

**MOLECULAR INTERACTIONS OF TECTONIN
PROTEINS IN HOST AND PATHOGEN RECOGNITION**

DIANA LOW HOOI PING

(B. Eng (Hons.), NUS)

**A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN COMPUTATION AND SYSTEMS BIOLOGY (CSB)
SINGAPORE-MIT ALLIANCE
NATIONAL UNIVERSITY OF SINGAPORE**

2009

ACKNOWLEDGEMENTS

There are several people without whom this thesis would not have been at all possible and whom I need to thank:

- Professor Ding Jeak Ling and Professor Chen Jianzhu, my supervisors, for the guidance that they have given to me since Day 1. They have provided me great opportunities to learn in many different environments and to learn from many individuals with diverse backgrounds. I would like to thank Prof. Ding firstly for helping me adapt to the world of biology, and secondly for her continuous and dedicated effort in imparting invaluable knowledge and skills as a researcher. I would like to thank Prof. Chen for his constant support of my work, advice and encouraging words always.
- Singapore-MIT Alliance for their financial support of my post-graduate studies via the SMA Graduate Fellowship.
- Dr. Sivaraman Jayaraman for his advice in my DLS experiments. Dr. Ganesh Anand, for his manifold efforts in every part of my HDMS work. Dr. Adam Yuan for the introduction to crystallography and help in the crystallization efforts and also with the X-ray diffraction data collection.
- Dr. Agnès Le Saux, my mentor – I owe her a big thank you for training me to run and design experiments, reading my progress reports and thesis, and giving me priceless advice on lab work. *Merci beaucoup, vous êtes une bonne maitresse.*
- Dr. Vladimir Frecer - for helping me with all the computational studies, reading the many manuscript drafts and working thru emails. Also thank you for the incredible time in Trieste – work and play was equally hard but was so much fun.

- My lab mates, past and present, for the camaraderie and the help rendered in all aspects of lab life, and life in general. To the late-night workers - your company, stories, anecdotes, advice and laughter are much appreciated.
- MIT-CCC, J.U.S.T. Us – Thank you for the prayer support & encouragement and for being constant reminders on where our hope, faith and focus should always be.
- Last but not least, I would like to thank the people closest to me – Dad, Mom, Chris and Grace - for always being there for me and listening to *everything*, even from afar, and for reminding me to always take it easy, and have a sense of humour in everything I do.

This thesis is dedicated

*“.. to Him who is able to do exceedingly abundantly above all that we ask or think,
according to the power that works in us”*

Eph 3:20

TABLE OF CONTENTS

Acknowledgements	i
Table of Contents	iii
Summary	vi
List of Tables	viii
List of Figures	viii
List of Abbreviations	xiii
List of Primers	xv
1 General introduction and overview	
1.1 Overview of the innate immune system	1
1.1.1 The prophenoloxidase pathway	4
1.1.2 The complement system	5
1.1.3 Serine proteases as activators and enhancers of immune response	6
1.2 Recognition of pathogens and activation of the innate immune system	9
1.2.1 Pathogen recognition receptors - CRP and GBP as key innate immune molecules	9
1.2.2 The beta-propeller structure and proteins	10
1.2.3 The interactome hypothesis	15
1.2.4 Pathogen associated molecular patterns	17
1.2.4.1 Lipopolysaccharide	18
1.2.4.2 Lipid A	21
1.2.5 Antimicrobial peptides	22
1.3 Horseshoe crab as an ideal experimental model host for innate immunity study	24
1.4 Overview of thesis	26
2 Materials and Methods	
2.1 Computational Analysis	
2.1.1 Bioinformatics analysis	27
2.1.2 Protein homology modeling	27
2.1.3 Construction of bacterial and saccharide structures	29
2.1.4 Protein-protein and protein-ligand docking	29
2.1.5 Identification of LPS-binding motifs	30
2.1.6 Design and synthesis of LPS-binding peptides	30
2.2 Preparative Methods	
2.2.1 Organisms	31
2.2.2 Biochemicals and enzymes	31
2.2.3 Medium and agar	32
2.2.4 Collection of cell-free hemolymph	33

2.2.5	Depyrogenation of equipment and buffers	33
2.2.6	Purification of GBP from cell-free hemolymph	33
2.3 Analytical Methods		
2.3.1	SDS-PAGE analysis	34
2.3.2	Western blot	34
2.3.3	Mass spectrometry	35
2.3.4	ELISA	36
2.3.5	Yeast 2-hybrid co-transformation assay	37
2.3.6	Yeast 2-hybrid library screening	39
2.3.7	Dynamic light scattering analysis	42
2.3.8	Protein crystallization	42
2.3.9	Amide hydrogen exchange mass spectrometry (HDMS) and data analysis	42
2.3.10	Surface plasmon resonance analysis	45
2.3.11	Pyrogene assay for endotoxicity	45
3 GBP, a representative Tectonin protein in innate immune defense		
3.1 Introduction		
3.1.1	Tectonin domains in beta-propeller repeats	47
3.1.2	Lectins	51
3.1.3	The Tectonin domain	55
3.2 Results and Discussion		
3.2.1	Biochemical properties of GBP	56
3.2.1.1	Purified GBP shows polymeric forms	56
3.2.1.2	GBP is a multimeric complex in solution	60
3.2.1.3	Purified GBP retains saccharide-binding function	62
3.2.2	The GBP structure	63
3.2.2.1	Crystallization of GBP and CRP for structure determination	63
3.2.2.2	Computational modeling of GBP and CRP structure	66
3.2.2.3	Saccharides and LA dock to similar sites in GBP	74
3.2.3	Molecular mechanism of GBP:LPS interaction	78
3.2.3.1	GBP interacts with LPS via sugar groups	78
3.2.3.2	The different lengths of LPS bind strongly to GBP	79
3.2.3.3	The interaction between GBP and LPS is independent of Ca ²⁺	81
3.2.3.4	GBP interacts with LPS via distinct interaction surfaces	82
3.2.4	Mechanism of action of GBP : CRP interaction	85
3.2.4.1	HDMS reveals that GBP interacts with CRP through a non-symmetrical protein-protein contact	85
3.2.4.2	Yeast 2-hybrid interaction analyses show consistent interaction domains with HDMS observations	90
3.2.4.3	Protein-protein docking reaffirms the feasibility of GBP:CRP binding region	91

3.2.5	Effects of infection on the GBP, CRP and LPS interactions	93
3.2.5.1	Infection enhances interaction between GBP and CRP to LPS	93
3.2.5.2	Infection causes irreversible conformational change to GBP and CRP	96
3.2.5.3	GBP and CRP binds and disrupts LPS micelles and exposes its endotoxicity	97
3.3	Summary	99
4	hTectonin – Discovery of a novel Tectonin protein in the human	
4.1	Introduction	101
4.1.1	Are the Tectonin proteins evolutionary conserved? Do PRR:PRR interactomes exist in the human system?	101
4.2	Results and Discussion	
4.2.1	hTectonin – a distantly related homolog of the invertebrate Tectonins	
4.2.1.1	hTectonin is widely present across the various species	104
4.2.1.2	hTectonin β -sheets are highly conserved	106
4.2.2	In search for interaction partners of hTectonin	
4.2.2.1	hTectonin gene is expressed in immune cell lines	107
4.2.2.2	hTectonin interacts with immune-related molecules in the leukocyte library	108
4.2.2.3	hTectonin interacts with ficolin	114
4.2.2.4	hTectonin protein expression increases in response to LPS stimulation	115
4.2.3	LPS-binding peptides in Tectonins	116
4.2.3.1	An algorithm was developed for large-scale screening of LPS- binding motifs in proteins	117
4.2.3.2	hTectonin contains LPS-binding motifs	119
4.2.3.3	Designed predicted LPS-binding peptides are hydrophilic, synthetically feasible and suitable for binding analysis.....	120
4.2.3.4	hTectonin- and GBP-derived Tectonin peptides bind LPS with high affinity	122
4.3	Summary	125
4.4	Common features between GBP and hTectonin	126
5	Conclusion	128
6	Future perspectives	132
7	References	135
8	Publications	
9	Appendix A	

