.3 - 1.7	-	0.99	0.83 - 1.1	1.9	1.5 - 2.3
T36N-N50Q-N73Q-N111Q	-	-	1.6	1.4 - 1.7	-	0.92	0.77 - 1.1	1.8	1.4 - 2.3
N50Q-N73Q-N111Q-N23Q	-	-	0.35	0.29 - 0.40	-	0.34	0.25 - 0.44	4.0	1.7 - 6.3
N73Q-N111Q-N23Q-T36N	-	-	1.5	1.3 - 1.6	-	0.91	0.81 - 1.0	1.9	1.4 - 2.4
N111Q-N23Q-T36N-N50Q	-	-	1.2	0.98 - 1.4	-	0.95	0.58 - 1.3	1.8	1.3 - 2.3

* The K_D of the second SSL6-binding phase could not be accurately determined because of the low R_{max} .

SSL5 and SSL6 both bind GPVI via interactions with *N*- and *O*-glycans

Besides various immune regulatory functions, it has also been demonstrated that SSL5 can modulate hemostasis by induction of non-native platelet activation. In plasma-free experiments, SSL5 induced platelet aggregation by interacting with platelet receptors GPIIb α (0.34 μ M) and integrin $\alpha_{IIb}\beta_3$ (0.82 μ M)²⁰. More recently, SSL5 has also been suggested to stimulate platelet activation via another platelet receptor: GPVI²¹. However, this claim is only substantiated by a single binding experiment using recombinant glycosylated receptor ectodomains, and does not address the lectin-capacities of SSL5.

We produced glycosylated and mannosylated forms of three GPVI-ectodomain-Fc fusions, all containing the D1-D2 domains and different parts of the *O*-glycosylated stalk region: *GPVI-D1D2* (without stalk), *GPVI-234* (includes 30 residues of the stalk) and *GPVI-261* (includes the complete 57-residue stalk). SDS-PAGE analysis of expressed proteins (Figure S4) revealed higher molecular weights than expected from protein mass (Table 3), likely arising from the presence of two *N*-glycans, located in the D1- and Fc-domains, and a number of *O*-linked glycans in the stalk-region. The number of Ser/Thr-residues in the stalk-regions of our GPVI-234 and GPVI-261 constructs, 10 and 22, respectively, would limit the maximum number of *O*-glycans in these constructs accordingly. The molecular weight differences observed for these constructs indicate that these residues are indeed glycosylated, independent of the HEK-cells used for expression.

Table 3: Molecular weight analysis of GPVI ectodomain expressions

Protein	<i>g</i> -GPVI					<i>m</i> -GPVI					Potential M_w (kDa)
	Theoretical M_w (kDa)		M_w on gel	ΔM_w	ΔM_w	Theoretical M_w (kDa)		M_w on gel	ΔM_w		
	Protein	<i>N</i> -Glycans	Total	(kDa)		(kDa)	Protein	<i>N</i> -Glycans	Total	(kDa)	
GPVI-D1D2-TEV-Fc	48.0	5.2 (2)	53.2	56.8	3.6	48.0	2.8 (2)	50.8	55.9	5.0	– (0)
GPVI-234-TEV-Fc	51.0	5.2 (2)	56.1	68.4	12.3	51.0	2.8 (2)	53.8	67.6	13.8	6.6 (10)
GPVI-261-TEV-Fc	53.8	5.2 (2)	59.0	76.4	17.4	53.8	2.8 (2)	56.6	75.0	18.4	14.5 (22)
GPVI-D1D2 short Fc	44.8	5.2 (2)	50.0	53.3	3.4	44.8	2.8 (2)	47.6	51.4	3.8	– (0)
GPVI-234 short-Fc	47.7	5.2 (2)	52.9	62.7	9.7	47.7	2.8 (2)	50.5	60.9	10.3	6.6 (10)
GPVI-261 short-Fc	50.6	5.2 (2)	55.8	71.3	15.6	50.6	2.8 (2)	53.4	70.4	17.1	14.5 (22)

To identify potential regions of GPVI interacting with SSL5, we measured SSL5-binding to the various GPVI-ectodomain constructs in both glycoforms (Figure 6A). *g*-GPVI D1D2 binds SSL5 with an apparent affinity of 1.3 μ M (Figure 6C and Table 4) and a total capacity that corresponds with one molecule SSL5 per molecule of GPVI (Figure 6E). Responses observed for *m*-GPVI D1D2 were substantially smaller, and did not reach equimolar SSL5-GPVI binding (Figure 6E). Apparently, in absence of complex *N*-linked glycosylation in the D1-domain of GPVI, SSL5 does not interact with the extracellular Ig-like domains of GPVI. SSL5-binding to the glycosylated form of these domains must therefore be exclusively mediated by SSL-glycan interactions.

The number of SSL5-molecules binding GPVI increased to three if part of the stalk-region was included (*g*-GPVI-234), or four if the full-length stalk was present (*g*-GPVI-261; Figure 6E and Table 4). Simultaneously, the affinity for SSL5-binding increased, with K_D -values in the range of 0.66-1.0 μ M; values that are similar to the K_D of SSL5-PSGL-1 binding

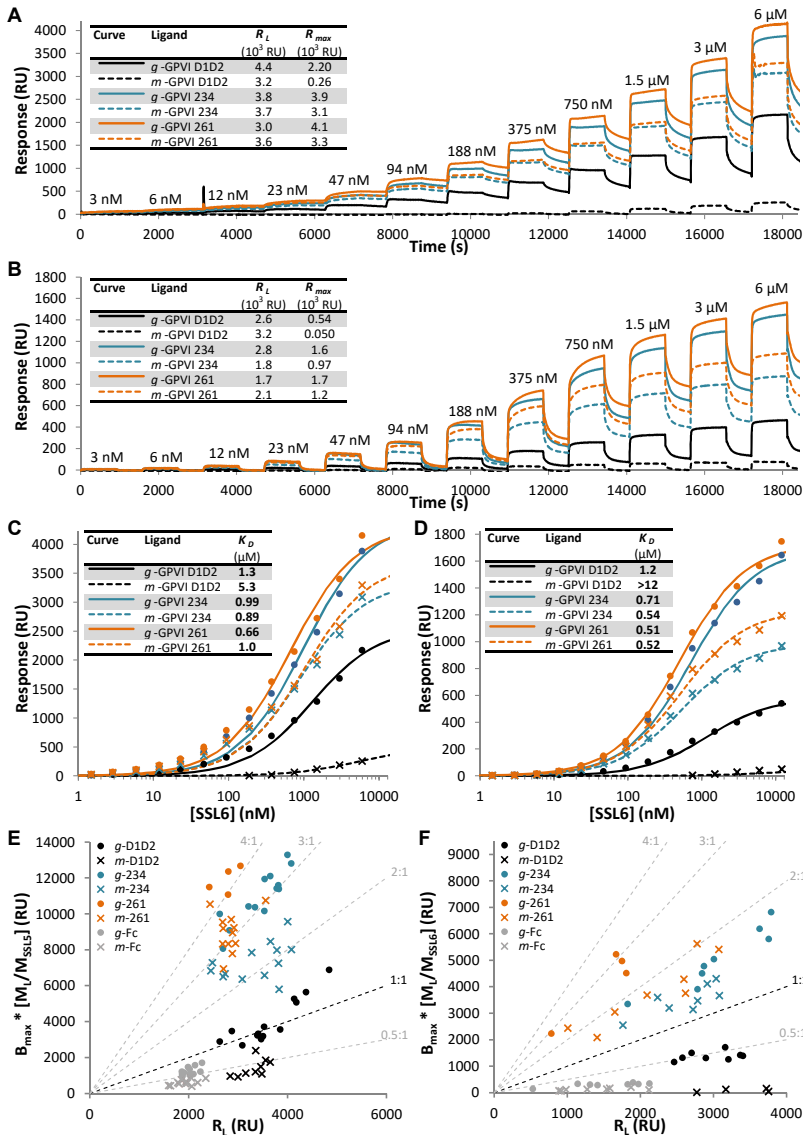


Figure 6: SSL5- and SSL6 binding to GPVI ectodomains. (A-B) Kinetic titration experiment using successive injections of stepwise increased concentrations (3 nM - 6 μ M) of SSL5 (A) or SSL6 (B) to immobilized GPVI-ectodomains. Responses (response units, RU) of SSL-binding were measured to the glycosylated (solid lines) and mannosylated (dashed lines) forms of GPVI-D1D2 (black), GPVI-234 (light blue), and GPVI-261 (orange lines). SSL5-binding was measured in three different experiments with at least four independent GPVI-regions per measurement, SSL6-binding was measured in two different experiments with equal GPVI-regions, of which curve per representative region is shown here. Titration cycles consisted of a 900-sec. association phase directly followed by a 750-sec. dissociation phase. The inset shows the initial ligand capture responses (R_i) and the maximal SSL-binding (R_{max}) reached for the depicted curves. (C-D) Non-linear regression to a 1:1 binding model of end-of-injection points measured for SSL5- (C) and SSL6-binding (D) to these GPVI-constructs. (E-F) Relation between maximum SSL5- (E) or SSL6- (F) binding (B_{max}) vs. the initial ligand capture responses (R_i) of immobilized GPVI-D1D2 (black), GPVI-234 (light blue), GPVI-261 (orange), or Fc (gray). For an equimolar comparison of the initial ligand and analyte responses, B_{max} -values have been corrected by the molecular mass ratio $[M_L/M_{SSL}]$ between the corresponding ligands and the SSL-analyte.

(0.82 μM)¹¹. The mannosylated forms of these proteins bound one molecule SSL5 per GPVI-molecule less, which is in agreement with one extra SSL5-molecule binding to the D1-domain *N*-glycan of *g*-GPVI-234 and -261. The putative number of sialylated *O*-glycans in the GPVI-stalk far exceeds the number of SSL5-molecules that can bind GPVI. Possibly, *O*-glycosylation is incomplete, or the number of SSL5-molecules that can simultaneously bind GPVI is limited by sterical factors.

SSL6-binding to GPVI-constructs very much resembles binding of SSL5 (Figure 6B, D, and F). Both SSLs bind *g*-D1D2 with equal affinity. SSL6 binds somewhat stronger to GPVI-proteins including (part of) the stalk-domain, and its total binding is slightly higher (Figure 6F). Overall, our binding experiments do not indicate GPVI-specific interaction for both SSL5 and SSL6. Following from that, GPVI-binding mediated by SSL5 or SSL6 involves interactions of their sLe^x-binding sites with *N*- and *O*-glycans in GPVI. Although we could reproduce previous results by Hu et al.²¹ demonstrating binding of SSL5 to glycosylated GPVI, their conclusion that GPVI is specifically targeted by SSL5 is wrong, since SSL6 binds GPVI similarly.

Table 4: Kinetic parameters of SSL6- and SSL5-binding to GPVI ectodomains

Protein	GPVI-SSL5			GPVI-SSL6		
	K_p (μM)	95% interval (μM)	SSL5:GPVI	K_p (μM)	95% interval (μM)	SSL6:GPVI
<i>g</i> -GPVI-D1D2	1.3	1.16 - 1.44	1.1 \pm 0.2	1.2	1.0 - 1.3	0.5 \pm 0.1
<i>g</i> -GPVI-234	0.99	0.90 - 1.1	3.2 \pm 0.2	0.71	0.66 - 0.76	1.6 \pm 0.1
<i>g</i> -GPVI-261	0.66	0.54 - 0.78	4.3 \pm 0.3	0.51	0.48 - 0.54	2.8 \pm 0.3
<i>g</i> -Fc	1.5	1.2 - 1.7	0.6 \pm 0.1	0.77	0.68 - 0.86	0.2 \pm 0.1
<i>m</i> -GPVI-D1D2	5.3	3.6 - 7.0	0.4 \pm 0.1	>12	-	0.1 \pm 0.1
<i>m</i> -GPVI-234	0.89	0.80 - 0.98	2.2 \pm 0.4	0.54	0.50 - 0.57	1.3 \pm 0.2
<i>m</i> -GPVI-261	1.0	0.90 - 1.1	3.2 \pm 0.5	0.52	0.48 - 0.56	1.8 \pm 0.3
<i>m</i> -Fc	1.6	1.4 - 1.8	0.3 \pm 0.1	0.69	0.57 - 0.80	0.1 \pm 0.1

Discussion

Infectious organisms often use glycan-binding mechanisms for cell attachment, progression and pathogenicity³⁶, such as influenza virus hemagglutinins that promote host-cell entrance³⁷ and *E. coli* fimbriae that anchor the bacterium to the urothelial surface³⁸. *S. aureus* secretes a range of virulence factors including the SSLs that interfere with the host cell immune system and prevent recognition and subsequent clearance of the bacterium^{4,39}. A subset of these SSLs, SSL2-6 and -11, bind oligosaccharides sLe^x and sLacNac, which are abundantly present on the immune cell surface²². The capacity of SSLs in this lectin-subset to associate with glycan structures supports the hypothesis that SSL-specific interference in vital defense processes of the host is mediated by glycan-dependent mechanisms.

In most cases, contributions of SSL-glycan binding have been estimated based on mutations in the sLe^x-binding site of SSL^{13,40}, neuraminidase-treatment of the target protein^{9,11,14}, or comparison with other sLe^x-containing substrates¹². This can, however, result in contradictory results, as obtained for e.g. TLR2-inhibition by SSL3^{13,27,40} and Fc α RI-binding by SSL11¹², and may ultimately lead to the misidentification of immune

targets if SSL-binding would purely rely on SSL-sLe^x interactions. On the other hand, since innate immune mechanisms also often strongly depend on glycan-binding mechanisms, interference by non-specific SSL-sLe^x binding itself could drastically alter the vital processes of the host-cells. Both situations encourage further investigations on the sialobiology of these immune evasion molecules.

We analyzed SSL-specificity and sLe^x-contributions in three cases of previously reported host-microbe interactions: SSL3-TLR2^{13,27}, SSL6-CD47¹⁴, and SSL5-GPVI²¹. For two receptors, TLR2 and CD47, binding could not be mimicked by one of the other lectin-SSLs. Receptor-binding in these cases is not based on common sLe^x-binding properties, but on protein-protein interactions. By contrast, GPVI bound SSL5 and SSL6 equally well, and their interaction is devoid of a protein-protein specific component. Our binding experiments show that SSL3 and SSL6 specifically target their respective receptors by protein-protein interactions.

It is intriguing that both SSL3 and SSL6 interact only marginally with TLR2-glycans in our experiments, suggesting the absence of functional glycan structures for SSL-binding. These data support the notion that the role of sLe^x-binding by SSL3 must be binding to non-TLR2 glycans at the immune cell surface. Except for matrix metalloproteinases¹⁸, all immune targets of lectin-SSLs thus far published are situated at the host cell surface. This suggests a common role of sLe^x-binding by these SSLs to pre-concentrate in regions that determine immune activation by the host.

SSL6-CD47 binding appears to have glycan-dependent and glycan independent contributions, of which the latter implies that CD47 is a specific target of SSL6. Glycan-dependent contributions would have been expected on the basis of the highly glycosylated state of its extracellular IgSF-domain: the sLe^x-binding site of SSL6 would be able to reach one of the CD47 glycan antennas in almost all orientations of binding. We show that one glycan in addition to the structurally important N34-glycan is required for cooperative high-affinity SSL6-binding. To exclude that the mixed interaction observed in our experiments is an artefact of the dimeric CD47-form, these results must also be verified using monomeric CD47, the functional of receptor-form *in vivo*³⁵.

Overall, this novel experimental approach using receptor-Fc fusions and multi-channel SPR binding experiments provides a solid basis for further studies to sialic acid-dependent SSL-interactions. Our findings demonstrate potential to discriminate between glycan-enhanced (SSL6-CD47), glycan-independent (SSL3-TLR2), and glycan-only interactions (SSL5-GPVI). These results may therefore contribute to a better understanding of the host-pathogen interaction.

Methods

Expression and purification of SSL3, SSL5, SSL6

Staphylococcal superantigen-like proteins SSL3, SSL5 and SSL6 have been expressed in *E. coli* Rosetta-gami(DE3)-*pLysS*, and purified as previously described^{11,13,14}.

Cloning, expression, and purification of Fc-fusion proteins

Constructs of receptor ectodomains followed by an Fc-tag were *in frame* cloned in a HEK293 expression cassette system with Cystatin signal peptide and using the 5' BamHI and 3' NotI restriction sites for insertion of target genes, and the 5' NotI and 3' SphI sites

for introduction of a segment coding for a TEV-cleavable cysteine-bridged Fc-tag ('TEV-Fc') or a non-cleavable and non-covalently linked short Fc-tag ('short-Fc'). The TEV-Fc tag contains the TEV-protease cleavable ENLYFQGS-site followed by the hinge-CH2-CH3 domains (res. 99-330) of the Ig gamma-1 chain C region (Uniprot ID: P01857). The short Fc-tag only contains the CH2-CH3 domains (res. 120-330) preceded by a four-amino acid AAAG-linker that results from translation of the NotI-site plus an additional glycine. An additional construct was made that only contains the immunoglobulin CH2-CH3 domains without receptor domains.

Fc-fusion proteins were expressed with an N-terminal Cystatin-S signal peptide in HEK293-EBNA1 and HEK293-EBNA1-S (GNTI⁻) cells (U-Protein Express), and purified six days post-transfection by Protein-A affinity chromatography (GE Healthcare). Proteins were eluted with 0.1 M glycine pH 3.0 and 100 mM NaCl and collected in tubes containing 1/10 volume of neutralization buffer (1 M Tris-Cl pH 9.0). Protein purities and yields were assessed by A280 absorption and SDS-PAGE.

Cloning and expression of Leukocyte surface antigen CD47 C33G and glycan mutants

A 5'BamHI-3'NotI fragment comprising the residues 19-134 of the extracellular immunoglobulin superfamily (IgSF-)domain of CD47 (Uniprot-ID: Q08722) was cloned using primers listed in Table 5, which concurrently introduced the Cys33Gly-mutation of the cysteine that is natively involved in disulfide bond formation with Cys263 in the most C-terminal extracellular loop. The Cys33Gly mutant is further designated as wild type CD47 in this manuscript, since this mutation was previously shown to be silent with respect to SIRPβ-binding³⁵, and, regarding its inaccessible location *in vivo* near the membrane, would likewise expected to be in the case of SSL-binding. In addition to the Cys33Gly mutation, single, double, triple, and quadruple point-mutations eliminating one, two, three, or four out of five N-linked glycosylation sites (on Asn23, Asn34, Asn50, Asn73, and Asn111) were made by successive rounds of QuikChange site-directed mutagenesis and the primers listed in Table 5. All mutations were verified by DNA-sequencing.