SUMMARY

Beta-propeller proteins exhibit diverse functions in catalysis, protein-protein interaction, cell-cycle regulation, and immunity. Tectonins, a sub-class of β -propeller family, have been implicated in bacterial binding. Our prediction revealed that the galactose-binding protein (GBP) in the horseshoe crab, *Carcinoscorpius rotundicauda*, is an all-beta sheet protein consisting of Tectonin domains. Studies have shown that upon binding to Gram-negative bacterial lipopolysaccharide (LPS), GBP interacts with C-reactive protein (CRP) and carcinolectin (CL5) to form a pathogen recognition complex. However, the molecular basis of interactions between GBP and LPS and how it interplays with CRP remains largely unknown. Here, we sought to unravel the mechanisms of interaction by examining the structure-function relationship, with a view to understanding the pathophysiological implications of the Tectonin domain-containing proteins and the possible conservation of this concept in the mammalian system. Through homology modeling, GBP was revealed to be a 6 β -propeller toroidal structure. Interestingly, the seemingly repetitive and identical domains were able to simultaneously bind LPS and CRP via separate domains, suggesting that the Tectonin domains can differentiate self/non-self, which is crucial to frontline defense against infection. Infection condition, which was mimicked by Ca^{2+} chelation, increased the GBP-CRP affinity by 1000-fold. Re-supplementing the system with physiological levels of Ca^{2+} did not reverse the protein-protein affinity to basal state, suggesting that the infection-induced complex had undergone irreversible conformational changes. GBP was also able to increase the endotoxicity of LPS, prompting suggestions that it probably disrupts LPS micelles and exposes the endotoxic potential of LPS. *In vivo*, this may translate into the ability of GBP to bind

LPS and perturb the Gram-negative bacterial outer membrane while serving to prime the immune system for stronger downstream immune response. Phylogenetic analysis revealed strong evolutionary conservation of the Tectonin homologues from the plasmodium to human. We identified the hTectonin, a then hypothetical protein in the human genome database, as a potential distant homolog of GBP. hTectonin, like GBP, formed β -propellers with multiple Tectonin domains. It is present in the human leukocyte and interacts with M-ficolin, a known human complement protein whose ancient homolog, CL-5, is the functional protein partner of GBP. Furthermore, the affinity of hTectonin-derived LPS-binding peptides is comparable to that of the GBP-derived peptides. By virtue of a recent finding of another Tectonin protein called the leukolectin in the human leukocyte we propose that the Tectonin proteins could play an important role in innate immune defense and that this function has been conserved over several hundred million years, from invertebrates to vertebrates.

LIST OF TABLES

Table No.	Title	Page
Chapter 1		
1.1	The various functions of β -propeller proteins	13
Chapter 3		
3.1	Purification summary of GBP	59
3.2	DLS measurements for GBP of molecular radius, diffusion coefficient and molecular weight in solution over time	61
3.3	List of crystallization kits used in attempts to crystallize GBP and CRP	64
3.4	Data obtained from CRP crystal diffraction	65
3.5	List of Phi-Psi outliers in the GBP model	68
3.6	List of Phi-Psi outliers in the CRP model	73
3.7	Computed binding energies for top scoring saccharides and lipid A poses docked to GBP	76
3.8	Rate constants and equilibrium dissociation constants of binding kinetics of ligands to GBP	80
3.9	Summary of H/ ² H exchange data for GBP	88
3.10	Summary of H/ ² H exchange data for GBP	89
Chapter 4		
4.1	Putative interaction partners of hTectonin identified via yeast 2-hybrid screening of the human leukocyte cDNA library	110
4.2	Dissociation constants of Tectonin peptides when bound to LPS, ReLPS, lipid A	125

LIST OF FIGURES

Figure No.	Figure Title	Page
Chapter 1		
1.1	Innate immunity	2
1.2	The innate immune system plays a critical role in adaptive immunity	4
1.3	The prophenoloxidase pathway in invertebrates	5
1.4	The complement cascade	6
1.5	The coagulation cascade in the horseshoe crab	7
1.6	Serine proteases regulate the assembly of PRRs to prompt the innate immune response	8
1.7	Structures of hCRP, TtCRP and LpCRP	10
1.8	The structure of TL-1	11
1.9	Oligomers of TPL-1 and GBP	11
1.10	An example of a beta-propeller fold formed by 4 anti-parallel β -strands	12
1.11	Examples of beta propeller proteins with different number of folds	14
1.12	Features of the β -propeller structure	14
1.13	CRP interacts with GBP upon host infection	15
1.14	A model for pathogen recognition assembly via interaction between CrOctin and other PRRs in the horseshoe crab	16
1.15	Protein interaction network of GBP and CRP	16
1.16	Bacterial PAMPs and their recognition by various PRRs	17
1.17	Schematic diagram of the cell wall of Gram-negative bacteria	19
1.18	Structural organization of LPS in the Enterobacteriaceae	20

1.19	Detailed chemical structure of the <i>Salmonella minnesota</i> LA	18
1.20	The different mechanisms of action of antimicrobial peptides	21
1.21	The horseshoe crab	24
Chapter 2		
2.1	Cloning of genes into yeast 2-hybrid vectors	38
2.2	Colony selection of successful transformants	39
2.3	Hydrogen exchange mass spectrometry	43
Chapter 3		
3.1	Identified conserved sequence motifs in known beta-propeller repeats such as the WD-repeat, Kelch, YWTD and others	49
3.2	The 3D structure of the heterotrimeric G-protein complex	50
3.3	The 3D structure of neuraminidase	51
3.4	Sialic acid structural analogues and their mechanism of action	51
3.5	Schematic examples of major types of animal lectins, based on protein structure	52
3.6	Mechanism of saccharide binding observed in symmetrical β -propeller fold structures	53
3.7	Plasmodium stage of <i>Physarum polycephalum</i>	55
3.8	Consensus sequence for the definition of a Tectonin domain	56
3.9	GBP is co-purified with partners at pH7.4	58
3.10	GBP is purified at pH8.8	58
3.11	Electrophoretic analyses of GBP	59
3.12	Mass spectrometry analysis of purified GBP bands	60

3.13	GBP exists in polymeric form	61
3.14	OD405 measurements during GBP purification	62
3.15	GlcNAc-removed GBP is able to re-bind Sepharose	63
3.16	Crystals of GBP and CRP	65
3.17	X-ray diffraction pattern of CRP crystal	65
3.18	Secondary structure prediction of GBP	66
3.19	The 6 internal tandem repeats in the sequence of GBP	66
3.20	Sequence alignment of GBP to Tachylectin-1	67
3.21	The Ramachandran plot of the GBP homology-modeled structure	68
3.22	Homology model of GBP structure	69
3.23	The protein sequence of GBP showing 6 Tectonin domain repeats	69
3.24	Cysteine residues of GBP	70
3.25	Surface properties of GBP	71
3.26	Prediction of flexible residues on GBP	71
3.27	The structure of the homology modeled CRP	72
3.28	Ramachandran plot for the structure of CRP	73
3.29	Saccharides dockin to the hydrophobic clefts and central cavity of GBP	75
3.30	LA docks to the hydrophobic cleft of GBP	77
3.31	The sugar moieties of LPS and LTA bind to GBP	79
3.32	GBP binds LPS with high affinity	80
3.33	The interaction between GBP and LPS is independent of Ca^{2+}	81
3.34	Example of HDMS mass shift in the GBP sample	82

3.35	HDMS analysis – surface interaction of GBP with either GlcNAc or LA	84
3.36	HDMS analysis – surface interaction of GBP with CRP	86
3.37	HDMS analysis – surface interaction of CRP with GBP	87
3.38	Yeast 2-hybrid analysis shows specific Tectonin domains of GBP interact with CRP	91
3.39	Guided docking of the GBP-CRP interaction	93
3.40	Binding affinity of infected GBP	94
3.41	Effect of infection upon GBP-CRP interaction with LA	95
3.42	Effect of calcium on infected proteins binding to LA	96
3.43	GBP and CRP exposes LPS endotoxicity	98
3.44	Molecular size of GBP polymers increase with addition of LPS	99
3.45	Model of micelle disruption to increase LPS endotoxicity	99
3.46	Proposed mechanism of GBP interactions and formation of the pathogen recognition complex for innate immune response	100
Chapter 4		
4.1	hTectonin is distantly related to the invertebrate Tectonins	105
4.2	The hTectonin gene is widespread across many species	106
4.3	hTectonin forms β -sheets in its Tectonin repeats	107
4.4	hTectonin cDNA is found in the human T cells (A549), monocytes (U937) and leukocytes	108
4.5	Selected hTectonin partners for interaction confirmation	114
4.6	hTectonin interacts with ficolin	115
4.7	hTectonin protein expression increases in response to LPS stimulation	116
4.8	Algorithm workflow of the LPS motif search application	118

4.9	The LPSMotif program	118
4.10	LPS-binding motifs in Tectonin proteins	119
4.11	hTectonin LPS-binding motifs conserved in other species	120
4.12	Hydrophobicity plots analysis of GBP and hTectonin peptides	121
4.13	Peptides derived from the Tectonin domains of GBP and hTectonin bind LA with high affinity	123
4.14	Peptides derived from the Tectonin domains of GBP and hTectonin also bind LPS and ReLPS with high affinity	124

LIST OF ABBREVIATIONS

ABTS	2,2'-azino-bis[3-ethylbenzthiazoline-6-sulfonic] acid
BLAST	Basic local alignment search tool
BSA	Bovine serum albumin
cDNA	Complementary deoxyribonucleic acid
CFU	Colony-forming unit
CL-5	Carcinolectin-5
CRD	Carbohydrate recognition domain
CRP	C-reactive protein
CTL	Cytotoxic T lymphocytes
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytosine triphosphate
DEPC	Diethylpyrocarbonate
dGTP	Deoxyguanosine triphosphate
DLS	Dynamic light scattering
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleoside triphosphate
DO	Drop out
dTTP	Deoxythymidine triphosphate
EDTA	Ethylene diamine tetra acetic acid
EGTA	Ethylene glycol tetra acetic acid
ELISA	Enzyme-linked immunosorbent assay
Gal	Galactose
GBP	Galactose-binding protein
GlcNAc	N-acetyl-glucosamine
h	Hours
HDMS	Hydrogen deuterium exchange mass spectrometry
HMC	Hemocyanin
hpi	Hours post-infection
IPTG	Isopropyl-1-thio- β -D-galactopyranoside
kb	Kilobase
kDa	KiloDalton
KDO	2-keto-3-deoxyoctonate
LA	Lipid A
LB	Luria-Bertani
LPS	Lipopolysaccharide
LTA	Lipoteichoic acid
M	Molar
m/z	Mass/charge

MALDI-TOF	Matrix-assisted laser desorption ionization-time of flight
min	Minute
MOE	Molecular Operating Environment
MOWSE	Molecular Weight Search
OD	Optical density
<i>P. aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
PAMP	Pathogen-associated molecular pattern
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PMSF	Phenylmethylsulphonylfluoride
PRR	Pattern recognition receptors
PSIPRED	Protein secondary structure prediction software
QDO	Quadruple drop out
ReLPS	Lipopolysaccharide that is truncated to the “e” portion of its RR-core-polysaccharide
s	Seconds
SD	Synthetic defined
SDS	Sodium Dodecyl Sulfate
SMART	Simple molecular alignment research tool
SPR	Surface Plasmon resonance
TBS	Tris-buffered saline
TBST	Tris-buffered saline with Tween-20
TEMED	Tetramethylethylenediamine
v/v	Volume/volume ratio
w/v	Weight/volume ratio
YPDA	Yeast extract, peptone, dextrose, adenine

LIST OF PRIMERS

Name	Sequence (5' – 3')	Amino acid sequence	Purpose
GBP gene specific primers			
GBP-1F-NdeI	gcccatatgGAATGGA CCCACATTAATGG A	AEWTHIN	Forward primer for cloning GBP from domain 1 into pGBKT7
GBP-2F-NdeI	cgcatatgAGCAACTG GATAAAAGTGGAA GG	SNWIKVE	Forward primer for cloning GBP from domain 2 into pGBKT7
GBP-3F-NdeI	cgcatatgGGAAGTTG GACACAAATCAA	GSWTQIK	Forward primer for cloning GBP from domain 3 into pGBKT7
GBP-4F-NdeI	cgcatatgGGGGAATG GGAATTGGTAGAT	GEWELVD	Forward primer for cloning GBP from domain 4 into pGBKT7
GBP-5F-NdeI	cgcatatgGGTGTCTG GGAGAACATCC	GVWENIP	Forward primer for cloning GBP from domain 5 into pGBKT7
GBP-6F-NdeI	ATTTTCGAGTCAG TCCCTTAActgagggc c	GQWVRLP	Forward primer for cloning GBP from domain 6 into pGBKT7
GBP-1R-EcoRI	gcgaattcttaACCGGTA CACGGTCG	RCTRPCTG	Reverse primer for cloning GBP from domain 1 into pGBKT7
GBP-2R-EcoRI	cggaattcttaAGTACCA TCCACTGGCC	YKRPVDGT	Reverse primer for cloning GBP from domain 2 into pGBKT7
GBP-3R-EcoRI	cggaattcttaATTACAA GGCTTAGGACATT TG	FKCPKPCN	Reverse primer for cloning GBP from domain 3 into pGBKT7
GBP-4R-EcoRI	gcgaattcttaACTGCCG TCAACGGG	YRRPVDGS	Reverse primer for cloning GBP from domain 4 into pGBKT7

GBP-5R-EcoRI	cggaattcttaTGAACAA GGTTTCTTGCAGC G	FRCKKPCS	Reverse primer for cloning GBP from domain 5 into pGBKT7
GBP-6R-XhoI	ggcctcgagttaTAGGA AAGGATTCAACCA GCA	DDIFESVP	Reverse primer for cloning GBP from domain 6 into pGBKT7
hTectonin specific primers			
BC053591-F- HindIII	gcgaagcttatgCCCAAC TCAGTGCTGTGG	ATGCCCAA	Forward primer for cloning hTectonin from into pGBKT7
BC053591-R- XhoI	gcgctcgagGCAGCAG ACGGGGCCATGGG C	GCTGCTGA	Reverse primer for cloning hTectonin from into pGBKT7

All primers were reconstituted in water to 100 μ M stock and stored at -20 °C.