Table 5. PCR-primers used for cloning of the CD47 C33G and glycan point mutants

Primer	(Q)-PCR	+/-	Sequence*
CD47-C33G-Fw	PCR	+	5'- GGATCC CAGCTACTATTTAATAAAAACAAAATCTGTAGAA-TTCACGTTT <u>GG</u> TAATGACACTGTCGTCATTCC-3'
CD47-Rv	PCR	-	5'- GCGGCCG CAACAACACGATATTTTAGCTCGATGATCG-3'
CD47-N23Q-Fw	Q-PCR	+	5'-GGATCCAGCTACTATTTT <u>CAG</u> AAAAACAAAATCTGTAGAATTCACG-3'
CD47-N23Q-Rv	Q-PCR	-	5'-CGTGAATTCTACAGATTTTGT <u>TTTCT</u> GAAATAGTAGCTGGGATC-3'
CD47-T36N-Fw	Q-PCR	+	5'-CACGTTTGGTAATGACA <u>AA</u> IGTCGTCATTCCATGC-3'
CD47-T36N-Rv	Q-PCR	-	5'-GCATGGAATGACGAC <u>ATT</u> GTGCATTACCAAACGTG-3'
CD47-N50Q-Fw	Q-PCR	+	5'-CTAATATGGAGGCACA <u>ACAG</u> ACTACTGAAGTATACG-3'
CD47-N50Q-Rv	Q-PCR	-	5'-CGTATACTTCAGTAGT <u>CTG</u> TTGTGCTCCATATTAG-3'
CD47-N73Q-Fw	Q-PCR	+	5'-CACCTTTGATGGAGCTCTA <u>CAGA</u> AGTCCACTGTCCCCACTG-3'
CD47-N73Q-Rv	Q-PCR	-	5'-CAGTGGGGACAGTGGACT <u>TCTG</u> TAGAGCTCCATCAAAGGTG-3'
CD47-N111Q-Fw	Q-PCR	+	5'-GCTGTCTCACACAGGAC <u>AGT</u> ACTACTGTGAAGTAACAG-3'
CD47-N111Q-Rv	Q-PCR	-	5'-CTGTTACTTACAAGTGTA <u>CTG</u> CTCTGTGTGTGAGACAGC-3'

* Mutated sites are underlined, flanking BamI- and NotI-sequences are shown in bold fonts.

CD47 proteins in dimeric form using the C-terminal TEV-Fc tag or short Fc-tag were expressed and purified as described above. Wild type CD47 and quadruple mutants were also expressed in monomeric form using constructs coding for the same CD47-region, directly followed by a non-cleavable His₆-tag at its C-terminus. His-tagged CD47 was purified by IMAC using Ni-NTA resin (GE Healthcare) and 10 mM HEPES-buffer pH 7.5 with 150 mM NaCl and 15 mM Imidazole as wash-buffer. Proteins were eluted in two steps with 100 mM and 500 mM Imidazole in HEPES-buffer, respectively, and purities and yields were assessed by A280 absorption and SDS-PAGE.

Cloning and expression of glycoprotein-VI (GPVI) ectodomains

Three variants of the GPVI ectodomain (Uniprot ID: Q9HCN6-3) were cloned as BamI-NotI fragments using the primers listed in Table 6: 1) GPVI D1-D2, only containing the immunoglobulin-like D1-D2 domains⁴¹ (res. 21-206); 2) GPVI-234, the D1-D2 domains followed by part of the *O*-glycosylated stalk⁴² (res. 21-234); 3) GPVI-261, the D1-D2 domains followed by the full length stalk (res. 21-261). These GPVI-variants were expressed as TEV-Fc- and short-Fc fusion proteins in both HEK293-EBNA1 and HEK293-EBNA1-S cells and purified as described above.

Table 6. PCR-primers used for cloning of GPVI-ectodomains.

Primer	GPVI-residues	+/-	Sequence*
GPVI-Fw		+	5'- GGATCCC AGAGTGGACCGCTCCCC-3'
D1D2-Rv	Gln21-Ser206	-	5'- GCGGCCG CAGAGGTTCTGTGACCAC-3'
234-Rv	Gln21-Phe234	-	5'- GCGGCCG CGAATGAGACGGTCAGTTCAGCG-3'
261-Rv	Gln21-Ser261	-	5'- GCGGCCG CGGACTCACCAGCTGGAGAGTC-3'

* Flanking BamI- and NotI-sequences are shown in bold fonts.

Cloning and expression of homodimeric Toll-like receptors (TLR) 1, 2, and 6

Ectodomains of murine Toll-like receptors 1 (Uniprot ID: Q9EPQ1, res. 25-583), 2 (Uniprot ID: Q9QUN7, res. 25-588)²⁷, and 6 (Uniprot ID: Q9EPW9; res. 28-585) that all comprise the respective LRR-repeats and native LLR-cap-domains were cloned as BamI-NotI or BglII-NotI fragments using the primers listed in Table 7. TEV-Fc and short-Fc fusion proteins were expressed in HEK293-EBNA1 and HEK293-EBNA1-S cells by plasmid dilution⁴³, using 10-fold dilutions of expression plasmid in noncoding dummy plasmid, and cotransfecting with a 40-fold dilution of PRAT4A as previously described for monomeric mTLR2²⁷. TLR-Fc homodimers were purified as described above.

Table 7. PCR-primers used for cloning of murine TLR-ectodomains.

Primer	Residues	+/-	Sequence*
mTLR1-Fw	Ser25-Val583	+	5'- GGATCCG AGAAATGTGAGCTTATCATAAAGAGG-3'
mTLR1-Rv	Ser25-Val583	-	5'- GCGGCCG CAACAGTATCACAGGATAGTGGAGACATG-3'
mTLR2-Fw	Gln25-Ala588	+	5'- AGATCTC AGGAGTCTCTGTCTCATGTGATG-3'
mTLR2-Rv	Gln25-Ala588	-	5'- GCGGCCG CAGCCTGGTGACATTCCAAGACG-3'
mTLR6-Fw	Phe28-Val585	+	5'- GGATCCT TCTCTAATGAACCTTGAGTCTATGG-3'
mTLR6-Rv	Phe28-Val585	-	5'- GCGGCCG CAACAGTATCACAGGACAGTGG-3'

* Flanking BamI-, BglII-, and NotI-sequences are shown in bold fonts.

Analytical size exclusion chromatography

Receptor-Fc fusion proteins were analyzed by analytical size exclusion chromatography on a Nexera X2 HPLC-system (Shimadzu) equipped with an Superdex200 5/150-column (GE Healthcare) and equilibrated with 10 mM HEPES-buffer pH 7.5 that also contains 300 mM NaCl. Protein fluorescence was detected using excitation and emission wavelengths of 275 and 354 nm, respectively. Gelfiltration chromatograms have been normalized to maximum fluorescence.

Immobilization of Fc-fusion proteins on ProteinA/G-SensEye SPR-chips

For SPR binding experiments, receptor proteins with either a TEV-Fc or short Fc-tag were immobilized to a gel-type Protein-A/G SensEye SPR-sensor (IBIS Technologies) using a Continuous Flow Microspotter (CFM; Wasatch Microfluidics). Receptor molecules were diluted in PBS to final concentrations of 0.5-2.0 $\mu\text{g}/\text{mL}$ and final volumes of 150 μL . Coupling reactions of 48 samples in a 4x12 distribution were performed in 60 cycles of 1 minute, preceded by a buffer flow of 5 minutes. SPR-chips were transferred to the Instrument for Biomolecular Interaction Sensing Multiplex 96 (IBIS-MX96; IBIS Technologies) immediately after completion of the coupling protocol and equilibrated with running buffer (PBS + 0.005% Tween-20). Regions of interest (ROIs) containing ligand samples and blanc spots were selected manually. Estimation of the response of ligand capture (R_L) was performed by the automated preparation protocol at the beginning of a titration experiment, comparing the response units (RUs) of individual ROIs with two nearby located blanc regions.

Surface plasmon resonance (SPR) binding experiments

SPR-binding experiments were performed on an IBIS-MX96 instrument equilibrated with PBS running buffer containing 0.005% Tween-20. Serial 2-fold dilutions of SSL3, SSL5, or SSL6 in running buffer were prepared, starting from 6 μM (SSL3 and SSL5) or 12 μM (SSL6) and down to 3 nM as lowest concentration. The response of SSL-binding in RU was measured in a kinetic titration experiment by 100 μL -injections of increasing concentrations of SSL-analyte. Consecutive reaction cycles consist of a 15 (binding to CD47-Fc or GPVI-Fc) or 60 minutes (binding to mTLR2-Fc) association phase (back-and-forth flowing), followed by a 10-15 minutes dissociation phase (unidirectional buffer flowing), and were continued without intermediate regeneration cycles. Baseline stabilization was performed by a series of blanc cycles prior to injection of the first SSL-analyte.

Processing of binding data

SPR-binding data was processed using the SprintX-software (v.1.11.1.4; IBIS Technologies) provided along with the IBIS-MX96 instrument. Responses of analyte binding to all 48 ROIs within one measurement were referenced to local blanc regions situated in between rows of ligand ROIs, to subtract signals caused by small buffer effects as observed for these reference ROIs at injections of concentrated analyte. Referenced chromatograms were zeroed at the end of the dissociation phase of the final blanc sample prior to injection of the first SSL-analyte. Further data processing and regression was performed using SigmaPlot software (Systat Software Inc).

Thermodynamical analysis of the binding data was done by measuring the end-of-injection response of curves at the end of association phases for each concentration of SSL-analyte. Per ligand, binding data of separate ROI replicates were globally fitted in

SigmaPlot by non-linear regression to formula 1 for 1:1 ligand binding, to obtain shared K_D -values and individual B_{max} -values per ROI.

$$(1) R_{eq} = \frac{B_{max} \times C}{K_D + C}, \text{ with } R_{eq} \text{ the response (RU) at equilibrium, } c \text{ the concentration of SSL-analyte (M), and } K_D \text{ the dissociation constant for SSL-binding.}$$

In the case of SSL6-CD47 binding, in which binding of SSL6 clearly appeared multiphasic as well as partially cooperative, data were globally fitted by non-linear regression to formula 2 for a two-site binding model, to obtain shared $K_{D,1}$ - and $K_{D,2}$ -values. To improve fitting of the first, cooperative SSL6-binding step, the $B_{max,1}$ -values were thereby restrained to the observed end-of-injection responses at 93 nM SSL6.

$$(2) R_{eq} = \frac{B_{max,1} \times C}{K_{D,1} + C} + \frac{B_{max,2} \times C}{K_{D,2} + C}$$

Cooperativity of analyte binding was analyzed using Hill's equation (formula 3). Data was normalized to B_{max} or the local equilibrium maximum $B_{max,n}$ in the case of multiphasic analyte binding, providing the fraction f of receptor binding sites bound to SSL-analyte at every analyte concentration. Data was plotted in the linearized form $x = \log([SSL])$ and $y = \log[f / (1-f)]$, with the observed dissociation constant $K_{D,obs}$ at the intersect with the x-axis.

$$(3) f = \frac{C^{n_H}}{K_D + C^{n_H}}, \text{ with } c \text{ the concentration of SSL-analyte (M), and } n_H \text{ the Hill coefficient.}$$

References

1. Lowy, F. D. Staphylococcus aureus infections. *N. Engl. J. Med.* **339**, 520–32 (1998).
2. Bassetti, M., Nicco, E. & Mikulska, M. Why is community-associated MRSA spreading across the world and how will it change clinical practice? *Int. J. Antimicrob. Agents* **34 Suppl 1**, S15–9 (2009).
3. Liu, G. Y. Molecular pathogenesis of Staphylococcus aureus infection. *Pediatr. Res.* **65**, 71R–77R (2009).
4. Koymans, K. J., Vrieling, M., Gorham, R. D. & van Strijp, J. A. G. in *Life Science Journal* **6**, 23–27 (2015).
5. Langley, R. *et al.* The staphylococcal superantigen-like protein 7 binds IgA and complement C5 and inhibits IgA-Fc alpha RI binding and serum killing of bacteria. *J. Immunol.* **174**, 2926–33 (2005).
6. Laursen, N. S. *et al.* Structural basis for inhibition of complement C5 by the SSL7 protein from Staphylococcus aureus. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 3681–6 (2010).
7. Bestebroer, J. *et al.* Functional basis for complement evasion by staphylococcal superantigen-like 7. *Cell. Microbiol.* **12**, 1506–1516 (2010).
8. Patel, D. *Functional and Structural Characterization of Staphylococcal Superantigen-Like Protein 10 (SSL10)*. Thesis, (The University of Auckland (Thesis), 2011).
9. Bestebroer, J. *et al.* Staphylococcal SSL5 inhibits leukocyte activation by chemokines and anaphylatoxins. *Blood* **113**, 328–337 (2009).
10. Walenkamp, A. M. E. *et al.* Staphylococcal superantigen-like 10 inhibits CXCL12-induced human tumor cell migration. *Neoplasia* **11**, 333–44 (2009).
11. Bestebroer, J. *et al.* Staphylococcal superantigen-like 5 binds PSGL-1 and inhibits P-selectin-mediated neutrophil rolling. *Blood* **109**, 2936–43 (2007).
12. Chung, M. C. *et al.* The crystal structure of staphylococcal superantigen-like protein 11 in complex with sialyl Lewis X reveals the mechanism for cell binding and immune inhibition. *Mol. Microbiol.* **66**, 1342–55 (2007).
13. Bardoel, B. W. *et al.* Evasion of Toll-like receptor 2 activation by staphylococcal superantigen-like protein 3. *J. Mol. Med.* **90**, 1109–1120 (2012).
14. Fevre, C. *et al.* Staphylococcus aureus proteins SSL6 and SEIX interact with neutrophil receptors as identified using secretome phage display. *Cell. Microbiol.* **16**, 1646–1665 (2014).
15. Ramsland, P. A. *et al.* Structural basis for evasion of IgA immunity by Staphylococcus aureus revealed in the complex of SSL7 with Fc of human IgA1. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 15051–6 (2007).
16. Patel, D., Wines, B. D., Langley, R. J. & Fraser, J. D. Specificity of staphylococcal superantigen-like protein 10 toward the human IgG1 Fc domain. *J. Immunol.* **184**, 6283–92 (2010).
17. Itoh, S. *et al.* Staphylococcal superantigen-like protein 10 (SSL10) binds to human immunoglobulin G

- (IgG) and inhibits complement activation via the classical pathway. *Mol. Immunol.* **47**, 932–938 (2010).
18. Koymans, K. J. *et al.* Staphylococcal superantigen-like protein 1 and 5 (SSL1 & SSL5) limit neutrophil chemotaxis and migration through MMP-inhibition. *Int. J. Mol. Sci.* **17**, 1–16 (2016).
 19. Itoh, S. *et al.* Staphylococcal superantigen-like protein 5 inhibits matrix metalloproteinase 9 from human neutrophils. *Infect. Immun.* **78**, 3298–3305 (2010).
 20. de Haas, C. J. C. *et al.* Staphylococcal superantigen-like 5 activates platelets and supports platelet adhesion under flow conditions, which involves glycoprotein Iba α and α IIb β 3. *J. Thromb. Haemost.* **7**, 1867–74 (2009).
 21. Hu, H. *et al.* GPVI and GPIba mediate staphylococcal superantigen-like protein 5 (SSL5) induced platelet activation and direct toward glycans as potential inhibitors. *PLoS One* **6**, e19190 (2011).
 22. Munro, J. M. *et al.* Expression of sialyl-Lewis X, an E-selectin ligand, in inflammation, immune processes, and lymphoid tissues. *Am. J. Pathol.* **141**, 1397–1408 (1992).
 23. Baker, H. M. *et al.* Crystal structures of the staphylococcal toxin SSL5 in complex with sialyl Lewis X reveal a conserved binding site that shares common features with viral and bacterial sialic acid binding proteins. *J. Mol. Biol.* **374**, 1298–308 (2007).
 24. Hermans, S. J. *et al.* Structural and functional properties of staphylococcal superantigen-like protein 4. *Infect. Immun.* **80**, 4004–4013 (2012).
 25. Taylor, M. E. & Drickamer, K. *Introduction to Glycobiology*. (Oxford University Press, 2011).
 26. Cummings, R. D. Structure and function of the selectin ligand PSGL-1. *Brazilian J. Med. Biol. Res.* **32**, 519–528 (1999).
 27. Koymans, K. J. *et al.* Structural basis for inhibition of TLR2 by staphylococcal superantigen-like protein 3 (SSL3). *Proc. Natl. Acad. Sci.* **112**, 11018–11023 (2015).
 28. Jin, M. S. *et al.* Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **130**, 1071–82 (2007).
 29. Kang, J. Y. *et al.* Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity* **31**, 873–84 (2009).
 30. Oldenburg, P. A. *et al.* Role of CD47 as a marker of self on red blood cells. *Science* **288**, 2051–4 (2000).
 31. Oldenburg, P.-A. CD47: A Cell Surface Glycoprotein Which Regulates Multiple Functions of Hematopoietic Cells in Health and Disease. *ISRN Hematol.* **2013**, 614619 (2013).
 32. Willingham, S. B. *et al.* The CD47-signal regulatory protein α (SIRP α) interaction is a therapeutic target for human solid tumors. *Proc. Natl. Acad. Sci.* **109**, 6662–7 (2012).
 33. Horrigan, S. K. Replication Study: The CD47-signal regulatory protein α (SIRP α) interaction is a therapeutic target for human solid tumors. *Elife* **6**, 1–6 (2017).
 34. Jaiswal, S. *et al.* CD47 Is Upregulated on Circulating Hematopoietic Stem Cells and Leukemia Cells to Avoid Phagocytosis. *Cell* **138**, 271–285 (2009).
 35. Hatherley, D. *et al.* Paired receptor specificity explained by structures of signal regulatory proteins alone and complexed with CD47. *Mol. Cell* **31**, 266–77 (2008).
 36. Varki, A., Cummings, R. D. & Esko, J. D. in *Essentials of Glycobiology, Third Edition* (2009).
 37. Matrosovich, M., Herrler, G. & Klenk, H. D. in 1–28 (2013). doi:10.1007/128_2013_466
 38. Hartmann, M. & Lindhorst, T. K. The Bacterial Lectin FimH, a Target for Drug Discovery - Carbohydrate Inhibitors of Type 1 Fimbriae-Mediated Bacterial Adhesion. *European J. Org. Chem.* **2011**, 3583–3609 (2011).
 39. Thammavongsa, V., Kim, H. K., Missiakas, D. & Schneewind, O. Staphylococcal manipulation of host immune responses. *Nat. Rev. Microbiol.* **13**, 529–43 (2015).
 40. Yokoyama, R. *et al.* Staphylococcal superantigen-like protein 3 binds to the Toll-like receptor 2 extracellular domain and inhibits cytokine production induced by *Staphylococcus aureus*, cell wall component, or lipopeptides in murine macrophages. *Infect. Immun.* **80**, 2816–25 (2012).
 41. Horii, K., Kahn, M. L. & Herr, A. B. Structural basis for platelet collagen responses by the immune-type receptor glycoprotein VI. *Blood* **108**, 936–42 (2006).
 42. Miura, Y., Takahashi, T., Jung, S. M. & Moroi, M. Analysis of the interaction of platelet collagen receptor glycoprotein VI (GPVI) with collagen. A dimeric form of GPVI, but not the monomeric form, shows affinity to fibrous collagen. *J. Biol. Chem.* **277**, 46197–204 (2002).
 43. Halff, E. F., Versteeg, M., Brondijk, T. H. C. & Huizinga, E. G. When less becomes more: Optimization of protein expression in HEK293-EBNA1 cells using plasmid titration - A case study for NLRs. *Protein Expr. Purif.* **99**, 27–34 (2014).