CHAPTER 1

GENERAL INTRODUCTION AND OVERVIEW

1.1 Overview of the innate immune system

The innate immune system serves to protect the host from microbial infection and is thought to be the dominant host defense system in many organisms (Litman et al. 2005). The innate response is usually triggered when microbes are identified by pattern recognition receptors (PRRs) present in the host (Figure 1.1), which recognize components that are conserved among broad groups of microorganisms (Medzhitov 2007). These microbe-specific molecules that are recognized by a given PRR are called pathogen-associated molecular patterns (PAMPs) and include bacterial carbohydrates (e.g. lipopolysaccharide (LPS)), nucleic acids (e.g. bacterial or viral DNA or RNA), bacterial peptides (flagellin), peptidoglycans and lipoteichoic acids (from Gram-positive bacteria), N-formylmethionine, lipoproteins and fungal glucans. PRRs also recognize damaged, injured or stressed cells that send out alarm signals, many of which are recognized by the same receptors as those that recognize pathogens (Matzinger 2002).

Although it does not give long-lasting protection to the host, the innate immune system is crucial in providing immediate defense against infection (Janeway and Medzhitov 2002) in all classes of plant and animal life, and is the only form of immune defense in the invertebrates. The invertebrate innate immune system employs several mechanisms to recognize and eliminate pathogens: (i) prophenoloxidase pathway to oxidatively kill invading microorganisms, (ii) complement-mediated antimicrobial action, (iii) blood coagulation to immobilize the invading microbes, (iv)

lectin-induced complement pathway to lyse and opsonize the pathogen, and (v) prompt synthesis of potent effectors, such as antimicrobial peptides (Ding et al. 2004).

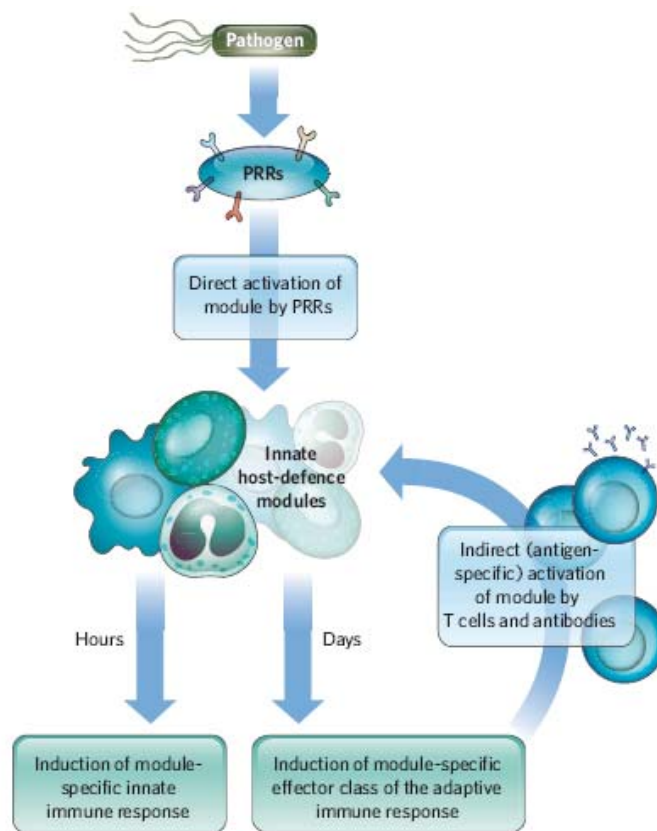


Figure 1.1: Innate immunity. PRRs of the innate immune system recognizes pathogens and activate downstream innate defense modules like coagulation pathway, complement pathway and also induce further action by the adaptive immune system. Figure adapted from Medzhitov (2007).

Inflammation is one of the first responses of the innate immune system to infection or tissue injury. Inflammation is stimulated by chemical factors (Sturtinova 1995) released by the affected cells and serves to establish a physical barrier against the spread of infection, and to promote healing of any damaged tissue following the clearance of pathogens. Another branch of innate immune response is through the complement pathway. It is a biochemical cascade of the immune system that helps, or “complements”, the ability of antibodies to clear pathogens or mark them for destruction by other cells (Janeway 2005). The complement cascade is made up of the classical, alternate and lectin-mediated pathways that are composed of many small

plasma proteins. Complement components can be found in many species evolutionarily older than mammals including plants, birds, fish and some species of invertebrates for example the horseshoe crab, which is dubbed the “living fossil”, with evolutionary history of several hundred million years (Zhu et al. 2005). The conservation of the complement system from the horseshoe crab to humans reflects the significance of the innate immune armament to curb pathogen invasion.

The delayed response of the adaptive immune system requires that the innate immune system continues to serve as the frontline defense system even for vertebrates. Research has shown that the adaptive immune system is not independent of the innate immune system (Medzhitov and Janeway 1997; Katsikis et al. 2007). Rather, it is influenced by recognition signals from the innate immune system as an input to help shape its development and utilizes the effector apparatus of the innate immune system as part of its ammunition towards invaders (Figure 1.2). Downstream consequences of the immune response include the recruitment of complement components of the innate immune system by antibodies – phagocytic macrophages and dendritic cells "present" antigens to T-cells to initiate both cell-mediated and antibody-mediated adaptive immune responses. The interaction of PAMPs and Toll-like receptors (TLRs) on dendritic cells causes them to secrete cytokines, including interleukin 6 (IL-6), which interfere with the ability of regulatory T-cells to suppress the responses of effector T-cells to the antigen (Kimball 1994).

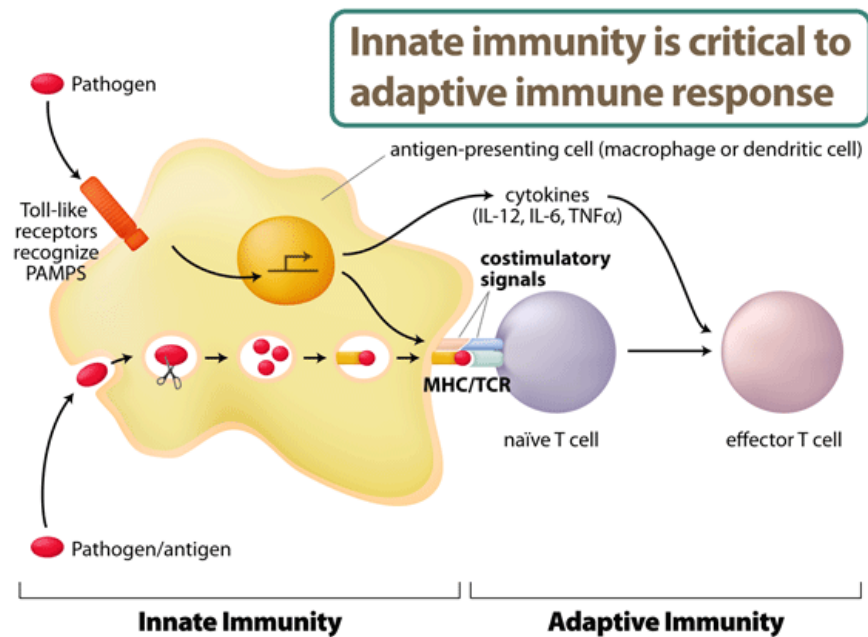


Figure 1.2: The innate immune system plays a critical role in adaptive immunity. Components of the innate immune system (e.g. the TLRs) recognize PAMPs; they process and prepare the pathogen to be presented to the T-cells of the adaptive immune system. Figure adapted from <http://research4.dfci.harvard.edu/innate/innate.html>.