Supplementary Figures

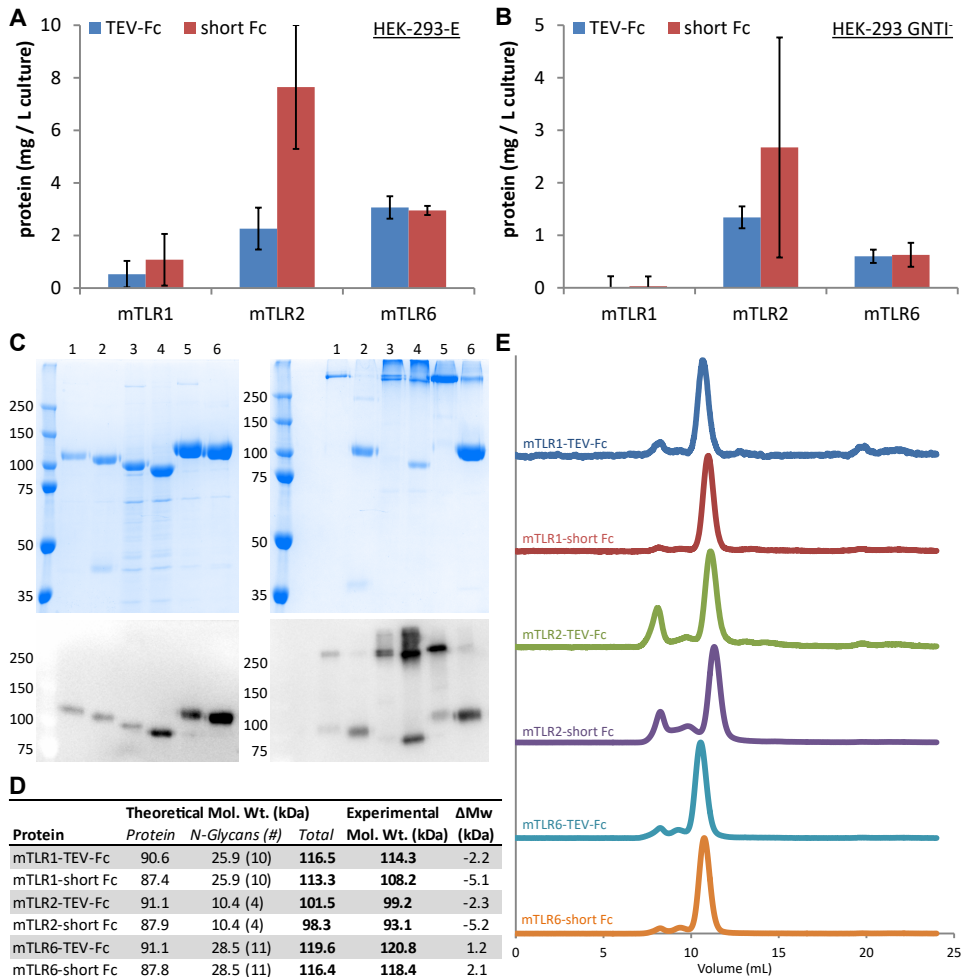


Figure S1: Expression of the LRR-domains of TLR1, -2, and -6. (A-B) Expression yields of Fc-tagged TLR1, -2, and -6 expressed in HEK293-E (*g*-TLR, A) or HEK293-S (*m*-TLR, B) and measured by A280 absorption. All constructs were expressed in three different experiments, and A280 absorption was measured three times per experiment. (C) SDS-PAGE analysis (top panels) of the purified glycosylated forms of TLR1-TEV-Fc (lane 1), TLR1-short Fc (lane 2), TLR2-TEV-Fc (lane 3), TLR2-short Fc (lane 4), TLR6-TEV-Fc (lane 5), TLR6-short Fc (lane 6), run under reducing conditions (top left panel) and non-reducing conditions (top right panel). Also cell-free culture media of the same expressions were analyzed with immunoblotting (bottom panels) and stained with the anti-human IgG-HRP conjugate. (D) Table showing the theoretical molecular weights of the expressed TLR-proteins and the experimental weights as observed on corresponding to the bands observed on SDS-PAGE gel in reduced form. (E) Analytical gel filtration chromatograms of glycosylated variants of these TLRs, showing equal retention volumes for TEV-Fc and short Fc-tagged receptor variants.

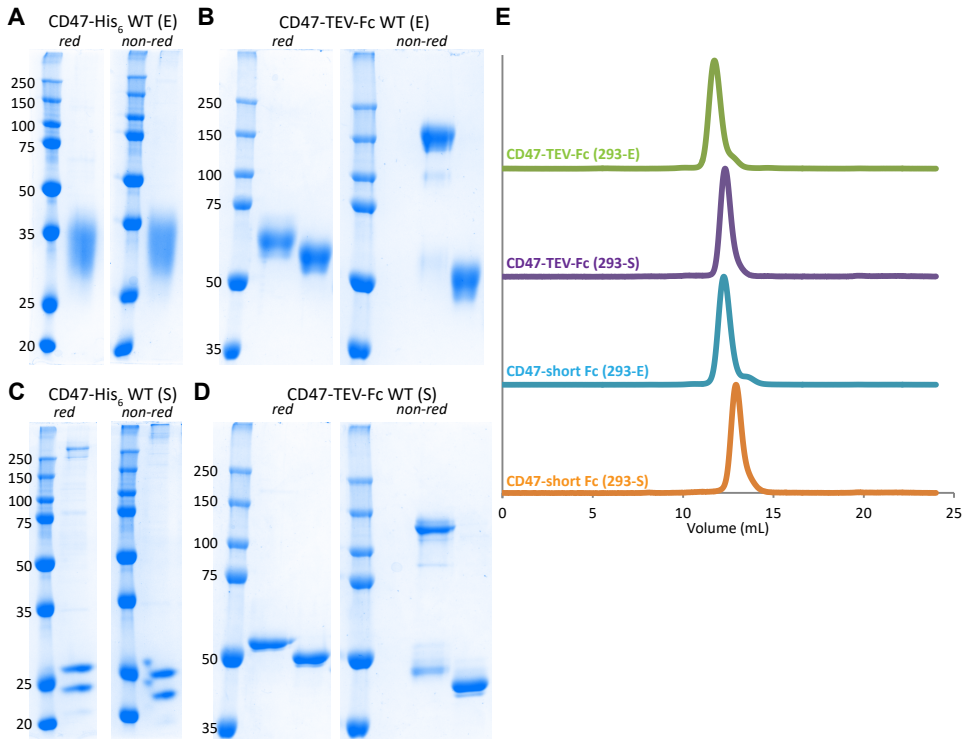
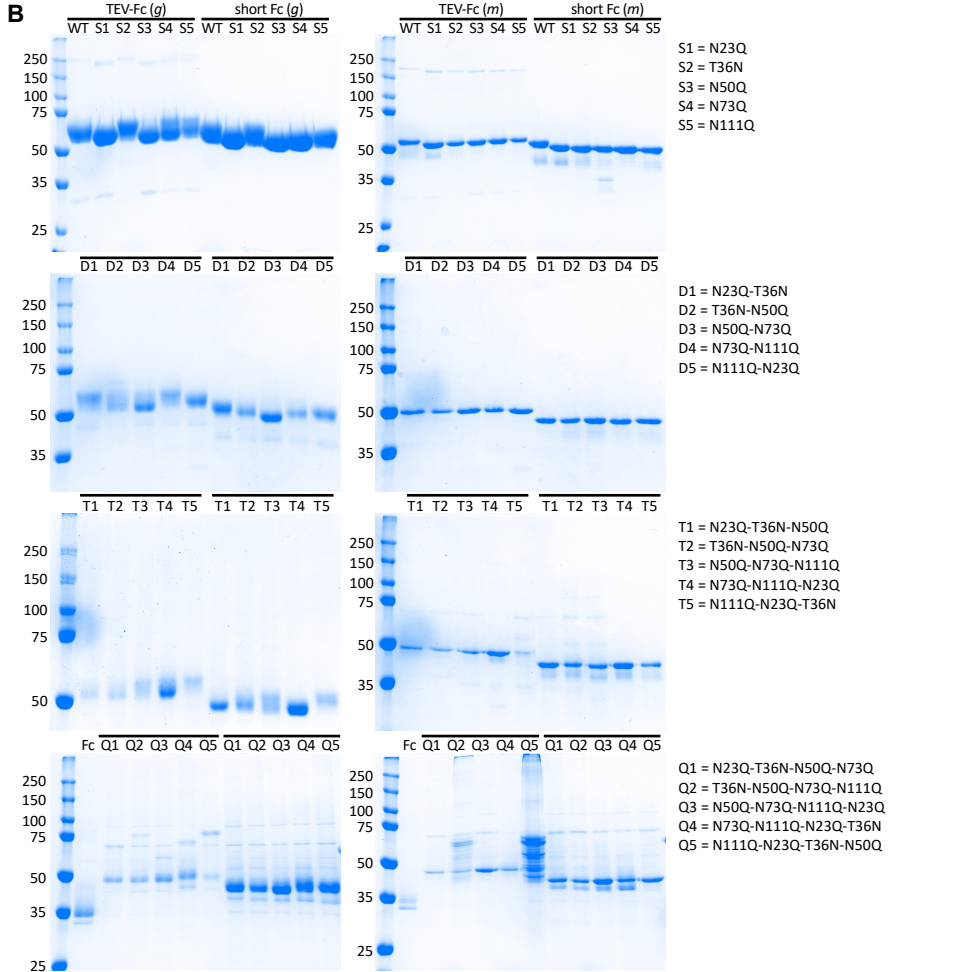
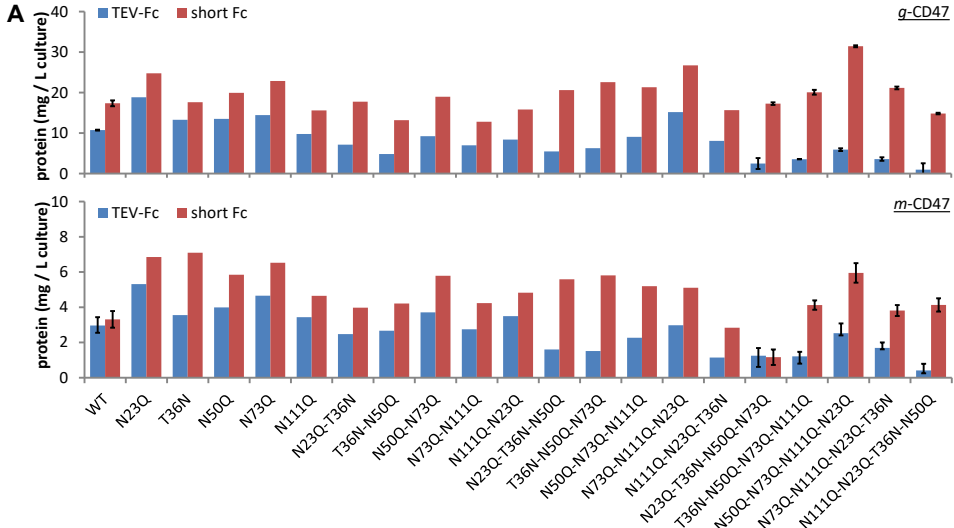


Figure S2: Expression of monomeric and dimeric CD47. (A-D) SDS-PAGE analysis of wild type CD47-His₆ (A & C), CD47-TEV-Fc (B & D) expressed in HEK-E and HEK-S cells and run both under reduced (left panels) and non-reduced (right panels) conditions. (E) Analytical gel filtration chromatograms of wild type CD47-TEV-Fc and CD47-short Fc expressed in HEK293-E and HEK293-S cells, respectively, showing similar retention volumes for all CD47 dimers. The CD47 dimers expressed in HEK-S are notably shifted, indicating lower molecular weight as observed by SDS-PAGE analysis.

Figure S3 (next page): Expression of CD47-TEV-Fc and CD47-short Fc mutants. (A) Expression yields of Fc-tagged CD47 mutants expressed in HEK293-E (*g*-CD47) or HEK293-S (*m*-CD47) measured by A280 absorption. Wild type CD47 and quadruple mutants were expressed in three different experiments, the remaining mutants were expressed and measured a single time. (B) SDS-PAGE analysis of TEV-Fc- and short Fc-tagged variants of wild type CD47, mutants and unfused Fc protein, expressed in HEK293-E (*g*) or HEK293-S (*m*). This analysis shows that wild type CD47 and the majority of the single and double mutants were stably expressed and protein purity was determined >95%. In these cases, glycosylated CD47 appears in multiple glycoforms, indicating that glycosylation of these proteins may not have reached completeness during expression. It is furthermore observed from the multiple bands corresponding to especially the TEV-Fc-tagged triple and quadruple mutants, that these proteins were less stably expressed and purity was determined <90%.



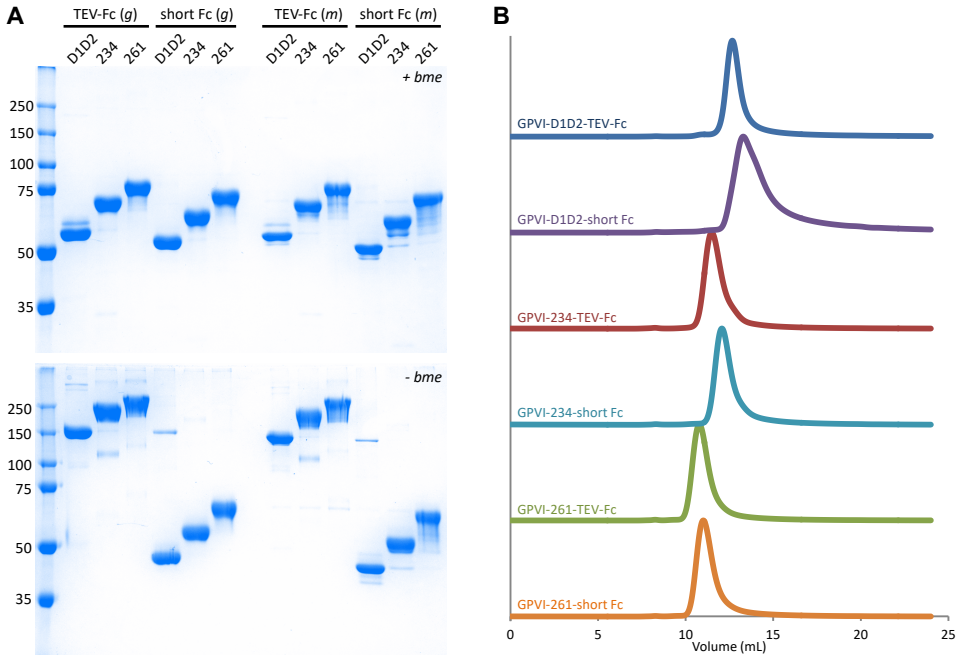


Figure S4: Expression of GPVI ectodomains with various stalk lengths. (A) SDS-PAGE analyses of TEV-Fc and short-Fc fusion protein expression of three GPVI-constructs in glycosylated (*g*) and mannosylated (*m*) glycoforms and run under reduced (top panel) and non-reduced conditions (bottom panel): ‘*D1D2*’, consisting of the D1-D2 domains only; ‘*234*’, the D1-D2 domains and also partially including the *O*-glycosylated stalk till res. 234; ‘*261*’, full GPVI-ectodomain consisting of the D1-D2 domains and the entire stalk till res. 261. **(B)** Analytical gel filtration chromatograms of Fc-tagged *g*-GPVI ectodomains indicating the dimeric state for both TEV-Fc and short-Fc proteins.

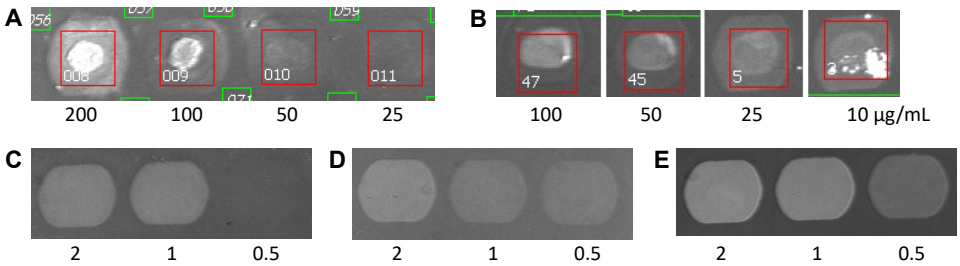


Figure S5: Immobilization of receptor ectodomains on SPR-chips. (A) Covalent amine-COOH coupling in sodium propionate buffer pH 5.0 of random free amino-groups TLR2 to a SPR-chip containing reactive carboxylic acid groups. **(B)** Biotin-streptavidin coupling in PBS of TLR2 after overnight biotinylation with BirA. **(C-E)** Fc-proteinA coupling in PBS of TLR2-Fc **(C)**, CD47-Fc **(D)**, and GPVI-Fc **(E)** by expression of Fc-fusion proteins. Protein concentrations ($\mu\text{g/mL}$) of the solutions used for immobilization are depicted below.