1.1.1 The prophenoloxidase pathway

The prophenoloxidase (PPO) pathway (Figure 1.3) is a major defense pathway in the invertebrates (Cerenius and Soderhall 2004). The activation of PPO to phenoloxidase (PO) results in the production of the highly reactive cytotoxic quinone which produces reactive oxygen species (ROS) to kill the microbial intruder effectively. As the toxicity of ROS poses a dilemma to the host's own survival, its production must be tightly controlled. When activated, the phenoloxidase oxygenates monophenols to o-diphenols and further to o-quinones, which are intermediate products for melanin formation. Melanin, serves as a shield to prevent the spread of the microbial intruder, while the highly reactive intermediate products (quinones) are antimicrobial and cytotoxic. Work in our lab also recent showed that the PPO activity of horseshoe crab hemocyanin (HMC), initially triggered by microbial proteases, can be further enhanced by PAMPs (Jiang et al. 2007). This demonstrates a direct antimicrobial

strategy by which the host opportunistically exploits the invading microbe's virulence factors to convert its respiratory proteins into potent ROS producers to effectively kill the intruder.

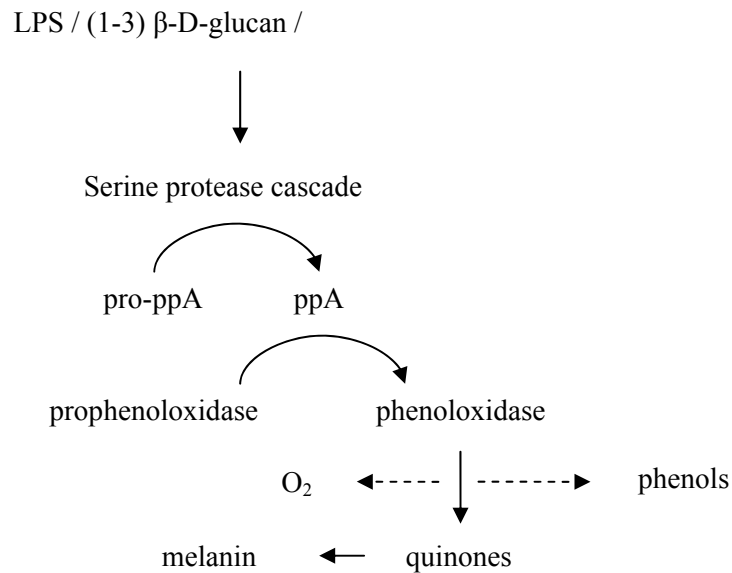


Figure 1.3 The phenoloxidase pathway in invertebrates. β -1,3-glucan, lipopolysaccharide and peptidoglycan are bound by pattern-recognition receptors (PRRs) which then trigger a serine protease cascade leading to the activation of the prophenoloxidase-activating enzyme (ppA) from its pro-form. The activated ppA which is also a serine protease then cleaves the prophenoloxidase into phenoloxidase. The phenoloxidase enzyme then converts monophenols into quinones via oxygenation. Quinones are the intermediate compounds for melanin formation and are highly reactive and toxic to cells. Adapted and modified from Cerenius and Soderhall (2004).

1.1.2 The complement system

One major pathway in innate immunity that can be activated by PRRs is the complement system. Even though it belongs to the innate immune system, it can be recruited and brought into action by the adaptive immune system. Thus, it also acts as an initiator of events to trigger downstream action related to the adaptive immune system.

Three biochemical pathways activate the complement cascade: the classical complement pathway, the alternative complement pathway, and the mannose-binding lectin pathway (Janeway 2005) (Figure 1.4). Major examples of PRRs capable of activating the complement system include C-reactive protein (CRP, classical complement pathway) and the ficolins (lectin pathway). The complement system consists of a number of small proteins found in the blood, normally circulating as inactive zymogens (enzyme precursors). When stimulated by one of several triggers, proteases in the system specifically cleave the proteins to release cytokines and initiate an amplifying cascade of further cleavages. The end-result of this activation cascade is massive amplification of the response and activation of the cell-killing membrane attack complex (MAC). Over 20 proteins and protein fragments make up the complement system, including serum proteins, serosal proteins, serine proteases and cell membrane receptors.

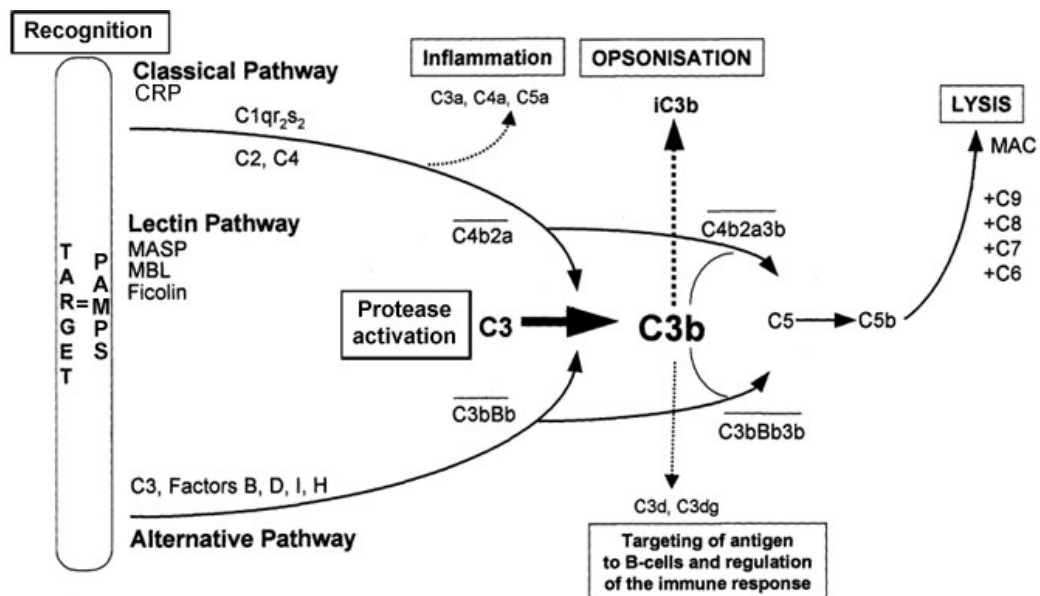


Figure 1.4: The complement cascade. Major PRRs like CRP and ficolin and proteinases like mannan-binding lectin-associated serine proteinase (MASP) can trigger the complement cascade via the classical, lectin and alternative pathways. Once activated, a series of reactions takes place which ultimately leads to downstream effector actions resulting in inflammation, opsonisation of pathogens or lysis of cells via the MAC. Figure adapted with modifications from DeFranco et. al. (2008).