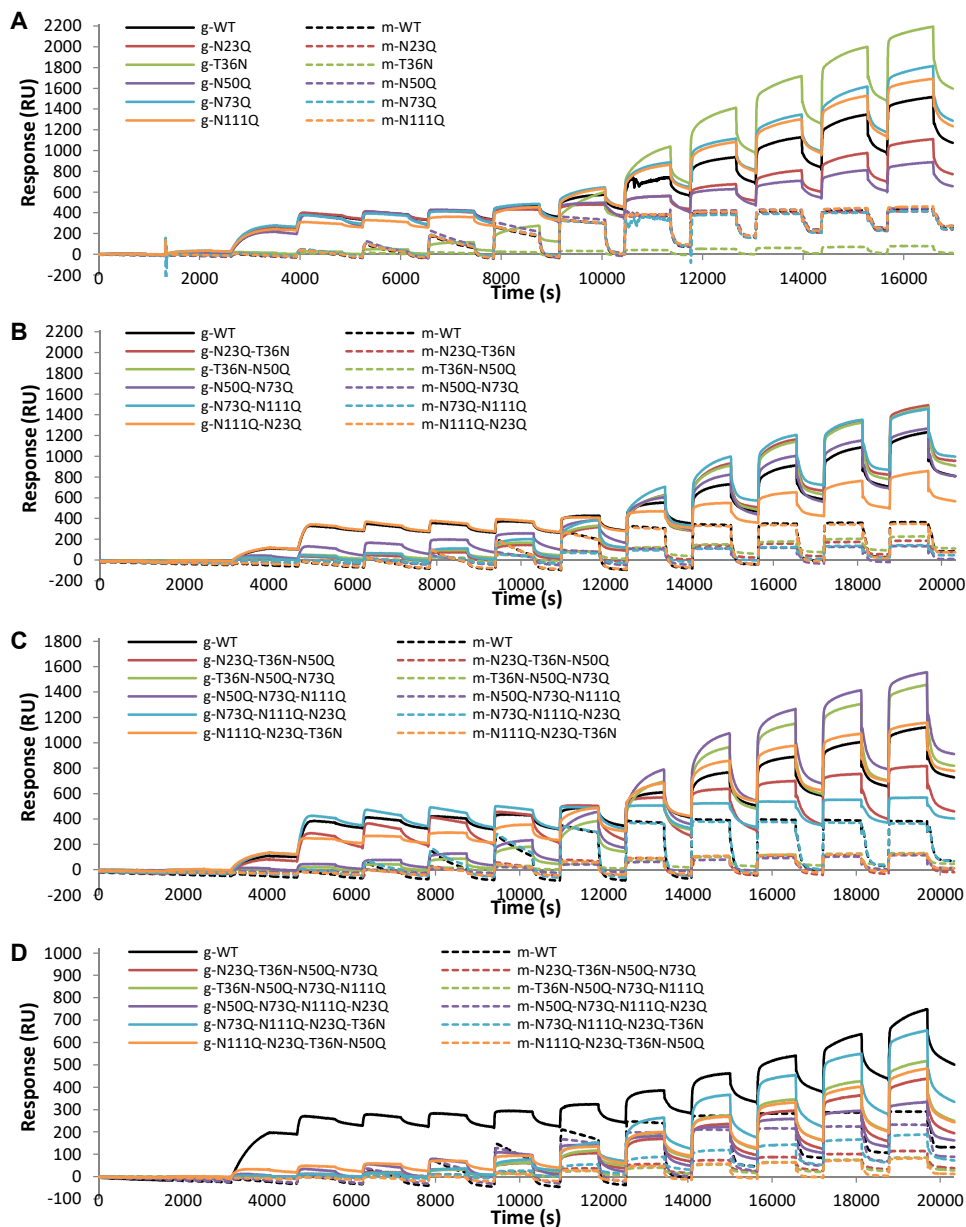


Figure S6: SSL6-binding to *g*-CD47 and *m*-CD47 glycan mutants. (A-D) Kinetic titration experiments using successive injections of stepwise increased SSL6-concentrations (3 nM - 6 μ M) to immobilized CD47-TEV-Fc wild type (black lines) and glycan single (A), double (B), triple (C) and quadruple (D) point mutants (colored lines). Responses (RU) of SSL6-binding were measured for regions containing glycosylated (*g*-)CD47 (solid lines) or mannosylated (*m*-)CD47 (dashed lines). Binding was measured to at least three independent ROIs per mutant, of which one representative curve is shown here.

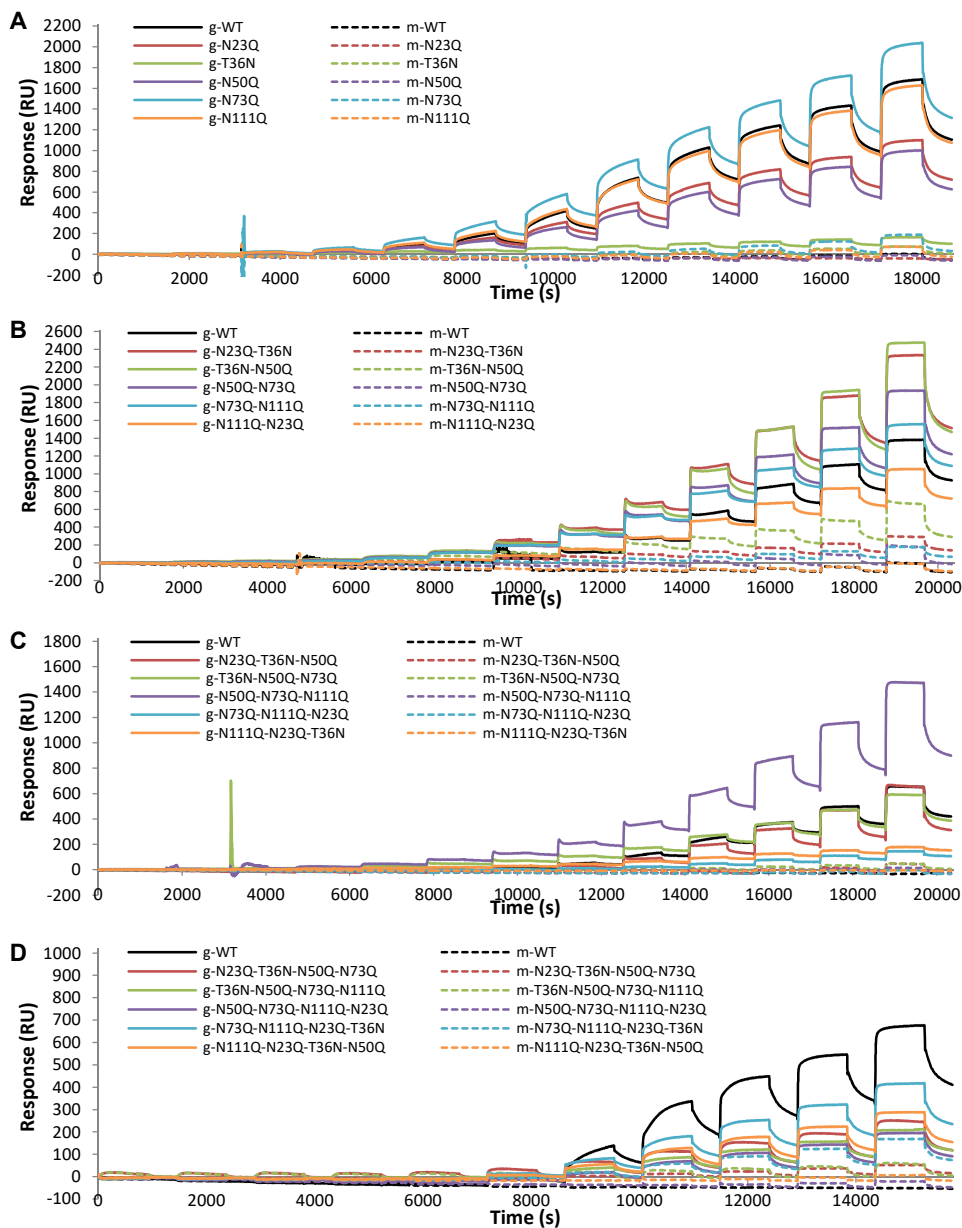
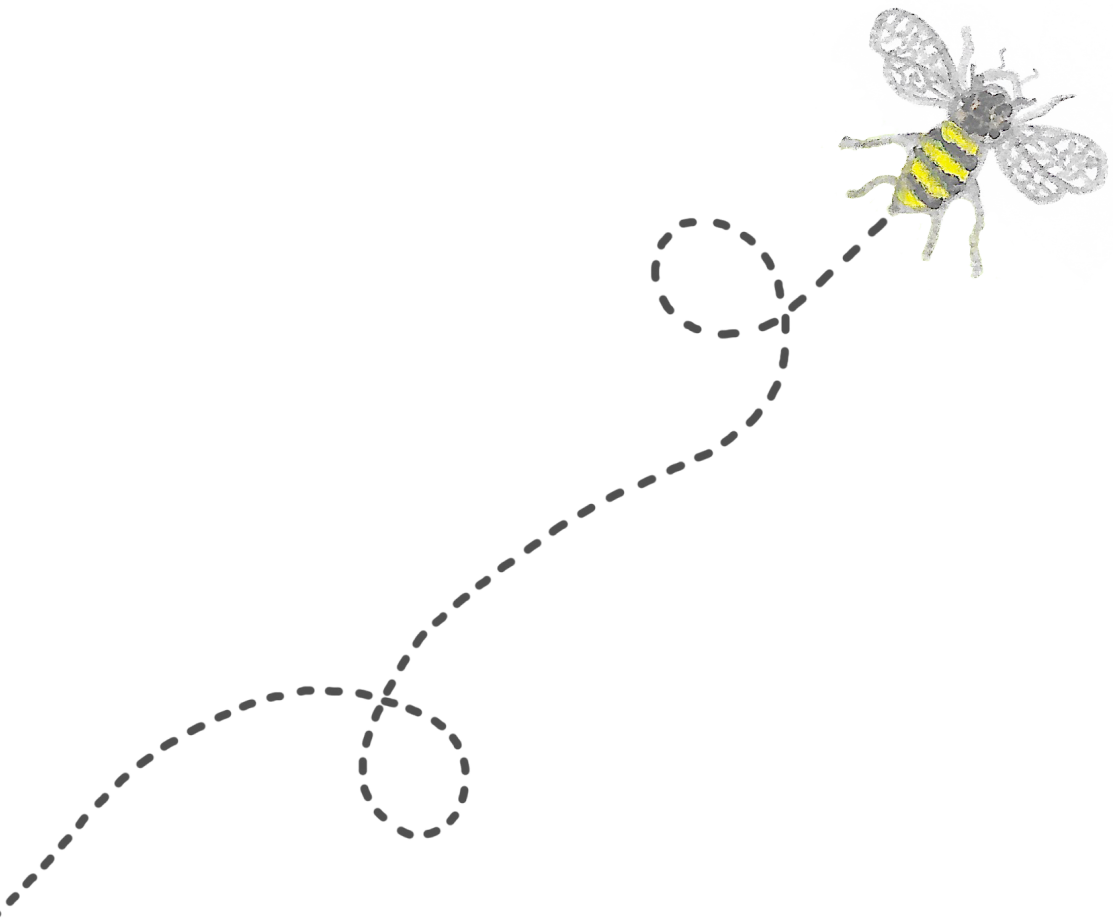
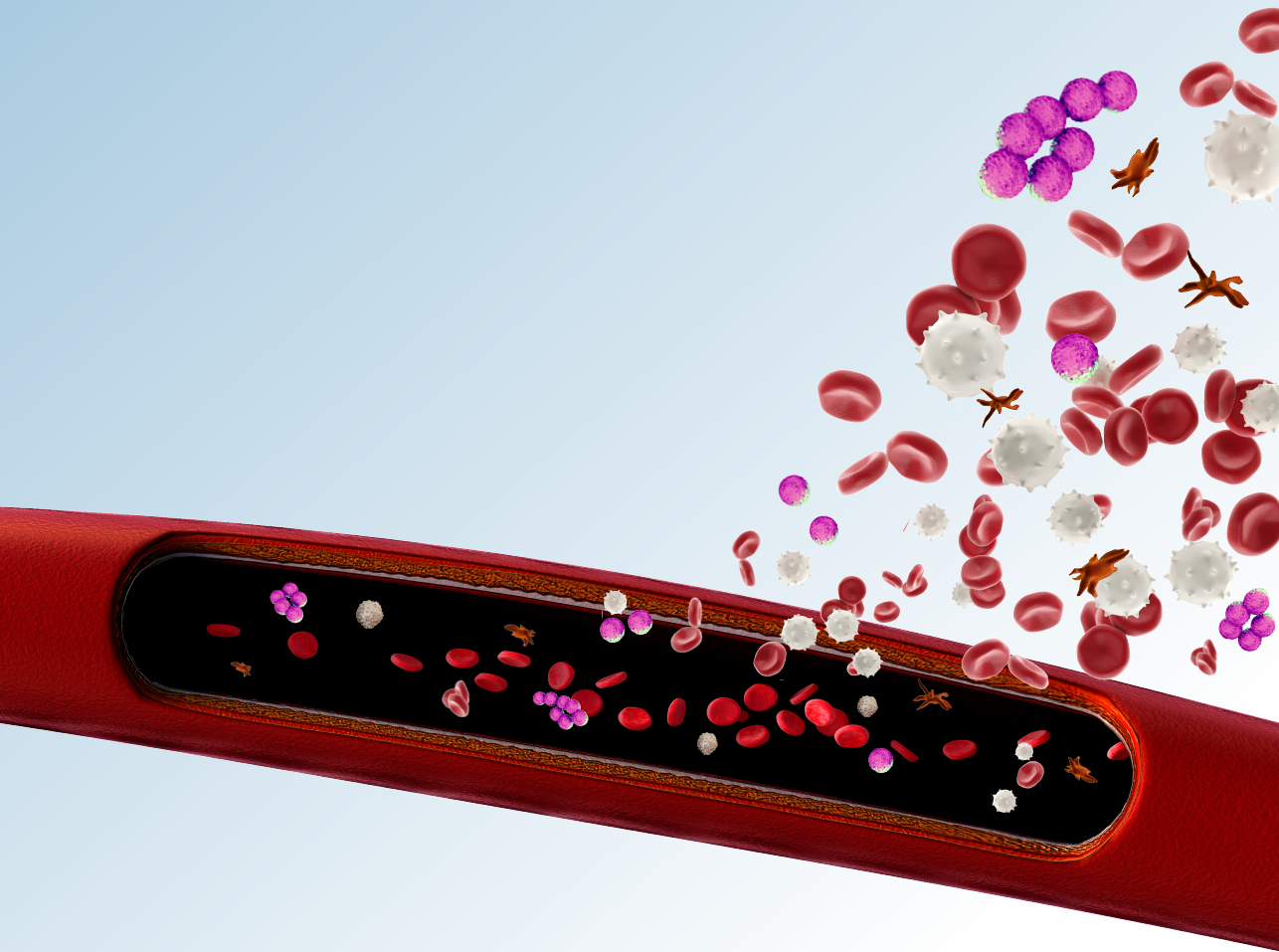


Figure S7: SSL5-binding to *g*-CD47 and *m*-CD47 glycan mutants. (A-D) Kinetic titration experiment using successive injections of stepwise increased SSL5-concentrations (3 nM - 6 μ M) to immobilized CD47-TEV-Fc wild type (black lines) and glycan single (A), double (B), triple (C) and quadruple (D) point mutants (colored lines). Responses (RU) of SSL5-binding were measured for regions containing glycosylated (*g*-)CD47 (solid lines) or mannosylated (*m*-)CD47 (dashed lines). Binding was measured to at least three independent ROIs per mutant, situated at different locations on the SPR-chip, of which one representative curve is shown here.



4

5



General discussion

Louris J. Feitsma

Crystal & Structural Chemistry, Bijvoet Center for Biomolecular Research, Department of Chemistry, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.

Blood vessel damage provokes adhesion and activation of circulating platelets to the site of tissue injury. One essential step in hemostasis is binding of platelet receptor glycoprotein VI (GPVI) to exposed collagens of the sub-endothelial matrix. Secretion of platelet granules from adhered and activated platelets trigger cross-linking to additional platelets and in turn stimulates their activation, ultimately leading to the clotting of platelets to a plug that seals off the damaged endothelial cell layer. At the same time, tissue damage reduces barriers for the transmission and progression of pathogens, including the principally commensal skin bacterium *Staphylococcus aureus*. Recognition of various pathogen-associated molecular patterns (PAMPs) by the family of Toll-like receptors (TLRs) is one of the primary steps in leukocyte activation, migration to the site of infection, and subsequent clearance of the pathogen. Both processes are of vital importance to arrest bleeding and to impede progression and pathogenicity of invading bacteria, respectively. Opposite to the latter process, bacteria such as *S. aureus* have developed intriguing mechanisms to evade detection and clearance by the host organism via the secretion of various toxins, such as the Staphylococcal superantigen-like proteins (SSLs), which interfere in diverse immune processes. These immune evasion mechanisms in combination with increased antibiotics resistance have successfully promoted the emergence of *S. aureus* as a pathogen. In this chapter, I discuss the novel findings that our structural and biochemical studies have provided for GPVI-mediated platelet activation, TLR2-mediated neutrophil activation, and SSL-mediated immune inhibition.

Platelet activation via interactions of GPVI and $\alpha_2\beta_1$ with collagen

Collagens of the extracellular matrix that become exposed upon vascular damage promote adhesion of circulating platelets to the site of lesion both directly via platelet receptors that can bind specific sequences in collagen, and indirectly via plasma glycoprotein Von Willebrand Factor (VWF), which has separate binding sites for collagen and platelet receptor GPIIb α . Initial platelet tethering and deceleration is mediated via GPIIb α of the GPI-IX complex and large VWF-multimers. Stable adhesion, activation and thrombus growth requires the collagen-binding contributions of integrin $\alpha_2\beta_1$ and glycoprotein VI (GPVI). $\alpha_2\beta_1$ is supposed to provide most of the adhesive strength, along with the VWF-GPI-IX-V complex, whereas GPVI appears primarily responsible for platelet activation through the Fc-receptor γ -chain to which it is constitutively associated in the platelet membrane. The separation between adhesive and signaling contributions under conditions of shear, however, may be less strong than estimated from static adhesion assays with monomeric receptors, and sufficient tethering may only be realized by simultaneous interactions of $\alpha_2\beta_1$ and GPVI with collagen¹.