1.1.3 Serine proteases as activators and enhancers of immune response

The serine proteases seem to be a recurrent member in many of the immune related pathways, as can be seen in the description of the prophenoloxidase and complement pathways above. Besides the prophenoloxidase and complement pathways, serine proteases are also involved in the clotting pathway of vertebrates and the coagulation pathway of the horseshoe crabs (Figure 1.5) (Ding and Ho 2001). In the horseshoe crab, when Gram-negative bacteria invade the hemolymph, hemocytes detect LPS molecules on their surfaces (Ariki et al. 2004). The hemocytes then release granular components including two serine protease zymogens, named Factors C and G, which are autocatalytically activated by LPS or (1→3)- β -D-glucan, which are major components of the cell walls of Gram-negative bacteria and fungi, respectively. Factor B zymogen is then activated by Factor C to its active form (Factor B'), which activates proclotting enzyme to clotting enzyme. Clotting enzyme then converts coagulogen to an insoluble coagulin gel (Kawasaki et al. 2000).

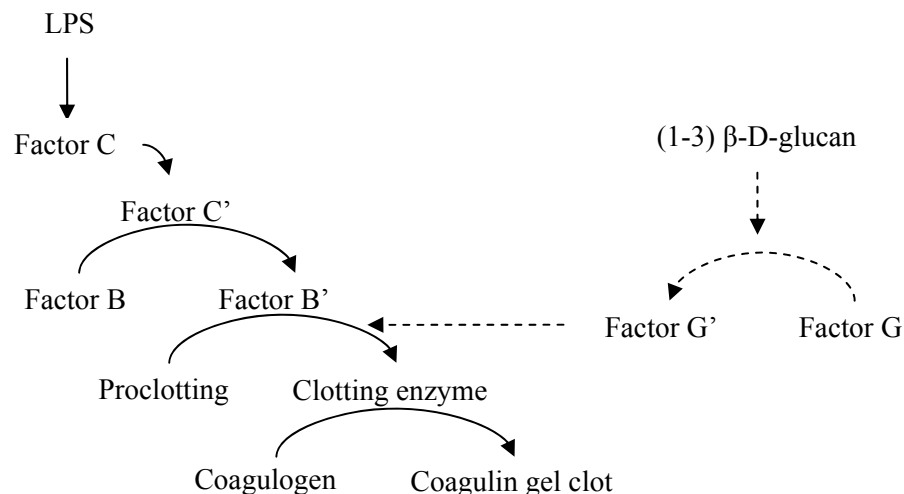


Figure 1.5: The coagulation cascade in the horseshoe crab. The coagulation pathway in the horseshoe crab can be triggered by LPS which autocatalyses the Factor C zymogen into an active serine protease, Factor C'. Factor C' then cleaves and activates another serine protease, Factor B, which in turn activates the proclotting enzyme. An insoluble clot forms when the clotting enzyme activates coagulogen into coagulin. Clotting can also be triggered by (1-3) β -D-glucan which acts via another serine protease zymogen Factor G. Activated Factor G then taps into the coagulation pathway via cleavage of proclotting enzyme. Adapted from Ding and Ho (2001).

Our recent work on the horseshoe crab revealed that 2 serine proteases, FC and C2/Bf found in the complement system, once activated, can influence and increase levels of recruitment of two PRRs - carcinolectin-5 (CL-5) and the galactose-binding protein (GBP) - to the bacterial surface (Le Saux et al. 2008). In turn, this phenomena mostly likely serves to enhance the complement pathway activity which involves another key PRR, CRP (Figure 1.6). Overall, GBP seems to be the key anchor in many reactions: (1) binding to bacterial LPS, (2) bridging of a vast network of PRRs, and (3) influential in the recruitment and binding to complement proteins which leads to key downstream interactions like the release of chemokines and cytokines.

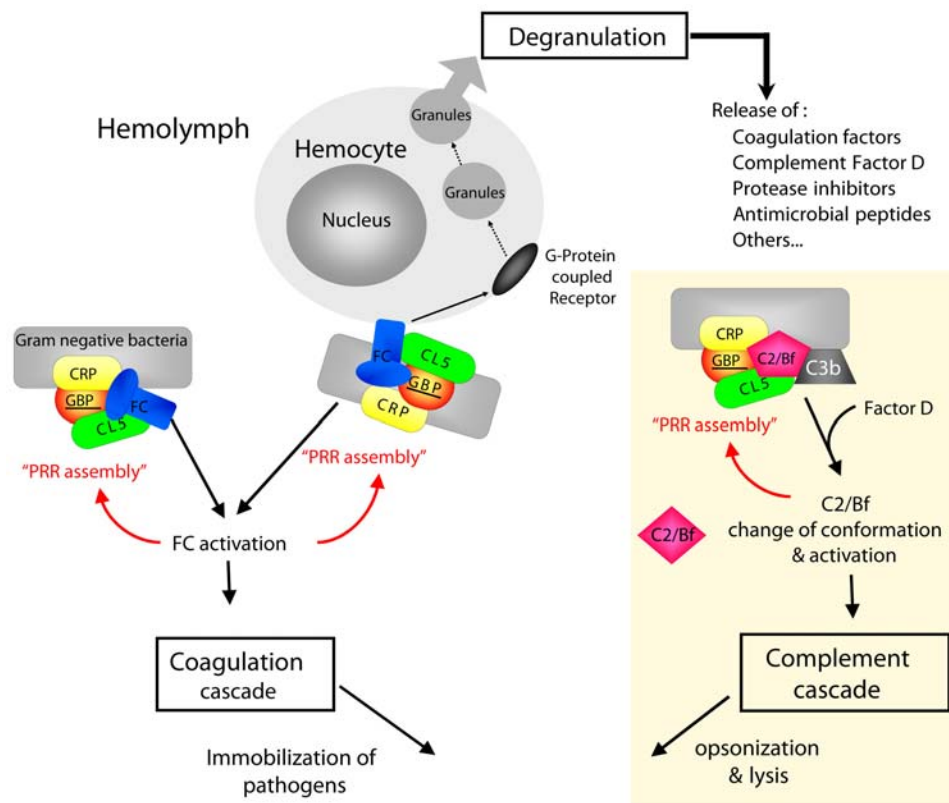


Figure 1.6: Serine proteases regulate the assembly of PRRs to prompt the innate immune response. Serine proteases are known for activating components of the complement and coagulation cascade. Le Saux et. al. (2008) showed that positive feedback of FC and C2/Bf in regulating the upstream pathogen-recognition assembly conceivably strengthens the immune response.

1.2 Recognition of pathogens and activation of the innate immune system

1.2.1 Pathogen recognition receptors - CRP and GBP as key innate immune molecules

In Section 1.1, we observed that CRP, and especially GBP are the main players in the innate immune response of the horseshoe crab. CRP is an acute phase plasma protein. The human CRP (hCRP) was first discovered by Tillett and Francis in 1930 as a substance in the serum of patients with acute inflammation that reacted with the C polysaccharide of pneumococcus (Thompson et al. 1999). The level of hCRP rises dramatically during acute inflammatory processes that occur in the body in reaction to the pneumococcus infection. It is thought to assist in complement binding to foreign and damaged cells and affect the humoral response to diseases. In the clinical setting, hCRP is used mainly as a marker of inflammation. Thus, measuring and charting hCRP levels can prove useful in determining the progress of the disease or indicate the effectiveness of treatment regimes. In the horseshoe crab, CRP is also believed to play an important role in innate immunity, as an early defense PRR against infections, making it a key pathogen recognition receptor (Ng et al. 2004; Ng et al. 2007).