Our crystal structures of the GPVI-collagen complex described in **chapter 2** of this thesis provide new insights into the collagen-binding mode of platelet-receptor GPVI. The structures reveal a collagen binding site situated across the D1-domain β -sheet, having the same location as the homologous site in Leukocyte-associated Immunoglobulin-like Receptor 1 (LAIR-1)². Our mutagenesis studies confirm the position of the primary collagen-binding site, and do not suggest a role for GPVI-residues located inside the D1-

domain groove between β -strands C' and E. Previous modeling studies with the structure of GPVI alone have therefore led to a misidentification of the primary collagen-binding site in the C'-E groove³. Subsequent studies in which interactions between potential anti-thrombotic reagents and GPVI have been modelled based on this previously suggested collagen-binding site⁴⁻⁸ might therefore be incorrect. One exception to this is a study on GPVI-antagonist losartan (DuP-753) that inhibits collagen-stimulated platelet aggregation⁹ and was shown by NMR-experiments to bind a region of GPVI that overlaps the primary collagen-binding site⁵. Their model of losartan-binding to GPVI, however, does not provide an obvious explanation for sterical inhibition of collagen-mediated platelet activation, possibly because of the use of geometrical boundaries in molecular dynamics experiments that are inconsistent with the location of the collagen-binding site. Unravelling of the primary site in GPVI and its collagen-binding mechanisms provides a major step forward for future structure-based design and improvement of anti-thrombotic drugs.

The development of drugs that specifically inhibit GPVI-mediated hemostasis may also require studies to minimize cross-reactivity with immune receptor LAIR-1. Our and previous work² provide the structural basis that support a common collagen-binding mechanism of GPVI and the functionally distinct receptor LAIR-1. Three residues in GPVI that have important interactions with collagen, W76, R38, and E40, are conserved between the GPVI and LAIR families. All residue positions in the collagen binding site of GPVI other than this triad are occupied by different amino acids in LAIR-orthologues, except for the human variants of LAIR-1 and -2, which appear more closely related to GPVI. The substantial weaker conservation within the LAIR-family is also reflected by the different collagen binding properties of human and mouse LAIR-1¹⁰, which constitutes a potential risk in the use of animal models for drug design in this situation. However, also a number of residues at the periphery of the binding site that potentially determine specificity for collagen sequences varies between human GPVI and human LAIR, and could be targeted for the development of GPVI- or LAIR-specific inhibitors.

In contrast to other collagen-binding proteins such as VWF-A3 and integrin $\alpha_2\beta_1$, the locations of GPVI-sites in collagen III could not be fully resolved using Toolkit-studies^{11,12}. The capacity of GPVI to bind Toolkit-III peptides¹¹ or GPO-peptides¹³ strongly correlates with the number of GPO-repeats present, but a number of anomalies measured in the set of Toolkit-III peptides suggest that also residues just outside a stretch of GPO-residues may affect GPVI-binding. The structures of the GPVI-collagen complexes demonstrate that GPVI requires a canonical collagen motif of 6-residues for binding: OGPOGP. This motif is found multiple times in triple-helical domains of collagen I and III, and suggest that GPVI-binding sites scatter throughout the entire fibril-forming domains. However, sites may not necessarily be composed by OGP-repeats only, as mutational analysis of Toolkit-peptide III-30 indicates that GPVI also binds the site constituted by the AGPOGP-sequence located centrally in that peptide. These findings demand for further studies addressing fine-specificity of collagen-motifs forming a GPVI-binding site.

Of all triple helical peptides containing a segment of the native collagen III sequence, peptide III-30 shows the highest capacity and affinity for GPVI-binding¹². The single OGPOGP-site present in the peptide overlaps the constant C-terminal region and does therefore not exist in the native collagen III-sequence. Although the binding to this peptide

in this assay is probably stronger than would be expected from its sequence in collagen III itself, the strong GPVI-binding to III-30 is intriguing, since two native AGPOGP-motifs in this peptide are flanked by a weak GAOGER- and a strong GMOGER-site for $\alpha_2\beta_1$ -binding in the native collagen III-sequence¹⁴. The two adjacent AGPOGP-sites within relative short distance (37 Å) on a single collagen triple helix can be occupied by two GPVI-molecules possibly as preformed platelet-associated dimer. The presence of four consecutive sites for platelet receptor binding, however, would support a model in which strong platelet attachment under *in vivo* conditions of shear is established by accumulative contributions of four receptors that directly bind to collagen.

The *in vivo* occurrence of collagen-binding by GPVI, $\alpha_2\beta_1$, and VWF strongly depends on the accessibility of receptor-sites at the surface of fibrillar collagen. Fiber diffraction studies have previously unraveled the molecular packing of collagen I helices in a fibril^{15,16}, but divergent opinions exist about the structure of the fibrillar surface. Models that describe defects in the fibrillar structure and that propose the presence at the fibril surface of only *N*- or *C*-terminal regions would suggest that – if the collagen-model is translated to collagen type III – the GPVI binding sites highlighted by peptide III-30 and the conserved $\alpha_2\beta_1$ -sites are inaccessible. Moreover, both ‘terminal surface models’ do not account for the presence of only VWF-binding site, which is predicted to be inaccessible in those models (see review¹⁷). At present, none of the existing surface models is sufficiently supported by experimental data that provide insights into the feasibility of receptor-binding, or these models must at least assume relatively fluidic surface conformations at sites of vascular damage to support for platelet binding.

Immune activation and modulation of Toll-like receptor 2

Skin injury is a sizable risk for the transmission and progression of pathogenic bacteria that demands for fast immune responses to preclude amplification of infections. Recognition of various patterns in pathogen-associated molecules is an important first step in activation of the innate immune system that ultimately leads to migration of leukocytes to the site of infection and subsequent clearance of the invading pathogen. The family of Toll-like receptors (TLRs) recognize various conserved patterns¹⁸ including lipoproteins and -peptides (TLR2 in combination with TLR1 or TLR6), lipopolysaccharide (TLR4), bacterial flagellin (TLR5), as well as non-self ssRNA (TLR7-8), dsRNA (TLR3), and dsDNA (TLR9). Immune activation induced by binding of bacterial lipopeptides to TLR2 is of vital importance during bacterial infections, which is notably reflected by the increased susceptibility of TLR2-deficient mice to septicemia due to *S. aureus*¹⁹, meningitis due to *S. pneumoniae*²⁰, and infection with *M. tuberculosis*²¹.

Crystal structures of the synthetic tripalmitoylated lipopeptide Pam₃CSK₄ bound to the TLR2-TLR1 complex and the dipalmitoylated Pam₂CSK₄ bound to the TLR2-TLR6 complex have revealed the molecular basis for selective TLR-activation by bacterial lipopeptides and -proteins^{22,23}. Other putative bacterial TLR2-agonists have primarily been identified after isolation from bacterial extracts, and include molecular compounds that are structurally very distinct from these cysteine-linked lipids, such as lipopolysaccharides, lipoteichoic acids, lipomannans, lipoarabinomannans, glycosylphosphatidylinositol,

zymosan and peptidoglycan. It is highly questionable for sterical reasons whether TLR2-TLR1 or TLR2-TLR6 complexes can associate with lipidic components that have more than three acyl chains (two chains binding to TLR2 and one to TLR1) or that have lipid head groups much larger than mercaptoglycerol. In view of the picomolar lipoprotein- or lipopeptide-concentrations needed for TLR2-sensing, however, all other putative TLR2-activating reagents were most likely contaminated by traces of natural lipoproteins or lipopeptides²⁴ and do not activate TLR2 themselves.

The crystal structures of the TLR2-TLR1-Pam₃CSK₄ and TLR2-TLR6-Pam₂CSK₄ complexes represent two activated states of TLR2 with a lipopeptide ligand bound and heterodimerized with TLR1 or TLR6^{22,23}. Both structures show a closed conformation of the lipopeptide pocket in TLR2, in which the lipid head groups and tails are buried inside the TLR-complex and only the peptides regions protrude into the solvent area. These closed conformations of the lipopeptide pocket in TLR2 are required to constitute the correct interfaces involved in complex formation with TLR1 or TLR6, but, unlike the more open conformation observed in other monomeric TLR2-ligand-structures²³, they are not likely facilitating lipopeptide transfer. These data would therefore support a two-step concerted mechanism in which binding of a lipopeptide/-protein is followed by a change in the TLR2-conformation that is needed for constitution of the specific complex with TLR6 or TLR1.

The exact sequential order of ligand binding and heterodimerization in an *in vivo* activation mechanism, however, cannot be determined with certainty from the existing structures alone, since structural information on the monomeric TLR2-Pam₂CSK₄ complex and the heterodimers without bound ligand is lacking. A previous study using fluorescence resonance energy transfer showed that TLR2 forms heterotypic associations with TLR1 and TLR6 in the membrane of human monocytes, both in the presence and absence of stimulating bacterial ligands²⁵. This would lead to an extended model in which TLR2 is co-localized with either TLR6 or TLR1 on resting cells, and forms a pre-existing dimer that modulates selectivity for di- or triacylated lipoproteins, respectively²². Structures of such non-activated heterodimers and the extent of interactions between the extracellular ligand-binding domains are not known, but must at least allow for ligand binding and therefore have to display substantial differences with respect to the activated complexes.

The crystal structure of the TLR2-SSL3-POPC complex presented in **chapter 3** contains the first proteinaceous inhibitor of TLR2: staphylococcal superantigen-like protein 3 (SSL3). SSL3 binds TLR2 near its lipopeptide pocket in the open pocket conformation, which obstructs TLR2-activation in two ways. First, SSL3-binding sufficiently reduces the size of the TLR2-pocket opening to inhibit transfer of lipopeptides to TLR2. Second, it involves a site on TLR2 that sufficiently overlaps regions responsible for TLR1/6-binding to preclude the formation of heterodimers with either TLR1 or TLR6. Yet, it is not known whether the low-nanomolar affinity and inhibitory capacity of SSL3 for TLR2 is also effective for disruption of a pre-existing or fully activated heterodimer, since affinity constants of both dimer types are lacking. One important aspect of SSL3-TLR2 binding that allows for SSL3-mediated inhibition at both stages of TLR2-activation is the approximately half-sized opening of the lipopeptide pocket after SSL3 has bound. Association of SSL3 with TLR2 is – because of the small remaining opening – not inhibited by the presence of a pre-bound ligand in the TLR2-pocket; subsequent formation a ternary SSL3-TLR2-ligand complex

is, if necessary for SSL3-binding, conducted by a conformational change of the ligand. Combined, SSL3 can inhibit TLR2 before and after a lipopeptide has bound, depending on the activation state of TLR2.

The formation of a ternary SSL3-TLR2-ligand complex is feasible and generally indiscriminative for the type and activating capacity of the ligand, but strongly depends on the order in which TLR2 associates with the ligand and SSL3: the ligand must bind first. The structure of the SSL3-TLR2 complex itself reveals the presence of a copurified phosphatidylcholine lipid (PC); an abundantly occurring membrane-component that must be recognized by TLR2 as self, incapable of inducing TLR2-activation, and sufficiently weakly bound not to inhibit binding of other activating ligands. Possibly, the lipopeptide pocket of TLR2 is permanently occupied by a diacyl-lipid such as PC, even in an inactive state. Also the presence of glycolipids is sterically feasible in the SSL3-TLR2 complex if it adopts a conformation similar to the previously observed conformation of LTA in TLR2 alone²³. We furthermore demonstrated that also the activating lipopeptide Pam₂CSK₄ can be engaged within a SSL3-TLR2 complex, but only if it is added prior to SSL3. Reversed addition, SSL3 and then Pam₂CSK₄, leads to strong inhibition of binding of the latter. Binding of SSL3 to TLR2 is apparently stable to such an extent, that release of SSL3 and transfer of ligands to the TLR2-pocket is kinetically highly unfavorable. Starting from the most plausible scenario of TLR2-activation – lipopeptide-binding precedes heterodimerization with TLR1/6 – the mode of SSL3-binding as observed in our structure then facilitates TLR2-inhibition in a dual mechanism: it prevents lipopeptide-binding or blocks dimerization, if lipopeptide-binding is already accomplished (Figure 1).

The strong capacity of SSL3 to inhibit TLR2-activation makes it an interesting starting point for the development of drugs against diseases that are linked to aberrant TLR2-activation, such as acute and chronic inflammation and auto-immune diseases²⁶. The design of small peptides that mimic the region of SSL3 involved in TLR2-binding may however be complicated by the extensive hydrophobicity of this region, and may necessitate the introduction of highly soluble groups in target compounds to counter problems with solubility. Furthermore, this region of SSL3 is composed of four OB-domain loops that show substantial flexibility in the unbound structure of SSL3, and the design of functional peptides requires approaches that use scaffolds coupled to a number of cyclic peptides, preferably an equal number of peptides as SSL3-loops involved in TLR2-binding. At present, bicyclic peptides have been synthesized that include the sequences of the β 2- β 3 loop and part of the β 1- β 2 loop of SSL3²⁷. Some of these peptides were shown to bind TLR2 and compete with SSL3-binding, and therefore create a promising starting point for further development of TLR2-targets. Functional inhibition of TLR2 has not yet been achieved using these peptides and may require longer or additional loop sequences and higher affinity for TLR2. Further studies may focus on the peptide structures that mimic longer fragments of the β 1- β 2 loop of SSL3, since especially this loops covers the pocket opening in TLR2 and binds to a TLR2-region strongly overlapping TLR1/6 sites.

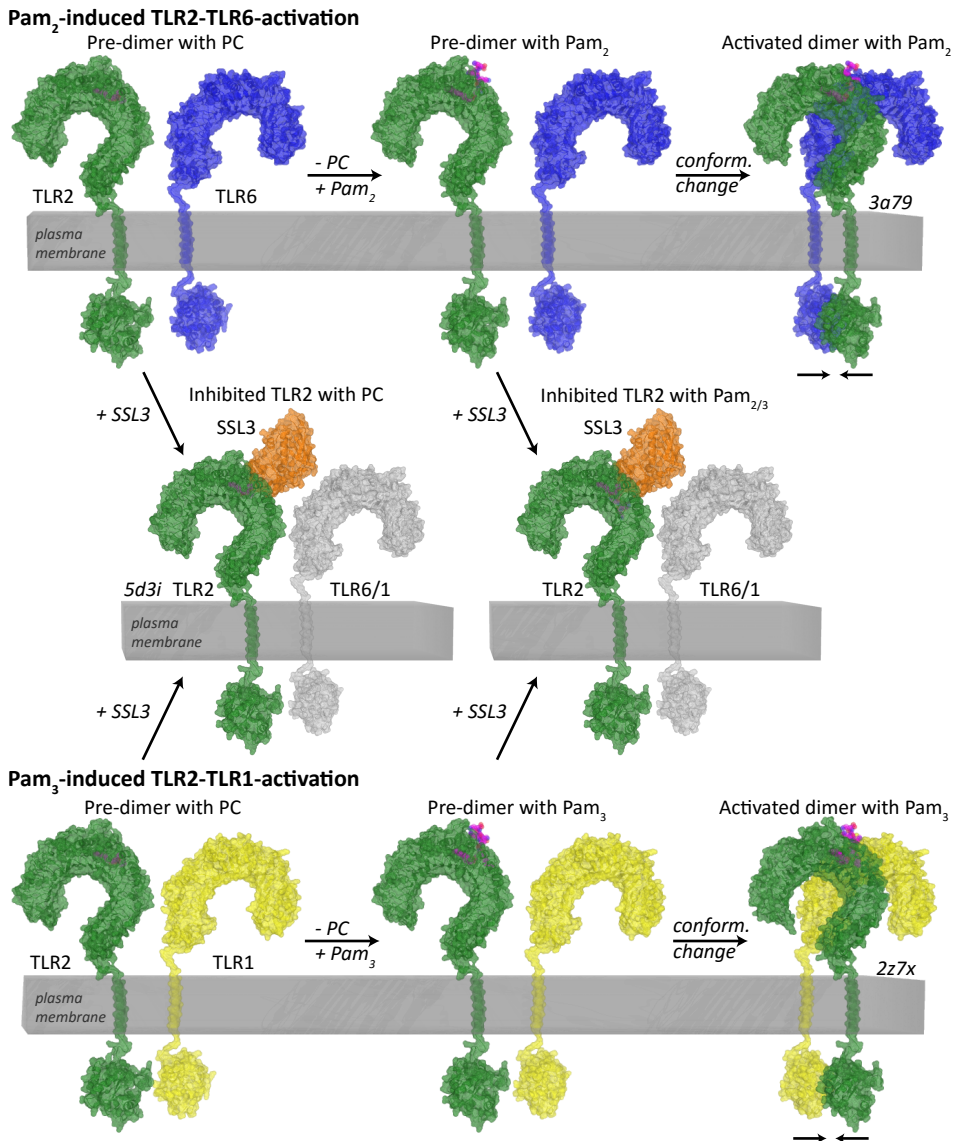


Figure 1: Proposed pathways of TLR2-activation and inhibition. Schematic overview of probable TLR2-activation pathways via Pam₂-selective TLR2-TLR6 heterodimers (left) and via Pam₃-selective TLR2-TLR1 heterodimers (right). TLR2-activation is achieved from membrane-associated pre-existing TLR2-TLR6 or TLR2-TLR1 dimers in two steps: exchange of non-activating PC-lipid to allow binding of the bacterial lipopeptide, followed by closing of the lipopeptide pocket in TLR2 that provides the correct interfaces for formation of the activated complexes with TLR6 or TLR1. SSL3 can inhibit at both stages of TLR2-activation, before and after the lipopeptide has bound, by formation of the SSL3-TLR2-PC and SSL3-TLR2-Pam_{2/3} complexes, respectively. TLR2 in these complexes with SSL3 cannot be further activated through lipopeptide binding and/or formation of the activated heterodimers. This overview is based on the ectodomain structures of the activated TLR2-TLR1-Pam₃ (PDB-ID: 2Z7X²²) and TLR2-TLR6-Pam₂ (PDB-ID: 3A79²³) complexes, as well as the SSL3-TLR2-PC inhibition complex presented in **chapter 3** (PDB-ID: 5D3I²⁸). PDB-codes are also indicated in the overview at the designated stages of activation and inhibition.