Structurally, hCRP is a 206 amino acid polypeptide which assembles into a radially symmetrical pentamer, characteristic of pentraxin family members (Thompson et al. 1999) (Figure 1.7A). Interestingly, the horseshoe crab CRP of the *Tachypleus tridentatus* species (TtCRP) was revealed to be hexagonal ring-like structure instead. A side-on orientation of TtCRP showed a dimeric rectangular structure. Although hexameric by structural homology, TtCRP is a member of the pentraxin superfamily as well (Iwaki et al. 1999) (Figure 1.52B). The *Limulus* horseshoe crab CRP (LpCRP)

has also been modeled as a hexameric structure on the basis of the hCRP, TtCRP and the human serum amyloid protein (hSAP), all of which belong to the pentrain family of proteins (Srinivasan et al. 1994) (Figure 1.5C).

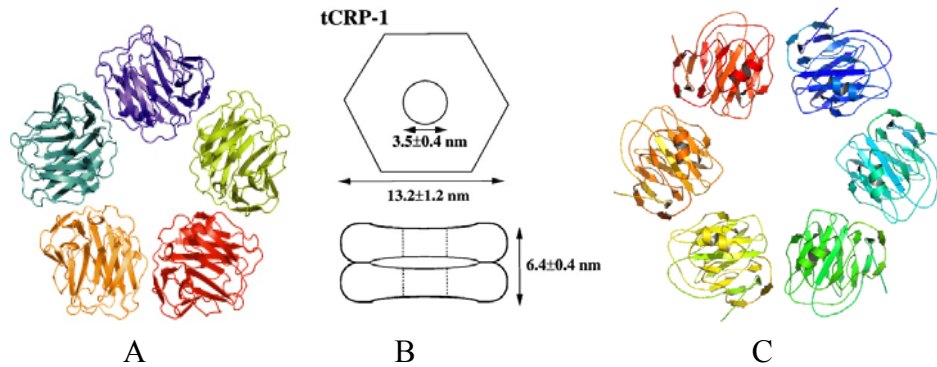


Figure 1.7: Structures of hCRP, TtCRP and LpCRP. (A) hCRP is a 206 amino acid polypeptide with a jelly-roll structure which assembles into a radially symmetric pentamer. (B) TtCRP was identified via electron microscopy to have a doubly-stacked, hexagonal ring-like structure. (C) The hexameric model of LpCRP based on hCRP, hSAP and TtCRP. Structures were adapted from Thompson (1999), Iwaki (1999) and Srinivasan (1994).

The *Carcinoscorpius rotundicauda* horseshoe crab CRP (hereby referred to as CRP) studied in our lab shares 27% homology to hCRP and 91% homology to the TtCRP. Despite not having a high homology to the mammalian protein, sequence analysis showed that the pentraxin region crucial for protein structure, C1q complement-binding sites and the calcium binding sites are conserved (Shrive et al. 1999). CRP was identified to be the predominant LPS-binding protein in the crab hemolymph (Ng et al. 2004). Its transcript level is up-regulated during infection although its protein levels were stably maintained, suggesting that it is one of the proteins involved in immune response against infection.

Unlike CRP which is well studied in the mammalian system, studies on GBP and closely homologous proteins have been limited to the invertebrates. GBP was first discovered in the Japanese horseshoe crab *Tachypleus tridentatus* and was so named

because it was found to bind to Sepharose which is a matrix rich in D-galactose and 3,6-anhydro-L-galactose (Chiou et al. 2000; Chen et al. 2001). Subsequently, it has been referred to as *Tachypleus* Plasma Lectin-1 (TPL-1). The extracellular TPL-1 shows a high homology (65%) to another intracellular protein, Tachylectin-1 (TL-1) (Saito et al. 1995), in the same organism. TL-1 harbours a six β -propeller structure (Figure 1.8) (Iwanaga and Lee 2005). The *Carcinoscorpius rotundicauda* horseshoe crab GBP (hereby referred to as GBP) shows 66% and 96% sequence homology to TL-1 and TPL-1 respectively. We predict that the GBP shares similar structure and properties to TL-1 and TPL-1 due to the high homology shown between their amino acid sequences.

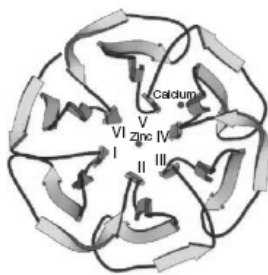


Figure 1.8: Graphical representation of the TL-1 structure. TL-1 is made up of 6 β -propellers (labeled I-VI). Each propeller is made up of 4 β -sheets. Figure adapted from Iwanaga and Lee (2005).

TPL-1 also shows some hints of possible oligomerization, approximately in monomeric and dimeric forms (Figure 1.9A). This could be related to their propensity to form clusters of interlocking molecules, which was suggested to immobilize and entrap the invading microorganisms (Chen et al. 2001). Similarly, our earlier studies have shown clues of this occurrence with GBP when monomers and dimers of GBP were observed in the horseshoe crab hemolymph (Figure 1.9B) (Ng et al. 2007).

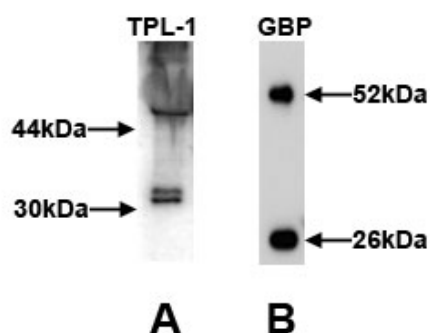


Figure 1.9: Oligomers of TPL-1 and GBP. (A) Purified TPL-1 shows existence in oligomeric form prompting suggestions that it could form clusters to immobilize and entrap bacteria. Figure adapted from Chen et. al. (2001) (B) GBP from the horseshoe crab hemolymph seems to indicate 2 major molecular weights. Figure adapted from Ng et. al. (2007) with modifications.

1.2.2 The beta-propeller structure in proteins

Beta propellers form one of the 6 major protein structural repeat families besides the β -trefoils, armadillo/HEAT, TPR-like, leucine-rich and ankyrin repeats (Andrade et al. 2001). Amongst the all- β folds, the group of β -propeller folds is especially interesting; these structures are modular in nature, have extreme diversity in sequence and function, and are found in organisms with very different phylogenetic origins. The modular nature of this fold is based on a simple building block: a four-stranded anti-parallel β -sheet (Figure 1.10). Because of the twist in the strands of the β -sheet, this modular unit resembles the blade of a propeller, which is the reason for the name given to the fold (Paoli 2001).



Figure 1.10: An example of a beta-propeller fold formed by 4 anti-parallel β -strands. Adapted from Paoli (2001).

The β -propeller architecture is generally observed in multi-domain proteins (Pons et al. 2003). It has been proposed that the β -propeller fold proteins serve as mediators of protein-protein interactions because many of these proteins have been observed to bind other molecules to carry out various functions (Table 1.1). Mutagenesis studies have also revealed that certain amino acids are vital for structural stability while others are involved with its functionality (Jawad and Paoli 2002). This therefore makes β -propellers a good candidate for drug design and protein engineering. Indeed, the separation of regions for function and structure makes it a very flexible protein for ease of manipulation (Pons et al. 2003). The β -propeller domain seems to be flexible

enough for acquiring a different number of blades to insert full domains into circularly arranged β -sheets, as well as to reduce or increase the number of strands into β -sheets (Paoli 2001; Jawad and Paoli 2002).