Immune evasion mechanisms of *Staphylococcus aureus*

The *Staphylococcus aureus* bacterium is predominantly a commensal skin bacterium, being continuously colonized in approximately 30% of the human population²⁹. In infectious state, *S. aureus* frequently causes mild skin and soft-tissue infections. Host invasion and influx into the bloodstream can also provoke *S. aureus* infections to progress into more serious diseases, such as endocarditis, sepsis, pneumonia, osteomyelitis, septic arthritis, and abscesses in organ tissues³⁰. Skin, soft-tissue and bloodstream infections related to *S. aureus* appear recurrent in 8-33% of the cases³¹, which emphasizes the marginal development of host protection against novel *S. aureus* infections³⁰. Key element in the progression of *S. aureus* and recurrence of *S. aureus*-related diseases is its ability to modulate both innate and adaptive immune responses of the host organism³²; processes that are of vital importance for the host organism to prevent against pathogen progression including *S. aureus*. Absence of functional vaccines in combination with exhaustive resistance against the currently available antibiotics elicited highly virulent *S. aureus* strains, such as various MRSA-strains³³. A growing number of *S. aureus* infections is difficult to treat using modern clinical techniques and demands for the development of novel therapies and strategies.

The host-microbe interaction between *S. aureus* and the infected organisms has been exposed to millions of years of ongoing mutual evolution, and have generated precise immune mechanisms for the detection and clearance of the bacterium on the one hand, and mechanisms to modulate immune activation on the other hand. The pathogenic emergence of *S. aureus* is primarily mediated through the secretion of virulence factors that interfere in immune processes of the host organism. The *S. aureus* secretome is estimated to consist of ~200 to 600 proteins³⁴, of which various proteins are currently known to interfere in the immune system³⁵. The design of drug cocktails that target the virulent *S. aureus* secretome could provide an elegant approach to interfere in the natural host-microbe interaction, thereby reducing bacterial toxicity and increasing the efficiency of immune responses by the host. Substantial progress in the design of such anti-toxins, however, must be preceded by a detailed molecular understanding of immune evasion mechanisms used by *S. aureus*, and the interaction must be significantly selective and specific to reduce potential risks of cross-reactivity.

Immune blocking and modulating activity is especially exhibited by the family of secreted staphylococcal superantigen-like proteins (SSLs), consisting of 14 homologous proteins of which 9 have been described to target one or multiple components in the immune system³⁵. One SSL has been reported to target platelet receptors involved in hemostasis³⁶, and suggest that currently unidentified SSL-targets may also exist outside immunity. SSL-targets have been detected via various methods, including neutrophil-stimulation³⁷, antibody-competition³⁸⁻⁴⁰, phage display⁴¹, pull-down analysis⁴²⁻⁴⁸, platelet aggregometry³⁶, matrix metalloproteinase activation⁴⁹, and surface plasmon resonance⁵⁰. However, not all studies have included a convincing post-identification analysis to substantiate – as far as the identification method did not provide this information – specificity of the SSL-target interaction, binding affinity, and ability to modulate activation or inhibition. It therefore remains to be seen in numerous cases if the described host-

microbe interactions are specific for the indicated SSL(s) and targets, or in other words, whether or not these SSLs have a unique site for interaction with the target receptor that is not present in (most of) the other SSL-members.

Molecular details of three SSL-target interactions have been unraveled in crystallography studies thus far: SSL7-ImmunoglobulinA (IgA)⁵¹, SSL7-Complement C5⁵², and SSL3-TLR2 presented in **chapter 3**. These studies revealed discrete binding sites for TLR2 and IgA in the SSL OB-domain, and for C5 in the SSL β -grasp domain. The highly distinct subdomain locations (Figure 2a) and amino acid compositions of sites (Figure 2b) that are generally not present in other SSLs, dictate that sites and immune modulating functions must have evolved separately. In view of the divergent evolution of SSL-molecules and their functions, the existence of another SSL, SSL4, capable of inhibiting TLR2 via an only modestly different site at the same location in the OB-domain is interesting, but not well-explained^{28,38}. Relatively high sequence identity between the OB- and β -grasp-domains of SSL3 and SSL4 sequences comprising the with respect to the other SSLs indicates a late branching and suggest that development of TLR2 inhibitory capacity has developed prior to gene duplication.

A shared element of the SSL-subset comprising SSL2-6 and SSL11, is the capacity to associate with sialyl Lewis^x (sLe^x)-containing glycans; complex glycan structures that are profusely present on the immune cell surface⁵³. The molecular details of interaction with the tetrasaccharide sLe^x are well-described for SSL4⁵⁴, SSL5⁵⁵, and SSL11⁵⁶. Immune inhibition by any of these six lectin SSLs has been shown to rely substantially on sLe^x-binding. Only inhibition of TLR2 by SSL3²⁸ and matrix metalloproteinase by SSL5⁴⁹ are exceptions to this,

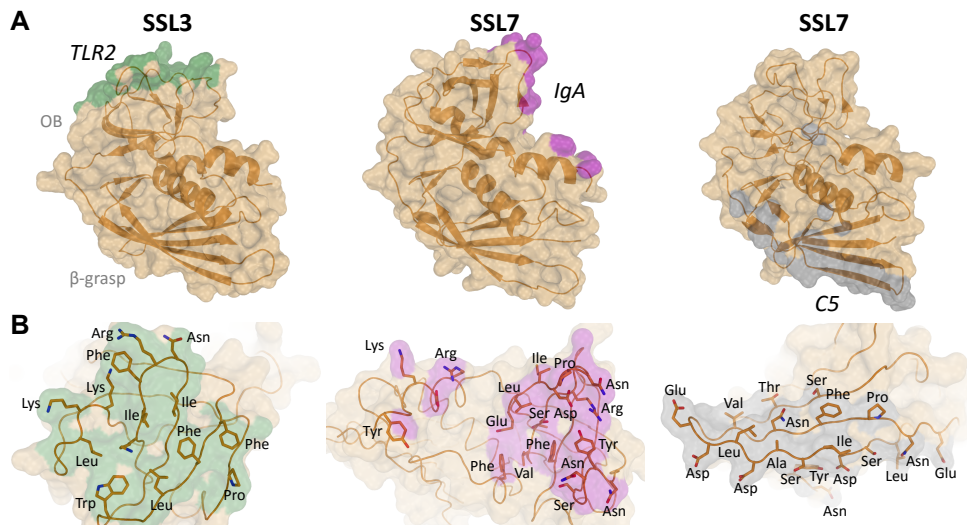


Figure 2: Location and nature of target-binding sites in SSLs (A) Structure of the Staphylococcal Superantigen-Like proteins as found in the SSL3-TLR2 complex (PDB-ID: 5D31²⁸; left panel), the SSL7-IgA complex (PDB-ID: 2QEJ⁵¹; middle panel), and the SSL7-complement C5 complex (PDB-ID: 3KLS⁵²; right panel) indicating the different locations of the binding sites for TLR2 (green), IgA (purple), and complement C5 (gray), respectively. **(B)** Detailed views of the TLR2-binding site of SSL3 (left panel), the IgA-binding site of SSL7 (middle panel), and the complement C5-binding site of SSL7 (right panel) showing the differences in shape and composition between SSL-sites.

and do not depend on interaction with target-attached glycans. Inhibition of all its other SSL5-targets, which are P-selectin glycoprotein ligand-1 (PSGL-1)³⁹, various chemokine and anaphylatoxin receptors³⁷, and platelet receptors^{36,50}, is significantly reduced upon neuraminidase treatment of these proteins removing terminal sialic acid from glycan moieties. In view of the small size of SSL5, about 25 kDa, it is highly unlikely that it has developed specific and separate binding sites for each of the putative targets. It has been suggested that SSL5 binds subgroups of targets via a common mechanism, since these targets share structural properties, such as the presence of sulfotyrosines in both PSGL-1 and GPIb α ³⁶, or the presence of a sialomucin stalk-domain in both GPIb α and GPVI⁵⁰. Evidence that SSL5-binding is facilitated by these groups is however lacking. Therefore, specific SSL-targets may have been misidentified if inhibitory effects or binding is solely a consequence of generic sLe^x-binding mechanisms.

The method presented in **chapter 4** provides an experimental setup that distinguishes between glycan-independent, glycan-enhanced, and glycan-only binding, and could assist in authentication of claimed SSL-target combinations. Our analysis of three SSL-receptor combinations showed that binding of SSL5 to GPVI is exclusively mediated by interaction with *N*- and *O*-linked glycans in GPVI, and a similar interaction could be established using SSL6 instead of SSL5. These results on the one hand indicate the presence on GPVI of *N*- and *O*-linked glycan structures, likely sLe^x or sLacNac, on GPVI to which both SSLs bind, on the other hand they reveal that GPVI is not targeted specifically by SSL5. GPVI appears therefore as a target of lectin-SSLs in general. Although it cannot be excluded that GPVI-binding by SSL-proteins induces GPVI-mediated platelet activation, this has in fact never been demonstrated. The other SSL-receptor pairs investigated, SSL3-TLR2 and SSL6-CD47, show a protein-protein component to the binding, which confirm published claims of specific receptor recognition. Binding of CD47 by SSL6 needs to be studied further to unravel the molecular basis of the interaction. In general, exclusion of generic lectin-glycan interactions and understanding the molecular details of specific SSL-receptor binding generates opportunities for the development of novel drugs targeting SSL virulence factors, in order to limit *S. aureus* virulence.

Conclusion

The crystal structures of GPVI-collagen and SSL3-TLR2 complexes have brought fundamental understanding of the molecular mechanisms of two important hemostatic and immune evasion processes a major step forward. Structural insights in GPVI-collagen binding improved our concepts of GPVI-mediated platelet activation and lead to a revision of existing assumptions on the location of the collagen-binding site in GPVI and the molecular details of the GPVI-collagen interaction. The GPVI-collagen complexes have also provided a molecular basis for understanding of the interaction of GPVI with (hydroxy) proline stretches in collagen, but our data cannot dictate a strict interpretation of the location of GPVI-binding sites in collagen III under physiological conditions. A number of conceptual challenges therefore remain. Long-standing questions about accessibility and locations of GPVI-binding sites in fibrillar collagens are not yet answered, and necessitate additional progress in the understanding of fibrillar structures, (degenerate) sequence motifs required for GPVI-binding, and the roles of platelet-associated GPVI-dimers.

The SSL3-TLR2 complex shows for the first time how activation of a Toll-like receptor is regulated by a bacterial virulence factor, and provides the structural basis for SSL3-mediated inhibition of immune activation via TLR2. This unraveled the mechanisms used by SSL3 to bind near the lipopeptide pocket of TLR2, to inhibit lipopeptide sensing by TLR2, and to prevent association of TLR2 with either TLR6 or TLR1. These data also reflect on the complex strategies used by *S. aureus* to evade immune activation of the host towards the bacterium; mechanisms primarily postulated on the basis of immune inhibition, but in numerous cases imprecisely understood on a molecular level. Characterization of SSL-glycan contributions to proposed host-inhibition interactions assists in authentication of target-specific inhibitory processes used by *S. aureus* to evade immunity of the host organism, filtering generic glycan-binding mechanisms used by SSLs for mobilization to the immune cell surface. Our novel biochemical method using distinct receptor glycoforms showed that GPVI is not a specific SSL5-target and specified contributions of interactions with target *N*-glycans for CD47- and TLR2-binding by SSL6 and SSL3, respectively. Combined, these structural and biochemical studies improved and conceivably revised concepts of vital processes in hemostasis, immune activation and evasion.

References

1. Farndale RW, Sixma JJ, Barnes MJ, de Groot PG. The role of collagen in thrombosis and hemostasis. *J. Thromb. Haemost.* 2004;2(4):561–73.
2. Brondijk THC, de Ruiter T, Ballering J, et al. Crystal structure and collagen-binding site of immune inhibitory receptor LAIR-1: unexpected implications for collagen binding by platelet receptor GPVI. *Blood.* 2010;115(7):1364–73.
3. Horii K, Kahn ML, Herr AB. Structural basis for platelet collagen responses by the immune-type receptor glycoprotein VI. *Blood.* 2006;108(3):936–42.
4. Kato-Takagaki K, Mizukoshi Y, Yoshizawa Y, et al. Structural and Interaction Analysis of Glycoprotein VI-binding Peptide Selected from a Phage Display Library. *J. Biol. Chem.* 2009;284(16):10720–10727.
5. Ono K, Ueda H, Yoshizawa Y, et al. Structural Basis for Platelet Antiaggregation by Angiotensin II Type 1 Receptor Antagonist Losartan (DuP-753) via Glycoprotein VI. *J. Med. Chem.* 2010;53(5):2087–2093.
6. Taylor L, Vasudevan SR, Jones CI, et al. Discovery of Novel GPVI Receptor Antagonists by Structure-Based Repurposing. *PLoS One.* 2014;9(6):e101209.
7. Liu W, Liu G, Zhou H, et al. Computer prediction of paratope on antithrombotic antibody 10B12 and epitope on platelet glycoprotein VI via molecular dynamics simulation. *Biomed. Eng. Online.* 2016;15(S2):152.
8. Bhunia SS, Misra A, Khan IA, et al. Novel Glycoprotein VI Antagonists as Antithrombotics: Synthesis, Biological Evaluation, and Molecular Modeling Studies on 2,3-Disubstituted Tetrahydropyrido(3,4- b) indoles. *J. Med. Chem.* 2017;60(1):322–337.
9. Grothusen C, Umbreen S, Konrad I, et al. EXP3179 Inhibits Collagen-Dependent Platelet Activation via Glycoprotein Receptor-VI Independent of AT1-Receptor Antagonism: Potential Impact on Atherothrombosis. *Arterioscler. Thromb. Vasc. Biol.* 2007;27(5):1184–1190.
10. Lebbink RJ, Raynal N, de Ruiter T, et al. Identification of multiple potent binding sites for human leukocyte associated Ig-like receptor LAIR on collagens II and III. *Matrix Biol.* 2009;28(4):202–10.
11. Jarvis GE, Raynal N, Langford JP, et al. Identification of a major GpVI-binding locus in human type III collagen. *Blood.* 2008;111(10):4986–4996.
12. Jung SM, Moroi M, Soejima K, et al. Constitutive dimerization of glycoprotein VI (GPVI) in resting platelets is essential for binding to collagen and activation in flowing blood. *J. Biol. Chem.* 2012;287(35):30000–30013.

13. Smethurst PA, Onley DJ, Jarvis GE, et al. Structural basis for the platelet-collagen interaction: the smallest motif within collagen that recognizes and activates platelet Glycoprotein VI contains two glycine-proline-hydroxyproline triplets. *J. Biol. Chem.* 2007;282(2):1296–304.
14. Raynal N, Hamaia SW, Siljander PR-M, et al. Use of synthetic peptides to locate novel integrin alpha-2beta1-binding motifs in human collagen III. *J. Biol. Chem.* 2006;281(7):3821–31.
15. Orgel JPRO, Irving TC, Miller A, Wess TJ. Microfibrillar structure of type I collagen in situ. *Proc. Natl. Acad. Sci. U. S. A.* 2006;103(24):9001.
16. Perumal S, Antipova O, Orgel JPRO. Collagen fibril architecture, domain organization, and triple-helical conformation govern its proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* 2008;105(8):2824–9.
17. Herr AB, Farndale RW. Structural insights into the interactions between platelet receptors and fibrillar collagen. *J. Biol. Chem.* 2009;284(30):19781–5.
18. Gay NJ, Symmons MF, Gangloff M, Bryant CE. Assembly and localization of Toll-like receptor signalling complexes. *Nat. Rev. Immunol.* 2014;14(8):546–558.
19. Takeuchi O, Hoshino K, Akira S. Cutting edge: TLR2-deficient and MyD88-deficient mice are highly susceptible to *Staphylococcus aureus* infection. *J. Immunol.* 2000;165(10):5392–5396.
20. Echchannaoui H, Frei K, Schnell C, et al. Toll-Like Receptor 2–Deficient Mice Are Highly Susceptible to *Streptococcus pneumoniae* Meningitis because of Reduced Bacterial Clearing and Enhanced Inflammation. *J. Infect. Dis.* 2002;186(6):798–806.
21. Sugawara I, Yamada H, Li C, et al. Mycobacterial infection in TLR2 and TLR6 knockout mice. *Microbiol. Immunol.* 2003;47(5):327–36.
22. Jin MS, Kim SE, Heo JY, et al. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell.* 2007;130(6):1071–82.
23. Kang JY, Nan X, Jin MS, et al. Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity.* 2009;31(6):873–84.
24. Zähringer U, Lindner B, Inamura S, Heine H, Alexander C. TLR2 - promiscuous or specific? A critical re-evaluation of a receptor expressing apparent broad specificity. *Immunobiology.* 2008;213(3–4):205–24.
25. Triantafilou M, Gamper FGJ, Haston RM, et al. Membrane sorting of toll-like receptor (TLR)-2/6 and TLR2/1 heterodimers at the cell surface determines heterotypic associations with CD36 and intracellular targeting. *J. Biol. Chem.* 2006;281(41):31002–31011.
26. Liu Y, Yin H, Zhao M, Lu Q. TLR2 and TLR4 in Autoimmune Diseases: a Comprehensive Review. *Clin. Rev. Allergy Immunol.* 2014;47(2):136–147.
27. Koymans KJ. SSLs in *Staphylococcal* Immune Evasion. Utrecht University (Thesis); 2015.
28. Koymans KJ, Feitsma LJ, Brondijk THC, et al. Structural basis for inhibition of TLR2 by *staphylococcal* superantigen-like protein 3 (SSL3). *Proc. Natl. Acad. Sci.* 2015;112:11018–11023.
29. van Belkum A, Melles DC, Nouwen J, et al. Co-evolutionary aspects of human colonisation and infection by *Staphylococcus aureus*. *Infect. Genet. Evol.* 2009;9(1):32–47.
30. David MZ, Daum RS. Community-associated methicillin-resistant *Staphylococcus aureus*: Epidemiology and clinical consequences of an emerging epidemic. *Clin. Microbiol. Rev.* 2010;23(3):616–687.
31. Kallen AJ. Health Care–Associated Invasive MRSA Infections, 2005–2008. *JAMA.* 2010;304(6):641.
32. Thammavongsa V, Kim HK, Missiakas D, Schneewind O. *Staphylococcal* manipulation of host immune responses. *Nat. Rev. Microbiol.* 2015;13(9):529–43.
33. Liu GY. Molecular pathogenesis of *Staphylococcus aureus* infection. *Pediatr. Res.* 2009;65(5 Pt 2):71R–77R.
34. Economou A. Bacterial secretome: the assembly manual and operating instructions (Review). *Mol. Membr. Biol.* 2002;19(3):159–169.
35. Koymans KJ, Vrieling M, Gorham RD, van Strijp JAG. *Staphylococcal* Immune Evasion Proteins: Structure, Function, and Host Adaptation. *Life Sci. J.* 2015;6(4):23–27.
36. de Haas CJC, Weeterings C, Vughs MM, et al. *Staphylococcal* superantigen-like 5 activates platelets and supports platelet adhesion under flow conditions, which involves glycoprotein Ibalpha and alpha IIb beta 3. *J. Thromb. Haemost.* 2009;7(11):1867–74.

37. Bestebroer J, Van Kessel KPM, Azouagh H, et al. Staphylococcal SSL5 inhibits leukocyte activation by chemokines and anaphylatoxins. *Blood*. 2009;113(2):328–337.
38. Baradoel BW, Vos R, Bouman T, et al. Evasion of Toll-like receptor 2 activation by staphylococcal superantigen-like protein 3. *J. Mol. Med.* 2012;90(II):1109–1120.
39. Bestebroer J, Poppelier MJG, Ulfman LH, et al. Staphylococcal superantigen-like 5 binds PSGL-1 and inhibits P-selectin-mediated neutrophil rolling. *Blood*. 2007;109(7):2936–43.
40. Walenkamp AME, Boer IGJ, Bestebroer J, et al. Staphylococcal superantigen-like 10 inhibits CX-CL12-induced human tumor cell migration. *Neoplasia*. 2009;11(4):333–44.
41. Fevre C, Bestebroer J, Mebius MM, et al. Staphylococcus aureus proteins SSL6 and SEIX interact with neutrophil receptors as identified using secretome phage display. *Cell. Microbiol.* 2014;16(11):1646–1665.
42. Itoh S, Yamaoka N, Kamoshida G, et al. Staphylococcal superantigen-like protein 8 (SSL8) binds to tenascin C and inhibits tenascin C–fibronectin interaction and cell motility of keratinocytes. *Biochem. Biophys. Res. Commun.* 2013;433(1):127–132.
43. Itoh S, Yokoyama R, Kamoshida G, et al. Staphylococcal superantigen-like protein 10 (SSL10) inhibits blood coagulation by binding to prothrombin and factor Xa via their γ -carboxyglutamic acid (Gla) domain. *J. Biol. Chem.* 2013;288(30):21569–80.
44. Langley R, Wines B, Willoughby N, et al. The staphylococcal superantigen-like protein 7 binds IgA and complement C5 and inhibits IgA-Fc alpha RI binding and serum killing of bacteria. *J. Immunol.* 2005;174(5):2926–33.
45. Itoh S, Hamada E, Kamoshida G, et al. Staphylococcal superantigen-like protein 10 (SSL10) binds to human immunoglobulin G (IgG) and inhibits complement activation via the classical pathway. *Mol. Immunol.* 2010;47(4):932–938.
46. Patel D, Wines BD, Langley RJ, Fraser JD. Specificity of staphylococcal superantigen-like protein 10 toward the human IgG1 Fc domain. *J. Immunol.* 2010;184(11):6283–92.
47. Itoh S, Hamada E, Kamoshida G, et al. Staphylococcal superantigen-like protein 5 inhibits matrix metalloproteinase 9 from human neutrophils. *Infect. Immun.* 2010;78(7):3298–3305.
48. Yokoyama R, Itoh S, Kamoshida G, et al. Staphylococcal superantigen-like protein 3 binds to the Toll-like receptor 2 extracellular domain and inhibits cytokine production induced by Staphylococcus aureus, cell wall component, or lipopeptides in murine macrophages. *Infect. Immun.* 2012;80(8):2816–25.
49. Koymans KJ, Bisschop A, Vughs MM, et al. Staphylococcal superantigen-like protein 1 and 5 (SSL1 & SSL5) limit neutrophil chemotaxis and migration through MMP-inhibition. *Int. J. Mol. Sci.* 2016;17(7):1–16.
50. Hu H, Armstrong PCJ, Khalil E, et al. GPVI and GPIb α mediate staphylococcal superantigen-like protein 5 (SSL5) induced platelet activation and direct toward glycans as potential inhibitors. *PLoS One*. 2011;6(4):e19190.
51. Ramsland PA, Willoughby N, Trist HM, et al. Structural basis for evasion of IgA immunity by Staphylococcus aureus revealed in the complex of SSL7 with Fc of human IgA1. *Proc. Natl. Acad. Sci. U. S. A.* 2007;104(38):15051–6.
52. Laursen NS, Gordon N, Hermans S, et al. Structural basis for inhibition of complement C5 by the SSL7 protein from Staphylococcus aureus. *Proc. Natl. Acad. Sci. U. S. A.* 2010;107(8):3681–6.
53. Munro JM, Lo SK, Corless C, et al. Expression of sialyl-Lewis X, an E-selectin ligand, in inflammation, immune processes, and lymphoid tissues. *Am. J. Pathol.* 1992;141(6):1397–1408.
54. Hermans SJ, Baker HM, Sequeira RP, et al. Structural and functional properties of staphylococcal superantigen-like protein 4. *Infect. Immun.* 2012;80(11):4004–4013.
55. Baker HM, Basu I, Chung MC, et al. Crystal structures of the staphylococcal toxin SSL5 in complex with sialyl Lewis X reveal a conserved binding site that shares common features with viral and bacterial sialic acid binding proteins. *J. Mol. Biol.* 2007;374(5):1298–308.
56. Chung MC, Wines BD, Baker H, et al. The crystal structure of staphylococcal superantigen-like protein 11 in complex with sialyl Lewis X reveals the mechanism for cell binding and immune inhibition. *Mol. Microbiol.* 2007;66(6):1342–55.

A

ddendum



Nederlandse samenvatting

Dankwoord

List of publications

Curriculum Vitae (NL)

Curriculum Vitae (EN)

Nederlandse samenvatting

Een intacte menselijke huid biedt eerstelijns bescherming om binnendringing van ziekteverwekkers te voorkomen en biedt tevens bescherming tegen beschadiging van organen, zoals het direct hieronder liggende bloedvatenstelsel. Zonder adequate maatregelen van het lichaam zelf, zou beschadiging van de huid en anders door bijvoorbeeld een snede met een scherp voorwerp leiden tot fors bloedverlies. Bovendien zou toegang van bacteriën zonder verdere beschermingsmechanismen in vrijwel alle gevallen leiden tot bacteriële infecties. Met name bacteriën die van nature door de mens meegedragen worden, zoals de Stafylokokken-bacterie (*Staphylococcus aureus*), vormen daarbij een groot risico. Zodra huidbeschadigingen ontstaan, treden dan ook direct processen in werking om beide situaties te voorkomen. Enerzijds worden in de bloedbaan circulerende bloedplaatjes ter hoogte van de beschadiging afgeremd en geactiveerd (hemostase), en wordt een bloedstolsel gevormd waarmee de opening wordt afgesloten. Anderzijds treden lokaal ontstekingsmechanismen (inflammatie) in werking, die ervoor zorgen dat witte bloedcellen aangetrokken worden tot de beschadiging. Bij detectie van micro-organismen wordt dan direct begonnen met het onschadelijk maken en opruimen van de binnendringer. Hier staat tegenover dat *Staphylococcus aureus* – de oorzaak van 50% van de bacteriële huidinfecties – een verscheidenheid aan mechanismen heeft ontwikkeld om activatie van het immuunsysteem te remmen ('immuunontwijking'). Daarmee ontwijkt deze bacterie detectie en vernietiging en kan hij zich ongestoord verder vermenigvuldigen in het lichaam. Een aantal cruciale activatieprocessen op het gebied van hemostase, inflammatie en immuunontwijking zijn recentelijk ontdekt en daarom ontbreken op dit moment veelal gedetailleerde mechanismen op moleculair niveau. Hoofdstuk 1 van dit proefschrift geeft een overzicht van recente ontwikkelingen binnen deze drie thema's. In de hoofdstukken 2 t/m 4 presenteer ik onderzoeksresultaten die nieuwe inzichten geven in respectievelijk bloedplaatjesactivatie via de Glycoproteïne VI-receptor, immuunontwijking door *S. aureus* via blokkering van de Toll-like Receptor 2 (TLR2) en via interactie met andere immuunreceptoren. Dit onderzoek draagt daarmee bij aan een beter fundamenteel begrip van de werking dit soort biologische processen. Daarnaast vormen de uitkomsten van dit onderzoek ook een uitgangspunt voor ontwerp en ontwikkeling van nieuwe remmers en antibiotica in de behandeling van o.a. trombose, auto-immuunziekten en infecties van meticillineresistente *S. aureus* (MRSA). Een samenvatting van de belangrijkste onderzoeksresultaten en de consequenties hiervan voor toekomstig medicijnontwerp worden besproken in hoofdstuk 5.

Inzichten in binding van bloedplaatjesreceptor glycoproteïne VI aan collageen

Onder een laag van epitheelcellen aan de binnenkant van het bloedvat bevindt zich de extracellulaire matrix (ECM), die voornamelijk uit collageenfibrillen bestaat. Collageenfibrillen zijn lange, gebundelde eiwitketens, die stevigheid en elasticiteit aan het bloedvat geven. Naast deze meer structurele rol, speelt de matrix ook een belangrijke rol in signalering van eventuele beschadigingen. Zodra de epitheelcellaag beschadigd raakt, komt de ECM aan het oppervlak en in contact met bloedplasma. Allereerst herkent het bloedplasma-eiwit Von Willebrand Factor (VWF) blootliggend collageen en gaat hiermee een stevige interactie aan. In tweede instantie binden vrij circulerende bloedplaatjes

aan collageen, zowel indirect aan geïmmobiliseerd VWF als direct door middel van twee receptoren, integrine $\alpha_2\beta_1$ en glycoproteïne VI (GPVI). Het contact tussen bloedplaatjes en VWF zorgt ervoor dat bloedplaatjes worden afgeremd ter hoogte van de beschadiging. De directe interacties met collageen worden algemeen verondersteld bij te dragen aan een stevige aanhechting aan de matrix ($\alpha_2\beta_1$) en activatie van bloedplaatjes en verdere reparatieprocessen (GPVI), hoewel dit onderscheid lastig is vast te stellen.

Veel onderzoek is erop gericht om in de toekomst vorming van bloedproppen te kunnen voorkomen op het moment dat ook collageen aan het oppervlak komt maar stolling niet gewenst is: bij vernauwingen of verstoppingen van bloedvaten, of direct na een hartaanval. Juist vanwege zijn veronderstelde rol in de activatie van bloedplaatjes, wordt GPVI gezien als een belangrijke speler in onderzoek naar nieuwe antitrombotische middelen. De moleculaire details omtrent de binding van GPVI aan collageen waren tot nu toe onbekend, hoewel er wel een aanname gedaan werd op basis van de structuur van het receptorgedeelte aan de buitenkant van de cel. In **Hoofdstuk 2** laat ik met behulp van de kristalstructuur van het GPVI-collageen complex zien dat die aanname niet juist was. Deze resultaten onthullen de juiste manier waarop GPVI-collageen binding plaatsvindt, maar geven ook op atomair (± 0.1 -1 nanometer) en moleculair niveau (± 1 -10 nanometer) meer inzicht in het receptorgedeelte dat hierbij betrokken is. Tevens wordt duidelijker welke plekken in collageen een aangrijpingspunt vormen voor deze receptor.

Activatie van Toll-Like Receptor 2 op witte bloedcellen wordt geblokkeerd door *S. aureus*

Het aangeboren deel van ons immuunsysteem bevat een aantal receptoren op witte bloedcellen, die een aantal patronen herkennen afkomstig van micro-organismen, zoals lipiden (bacteriën), stukken DNA/RNA (virussen) en delen van de zweepstaart (bacteriën). Deze receptoren behoren tot één familie van Toll-Like Receptoren (TLRs), maar hebben zich gaandeweg de evolutie gespecialiseerd in de herkenning van verschillende microbiologische patronen. Een belangrijke receptor in de herkenning van de *Staphylococcus aureus* bacterie is TLR2. TLR2 bindt, in combinatie met of TLR1 of TLR6, lipopeptides (een stukje peptide gekoppeld aan een lipide) afkomstig van het bacteriële buitenmembraan. Zelfs als slechts minuscule hoeveelheden van deze lipopeptides worden gedetecteerd, beginnen deze witte bloedcellen ontstekingsbevorderende eiwitten, ofwel cytokinen, uit te scheiden, zodat andere cellen direct in actie kunnen komen om de indringers definitief uit de weg te ruimen.

S. aureus produceert enkele honderden kleine moleculen en eiwitten, zogenaamde virulentiefactoren, om te interfereren in diverse vitale processen van zijn gastheer. Een eerdere studie heeft aangetoond dat *S. aureus* in staat is om juist de TLR2-receptor te blokkeren met twee van deze virulentiefactoren, Staphylococcal superantigen-like (SSL) eiwit 3 en 4. Hoe dit precies gebeurt, en wat het verschil verklaart in activiteit van beide eiwitten was tot dusver niet duidelijk. In **Hoofdstuk 3** presenteren we de kristalstructuur van TLR2 in complex met SSL3. Uit deze structuur blijkt dat SSL3 stevig, en precies op het actieve gedeelte van TLR2 bindt, waarmee de receptor vrijwel geheel wordt uitgeschakeld.

Gelet op de verschillen tussen SSL3 en SSL4, binden beide moleculen hoogstwaarschijnlijk op dezelfde plek, maar met een deels verschillend oppervlak. Dat kan het verschil in TLR2-remming door beide moleculen verklaren. Uit de structuur blijkt

dat TLR2 na binding van SSL3 niet meer in staat is om bacteriële lipopeptides te detecteren en ook geen partnerschap meer aan kan gaan met TLR1 of TLR6; twee stappen die allebei nodig zijn om het afweersysteem te activeren. De bacterie schakelt daarmee ontmaskering van zijn aanwezigheid via deze route uit. Deze kennis is enerzijds van belang om geheel nieuwe type antibiotica te ontwikkelen tegen antibioticaresistente vormen van *S. aureus*, zoals de MRSA-bacterie. Anderzijds creëert deze kennis ook nieuwe mogelijkheden voor de behandeling van chronische ontstekingsziekten waarin TLR2 een rol blijkt te spelen.

Interactie van *S. aureus* virulentiefactoren met andere receptoren

TLR2-activatie is niet het enige proces wat door virulentiefactoren van *S. aureus* wordt aangetast, ook andere moleculen vergelijkbaar met SSL3 en SSL4 beïnvloeden belangrijke mechanismen in onder andere het afweersysteem. Zo binden SSL7 en SSL10 aan complementeiwitten die betrokken zijn bij perforatie van het bacteriële buitenmembraan, bindt SSL6 aan de CD47-receptor die betrokken is bij internalisering van micro-organismen, en bindt SSL5 aan glycoproteïnes Iba (GPIba) en VI (GPVI) die betrokken zijn bij bloedplaatjesactivatie (zie Tabel 2 op pagina 15 voor een volledige lijst). De capaciteiten van *S. aureus* om detectie en opruiming door de gastheer te dwarsbomen, lijken zodoende uit vele facetten te bestaan.

Het beeld dat *S. aureus* interfereert in vele processen wordt wel enigszins vertekend doordat een deel van deze SSL-virulentiefactoren, namelijk SSL2 t/m SSL6 en SSL11, beschikt over eigenschappen om bepaalde koolhydraatstructuren te binden die zich op de meeste celoppervlakreceptoren bevinden. Zulke gemeenschappelijke eigenschappen zijn handig om de concentratie van virulentiefactoren aan het oppervlak van witte bloedcellen te verhogen, maar zijn niet in alle gevallen noodzakelijk voor de specifieke immuunontwijkende activiteit. Het is daardoor voor wetenschappers lastig te achterhalen of SSL-binding aan een receptor specifiek zorgt voor blokkering van deze receptor, of dat binding aan receptor-gebonden koolhydraten alleen een tussenstap is en het SSL-eiwit in werkelijkheid een ander doel heeft. Doordat dit onderscheid niet altijd goed te zien is, kunnen eenvoudig vals-positieve resultaten worden verkregen.

In **Hoofdstuk 4** beschrijf ik een nieuwe methode waarin bindingsexperimenten parallel uitgevoerd worden met receptormoleculen die intacte koolhydraatstructuren bevatten (glucosyl-ketens) en moleculen die een alternatieve koolhydraatstructuur bezitten (mannose-ketens). Door naast de receptor-specifieke SSL ook de binding te meten met een willekeurig ander SSL-eiwit, kunnen specifieke interactiepatronen beter herkend worden. Deze methode heb ik gebruikt om drie SSL-receptor combinaties te onderzoeken, die in eerdere studies met elkaar in verband zijn gebracht: SSL3-TLR2, SSL5-GPVI en SSL6-CD47. Hieruit blijkt onder andere dat de TLR2-gebonden koolhydraten geen rol spelen bij SSL3-binding; iets wat ook blijkt uit de structuur van het SSL3-TLR2 complex in hoofdstuk 3. Verder lijkt SSL5-binding aan GPVI enkel gebaseerd op interactie met de koolhydraatstructuren en ontbreken specifieke componenten. Dit in tegenstelling tot SSL6-binding aan CD47: de interactie tussen beide moleculen is afhankelijk van de aanwezigheid van glucosyl-ketens, maar bevat ook een specifieke bindingscomponent. Deze resultaten zijn dan ook erg interessant voor verder onderzoek naar de mechanismen van CD47-remming door SSL6.

Eiwitkristallografie

Voor het verkrijgen van de resultaten in hoofdstuk 2 en 3 is onder andere gebruik gemaakt van eiwitkristallografie. Kristallografie is een techniek die binnen het eiwitonderzoek gebruikt wordt om de precieze vouwing van de aminozuurketen in een eiwit te bepalen. Dit geeft dus een driedimensionaal beeld van hoe het eiwit er op nanometerschaal uitziet. Deze techniek kan echter ook gebruikt worden om de binding van kleine(re) moleculen aan eiwitten te bestuderen, of, zoals in dit proefschrift beschreven, om de interactie tussen eiwitten te bestuderen. De informatie die dat oplevert, is zeer informatief voor het ontrafelen van de precieze werking van biologische processen.

Een eiwit of twee eiwitten waarvan bekend is dat deze een stabiele interactie vormen, kunnen in een zo zuiver mogelijke vorm in een bijna verzadigde oplossing gebracht worden. Als deze oplossing onder gecontroleerde omstandigheden verder verdampt, kunnen deze eiwitten precipiteren in kristallijne vorm, d.w.z. het eiwit vormt een kristal met een regelmatige pakking van moleculen. Vanwege de periodiciteit in het kristal kunnen lichtverstrooiingsexperimenten uitgevoerd worden, waarbij de golflengte van het licht ongeveer dezelfde grootte heeft als de lengte van één periode in het kristal, oftewel met Röntgenstraling ($\pm 1 \text{ pm} - 10 \text{ nm}$). Uit deze lichtverstrooiing kan nauwkeurig worden teruggerekend wat de dichtheid op elke plek in de repeterende eenheid was. Met deze informatie kan vervolgens een driedimensionaal structuurmodel van het eiwit gebouwd worden.

Door middel van deze techniek kon ik de binding van GPVI aan collageen en de binding van SSL3 aan TLR2 bestuderen. De coördinaten en bijbehorende experimentele data van beide eiwitstructuren zijn vrij beschikbaar in het publieke domein of komen dat kort nadat de resultaten in een wetenschappelijk tijdschrift gepubliceerd zijn. Deze structuurinformatie is verzameld in de RCSB Protein Data Bank (www.rcsb.org) en kan daar bekeken en gedownload worden. In de publicatielijst op pagina 124 van dit proefschrift staan bij elke publicatie de bijbehorende eiwitstructuren vermeld. Tevens staan op de afzonderlijke titelbladen bij elk hoofdstuk ook de PDB-codes vermeld waarmee deze structuren makkelijk gevonden kunnen worden in de Protein Data Bank. De '3D View' optie geeft de mogelijkheid om zelf deze eiwitstructuren tot in detail te bekijken. Op verzoek geef ik daar met alle plezier verdere uitleg en toelichting bij!

Dankwoord

Goed, daar ligt mijn proefschrift dan voor je... Dat werk waar ik in het begin, zes en een half jaar geleden, nog zo enthousiast over sprak, steeds stiller over werd, en op een gegeven moment wellicht de indruk wekte dat het beter was er niet over te beginnen. Het werk, het onderzoek, waar je tegelijkertijd zo van houdt en diepongelukkig van wordt als het niet wil lukken of als er geen einde aan wil komen. Soms kun je je meer indenken dan je daadwerkelijk in het laboratorium voor elkaar kunt krijgen. Ik realiseer me dat dat veel van mijzelf gevraagd heeft, maar ook van iedereen om me heen met wie ik samenwerkte of wilde afspreken. Nu is dan eindelijk het moment om trots te zijn op het resultaat en te lezen waarom het nodig was deze energie hierin te stoppen. Het eindresultaat vind ik zelf erg mooi geworden, maar ik laat de inhoudelijke beoordeling hiervan uiteraard over aan anderen.

Omdat dit alleen dus nooit gelukt zou zijn, wil ik iedereen bedanken die in de voorbije jaren hieraan heeft bijgedragen of me gesteund heeft dit voor elkaar te krijgen. Dat zijn er velen, dat realiseer ik me, en daar ben ik zeer dankbaar voor. Het lukt nooit om hier iedereen apart te noemen, en dat zou misschien ook niet helemaal in mijn eigen stijl zijn: ik doe dat liever 'live' met een hapje of drankje. Op deze plek wil ik graag een aantal mensen bedanken vanwege hun directe bijdrage aan het proefschrift of vanwege fijne herinneringen uit deze periode.

Allereerst wil ik mijn promotor, **Piet Gros**, en copromotor, **Eric Huizinga**, bedanken voor de intensieve begeleiding gedurende het hele traject. Ik heb mij eigenlijk altijd erg op m'n gemak gevoeld, en als onderzoeker-in-spé me bevoorrecht gevoeld om in een lab te werken met vele faciliteiten en waar op onderzoeksgebied zó ontzettend veel mogelijk is. Daarnaast is het denk ik erg bijzonder dat alle promovendi besluiten in de eerste onderwijsweek van het collegejaar met elkaar op vakantie te gaan (onderwijs is goed gekomen, toch?). Ik heb dan ook bijzonder prettig met iedereen samengewerkt, en ik wil iedereen die ik bij Kristal & Structuurchemie heb leren kennen bedanken voor de fijne samenwerking.

In het bijzonder koester ik warme herinneringen aan de talloze reizen die we in deze periode hebben gemaakt. **Martin** en **Remco**, onze eerste reis naar het Zwarte Woud (2013) was wellicht wat nat, maar we zijn wetenschappers om daar dan weer een oplossing voor te vinden. Al met al heel geslaagd! **Deniz & Camilla**, our trip to Istanbul (2014) truly belongs to the most memorable moments during my PhD-time in Utrecht! Deniz, also many thanks to your family for a warm and hospitable welcome in their residence. **Nadia**, also our trip to Brittany (2015) was wonderful, and you showed us so much the beauty of 'your' place. **Joek**, de laatst overgebleven koffiedrinker met smaak, dank je wel dat je ons naar Texel en Marken hebt gebracht! **Hedwich & Matti**, fijn om eindelijk eens de oversteek van Holwerd naar Ameland te voet afgelegd te hebben. **Federico**, it was an honor to visit Pavia and to give a lecture there. **Peng, XG, Tim, Nadia, Camilla, Hedwich**, and **Revina**, CCP4 study weekend in Nottingham and Chinatown in London... **Deniz**, I almost forgot making puzzles and eating raclette in Zürich... **Federico, Hedwich**, and **XG**, remember cheese fondue in Grenoble... **Peng & Christoph**, Weihnachtsmarkt und Glühwein in Aachen... **Els, Harma, Viviana** en **Dimphna**, jullie labuitjes naar Arnhem en

Den Bosch waren fantastisch! **Trusanne**, jij hebt het voor elkaar gekregen ons te laten reizen naar Zwijndrecht en Elburg... **Els, Marian, Hedwich**, waar zijn we niet geweest...?!

Kirsten, met jou heb ik al die tijd nog wel het meest intensief in het lab samengewerkt, aan TLR2 en SSL3, of vice versa. Ons onderzoeksproject, of was het misschien eigenlijk niet helemaal van ons? Volgens mij hebben we iets heel bijzonders voor elkaar gebokst. Het helpt daarbij om grenzeloos vertrouwen in elkaar te hebben als het project een tijd lang helemaal niet wil vlotten. **Carla** en **Jos**, ik heb buitengewoon plezierig met jullie gediscussieerd over SSL-projecten en dit heeft enorm geholpen deze succesvol te maken. **Mark**, wij hebben samen ontzettend veel esterase-structuren weten te verkrijgen. Een buitengewoon goede en fijne samenwerking dus, al had ik niet altijd voldoende tijd naast de andere projecten. **Raoul**, ook jou wil ik bedanken voor deze hartelijke samenwerking en de nuttige – soms pittige – discussies. **Dominique**, ontzettend bedankt voor jouw hulp bij het GPVI-project. Volgens mij heb je daar veel plezier aan beleefd. **Richard**, I appreciate our discussions over the years, which have been very beneficial to bring the GPVI-project further as it is now, and I want to sincerely thank you for those moments.

Pia, Maïke en **David**, het grootste deel van deze zes en een half jaar, heb ik de woensdagavond met jullie doorgebracht. Tikje aan de late kant misschien... zo nu en dan... maar heel fijn om dan even heel iets anders te doen dan wetenschap: zingen, spelletje regenwormen spelen, en rijstsalade eten met karamel-zeezout chocolade toe. Ook wil ik alle zangers van kamerkoor Next bedanken voor dit prachtige nieuwe koor, en ben ik blij dat ik daar deel van mag uitmaken (door uiteraard de uitstekende overtuigingskracht van **Cécile**). **Fokko**, ik denk dat het heel bijzonder is om in zo'n kort tijdsbestek een jong koor zulke uitdagende programma's te laten uitvoeren. **Nicolle** en **Lucas**, 'onze' uitvoering van de Mariavespers was echt fantastisch! Daarnaast heb ik fijne herinneringen aan twee keer een Italië-tournette, waarvoor ik **Federico, Dario** en **Luciano** wil bedanken, en aan Milaan met **Suzanne** en Sestri-Levante met **Trusanne**.

Tenslotte wil ik nog mijn fantastische paranimfen, vrienden en familie noemen. **Revina** and **Deniz**, as ultimate troubleshooters, I want to thank you both for the great help in organizing the defense! **Jeroen**, we wonen tegenwoordig behoorlijk ver uit elkaar, zo'n 200 kilometer, maar we hebben op zich wel bewezen dat die afstand gewoon per fiets af is te leggen. Dat was een heroïsche tocht! **Joram** en **Elizabeth**, het is altijd fijn om bij jullie te eten en logeren. Nu wonen jullie ook nog eens heel handig naast het station ;). **Beitske**, het is ontzettend fijn om wat Vivaldi te spelen, te kletsen over dingen waar we dan weer allebei tegen aanlopen, en Indonesisch te eten (dat laatste moeten we vaker doen) ;). **Trusanne**, jij staat al in praktisch elke bovenstaande alinea vermeld, wat heel bijzonder is, en we delen dus misschien wel meer dan we ons voor kunnen stellen. En dan heb jij ook nog een belangrijk aandeel in de opmaak van dit proefschrift en de illustraties. Ik weet dat je diep vanbinnen graag meer in de wetenschappelijke hoofdstukken had willen voorkomen, maar je bijdrage op deze manier en in onze goede vriendschap is ook van grote waarde. Veel, veel dank daarvoor. Als allerlaatste, gaat misschien wel mijn grootste dank uit naar **pap, mam, Gerard, Harma, Renze, Remco** en **Angélique**. Zonder jullie steun op alle mogelijke manieren was dit niets geworden. Graag blijf ik samen (race)fietsen (niet al te scherpe bochtjes graag), op concerten komen, naar de opera te gaan, oppassen of naar de dierentuin gaan met **Evelijn, Tine** en **Sjoerd**.

List of publications

Structural insights into collagen-binding by platelet receptor Glycoprotein VI

Louris J. Feitsma, T. Harma C. Brondijk, Gavin Jarvis, Dominique Hagemans, Dominique Bihan, Natasia Jerah, Marian Versteeg, Richard W. Farndale, and Eric G. Huizinga
Manuscript submitted.

PDB-codes of deposited crystal structures: [5OU7](#), [5OU8](#), and [5OU9](#)

Betacoronavirus Adaptation to Humans Involved Progressive Loss of Hemagglutinin-Esterase Lectin Activity

Mark J.G. Bakkers, Lang, Yifei, Louris J. Feitsma, Ruben J.G. Hulswit, Stefanie A.H. de Poot, Arno L.W. van Vliet, Irina Margine, Jolanda D.F. de Groot-Mijnes, Frank J.M. van Kuppeveld, Martijn A. Langereis, Eric G. Huizinga, Raoul J. de Groot

Cell Host & Microbe vol. 21 no. 3 (2017); p. 356-366; doi: [10.1016/j.chom.2017.02.008](#)

PDB-code of deposited crystal structure: [5N11](#)

Coronavirus receptor switch explained from the stereochemistry of protein-carbohydrate interactions and a single mutation

Mark J. G. Bakkers, Qinghong Zeng*, Louris J. Feitsma*, Ruben J. G. Hulswit, Zeshi Li, Aniek Westerbeke, Frank J. M. van Kuppeveld, Geert-Jan Boons, Martijn A. Langereis, Eric G. Huizinga, and Raoul J. de Groot

PNAS vol. 113 no. 22 (2016); p. E3111-E3119; doi: [10.1073/pnas.1519881113](#)

PDB-codes of deposited crystal structures: [4ZXN](#), [5JIL](#), [5JIF](#)

Structural basis for inhibition of TLR2 by staphylococcal superantigen-like protein 3 (SSL3)

Louris J. Feitsma*, Kirsten J. Koymans*, T. Harma C. Brondijk, Piet C. Aerts, Eddie Lukkien, Philip Lössl, Kok P. M. van Kessel, Carla J. C. de Haas, Jos A. G. van Strijp#, and Eric G. Huizinga#

PNAS vol. 112 no. 35 (2015); p. 11018–11023; doi: [10.1073/pnas.1502026112](#)

PDB-codes of deposited crystal structures: [5D3D](#), [5D3I](#)

A peptide mimic of the chemotaxis inhibitory protein of *Staphylococcus aureus*: towards the development of novel anti-inflammatory compounds

Anton Bunschoten, Johannes H. Ippel, John A.W. Kruijtzter, Louris J. Feitsma, Carla J.C. de Haas, Rob M.J. Liskamp, Johan Kemmink

Amino Acids vol. 40 no. 2 (2011); p. 731-740; doi: [10.1007/s00726-010-0711-3](#)

CHIPS binds to the phosphorylated N-terminus of the C5a-receptor

Anton Bunschoten, Louris J. Feitsma, John A.W. Kruijtzter, Carla J.C. de Haas, Rob M. J. Liskamp, Johan Kemmink

Bioorganic & Medicinal Chemistry Letters vol. 20 no. 11 (2010); p. 3338-3340; doi: [10.1016/j.bmcl.2010.04.028](#)

* these authors contributed equally to this work

these authors share senior authorship

Curriculum vitae (NL)



Louris Jakob Feitsma werd geboren op 31 oktober 1984 te Veenwouden. Zijn VWO-opleiding (profielen Natuur & Techniek en Natuur & Gezondheid) volgde hij aan het Lauwers College. Gedurende deze periode raakte hij al gefascineerd doorscheikunde, en haalde dan ook twee keer de finaleronde van de Nederlandse Chemie Olympiade in 2002 (11^e plaats) en 2003 (9^e plaats). Dit heeft zijn verdere carrière sterk beïnvloed heeft. Na het VWO startte hij in 2003 met de studie scheikunde aan de Rijksuniversiteit Groningen. Zijn bachelorprogramma focuste op biochemie en fysische chemie en deed hij een stageonderzoek in de groep van Dr. Marco Fraaije, waarin hij onderzoek deed naar enzymatische oxidatie van koolhydraten. Het master-programma Moleculaire Biologie en Biotechnologie (profiel Biochemie) volgde hij ook aan de Rijksuniversiteit Groningen. In zijn eerste onderzoeksstage in de Eiwitkristallografiegroep van Prof. Dr. Bauke Dijkstra werkte hij aan de structuuranalyse van ketosteroid dehydrogenase 1, één van de enzymen die de *Rhodococcus erythropolis* bacterie nodig heeft om cholesterol af te breken. Vervolgens ging hij naar Utrecht, om zijn 2^e onderzoeksstage te doen in de groep van Prof. Rob Liskamp, waar hij werkte aan de synthese van alternatieve peptiden van de C5a-receptor, die mogelijk de schadelijkheid van het CHIPS-eiwit (Chemotaxis-inhibitory proteïn van *Staphylococcus aureus*) kunnen beïnvloeden. Hij studeerde in 2010 af in Groningen, en vervolgde hij zijn loopbaan als promovendus in de groep van Spinozaprijswinnaar Prof. Piet Gros, onder begeleiding van Dr. Eric Huizinga. Zijn onderzoek legde zich voornamelijk toe op kristallisatie en structuurbepaling van belangrijke eiwitten en eiwitcomplexen betrokken bij hemostase en infectie & immuniteit. Samenwerkingen met de onderzoeksgroepen van Prof. Dr. Jos van Strijp (Universitair Medisch Centrum Utrecht), Dr. Raoul de Groot (Faculteit Diergeneeskunde) en Prof. Dr. Richard Farndale (Universiteit van Cambridge) hebben geresulteerd in een aantal publicaties in prominente wetenschappelijke tijdschriften. Onlangs vervolgde Louris zijn carrière in Leiden bij het bedrijf ZoBio, waar hij zijn ervaringen op het gebied van eiwitkristallografie inzet als onderzoeker. Naast eiwitkristallografie is Louris een groot liefhebber van de muziek van componisten uit voorbije eeuwen. Hij speelt orgel/klavecimbel en is daarnaast regelmatig te horen als koorzanger of solist.

Curriculum vitae (EN)



Louris Jakob Feitsma was born on October 31st, 1984, in Veenwouden, The Netherlands. He followed secondary school education at Lauwers College in Buitenpost. During this period, he already started to exhibit his fascination for chemistry and reached two times the final stage of the Dutch National Chemistry Olympiad in 2002 (11th place) and 2003 (9th place). This highly influenced his further career development afterwards. In 2003, he started his study Chemistry at the University of Groningen. During his Bachelor, he mainly followed courses on the topics of biochemistry and physical chemistry. He did his internship in protein chemistry within the research group of Dr.

Marco Fraaije, focusing on enzymatic oxidation of carbohydrates. He continued his master program in Molecular Biology and Biotechnology (profile Biochemistry) at the University of Groningen. His first internship focused on protein crystallography within the group of Prof. Dr. Bauke Dijkstra, in which he worked on a structural analysis of ketosteroid dehydrogenase 1; one of the enzymes required by soil bacterium *Rhodococcus erythropolis* to process cholesterol. Afterwards, Louris moved to Utrecht and started an internship in the group of Prof. Rob Liskamp. During this project, he worked on the synthesis of peptide mimics of the C5a-receptor that can potentially modulate virulence of Chemotaxis-inhibitory protein of *Staphylococcus aureus* (CHIPS). In 2010, Louris obtained his master's degree in Biochemistry and continued with a PhD program under the supervision of Dr. Eric Huizinga in the protein crystallography group of Prof. Piet Gros in Utrecht. His research focused on crystallization and structure determination of several key proteins and protein complexes in hemostasis and infection & immunity. Collaborations with the research groups of Prof. Dr. Jos van Strijp (University Medical Center Utrecht), Dr. Raoul de Groot (Utrecht University), and Prof. Dr. Richard Farndale (University of Cambridge) resulted in a number of publications within prominent peer-reviewed scientific journals. Louris recently continued his career in Leiden working as a scientist at ZoBio and applying his experience in protein crystallography as obtained during his PhD training. Besides protein crystallography, Louris is fascinated by music of ancient classical composers, plays organ/harpsichord, and can be frequently heard in vocal ensembles or as a soloist.