Table 1.1: The various functions of β -propeller proteins. List of representative solved protein structures with beta-propeller domains together with their function, origin and PDB ID. Adapted from Fulop and Jones(1999).

Number of blades	Name	Function	Sequence motif	'Velcro, closing'*	PDB entry code
Four	Haemopexin†	Haem binding and delivery	Conserved hydrophobic residues	S–S bridge	1hxn, 1qhu
Four	Collagenase	Collagen cleavage	Conserved hydrophobic residues	S–S bridge	1fbl
Five	Tachylectin-2	<i>N</i> -acetylglucosamine and <i>N</i> -acetylgalactosamine binding	DNWL, IGxGGW	3 + 1	1tl2
Six	Neuraminidase and sialidase	Terminal sialic acid cleavage from glycoconjugates	Aspartate box	1 + 3	7nn9, 2sil
Six	Glucose dehydrogenase	Oxidation of sugars to lactons	Two different hydrogen-bonding motifs	1 + 3	1qbi
Seven	Methylamine dehydrogenase	Conversion of methylamine to ammonia and aldehyde	None	3 + 1	1mda
Seven	Galactose oxidase	Oxidation of alcohols to aldehydes	<i>kelch</i> motif	1 + 3	1gof
Seven	G protein β subunit	Signalling	WD repeat	1 + 3	1gg2
Seven	Regulator of chromosome condensation (RCC1)	Nucleocytoplasmic transport and cell-cycle regulation	VYxWG	2 + 2	1a12
Seven	Prolyl oligopeptidase	Peptide bond cleavage	None	None	1qfm
Seven	N-terminal fragment of clathrin heavy chain	Lipid and protein transport between intracellular membrane compartments	Conserved hydrophobic residues	1 + 3	1bpo
Eight	Methanol dehydrogenase	Oxidation of methanol to formaldehyde	W-docking motif	1 + 3	4aah
Eight	Cytochrome <i>cd</i> ₁	Reduction of nitrite to nitric oxide and/or oxygen to water	None	3 + 1	1qks

A propeller structure can be formed from 4 to 8 highly symmetrical individual propellers (Figure 1.11). The propellers form not a cylindrical but rather a cone-like structure, due to how the position of the β -sheets are splayed out. The cavity formed by one end of the cone is mainly observed to be involved in catalysis (Paoli 2001), while the central tunnel is solvent accessible, usually filled with water molecules. Interaction with ligands or other proteins usually takes places between the propeller folds (Figure 1.12).

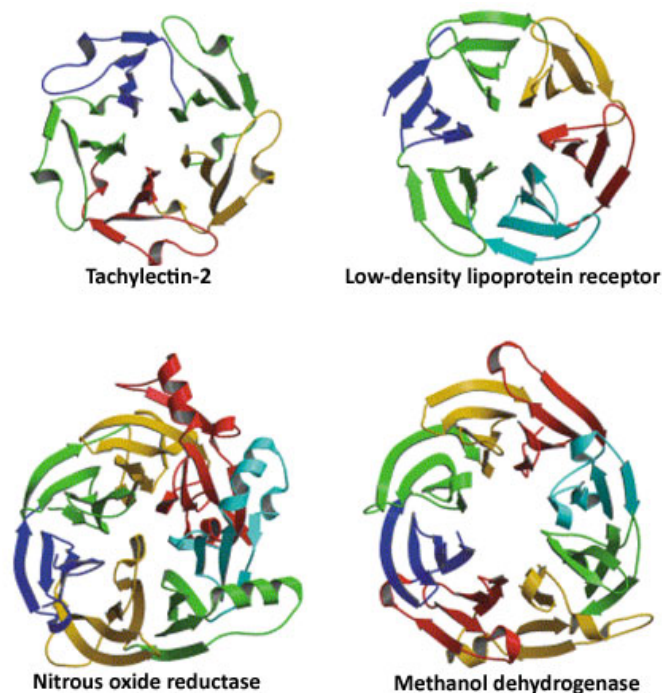


Figure 1.11: Examples of β -propeller proteins with different number of folds. A β -propeller protein can be formed with 4 to 8 propeller domains – Tachylectin-2 (5), lipoprotein receptor (6), nitrous oxide reductase (7) and methanol dehydrogenase (8). The individual domain sequences can be highly homologous or may contain important conserved residues in several key sequence locations. Some proteins, like the nitrous oxide reductase, also contain alpha-helices on top of its basic 7-propeller fold structure. Adapted from Jawad and Paoli (2002).

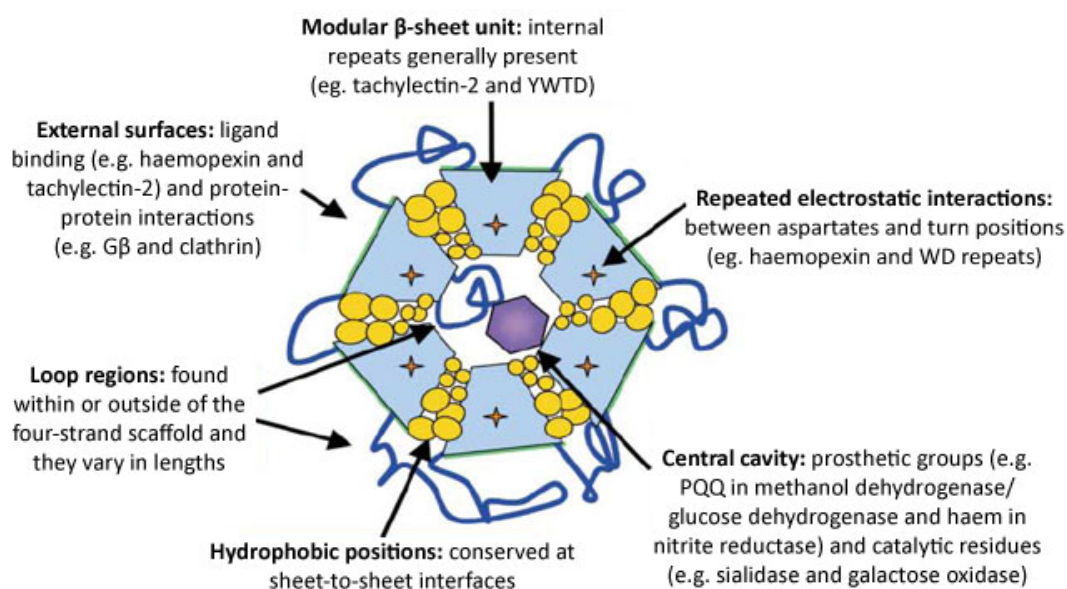


Figure 1.12: Features of the β -propeller structure. Simplistic conceptual diagram summarizing the various properties and features of members of the propeller fold. A six-bladed array is shown as a representative of the β -propeller scaffold. Adapted from Jawad and Paoli (2002) with modifications.

1.2.3 The interactome hypothesis

Our studies have shown that CRP interacts with GBP (Ng et al. 2007). Interestingly, they only interact when the host is infected (Figure 1.13). This suggests that infection primes GBP to bind to CRP or vice versa, and it is postulated that both these proteins form a core of PRR complex to recruit other plasma PRRs to form a "PRR-interactome" to result in an enhanced host response against the invading pathogen.

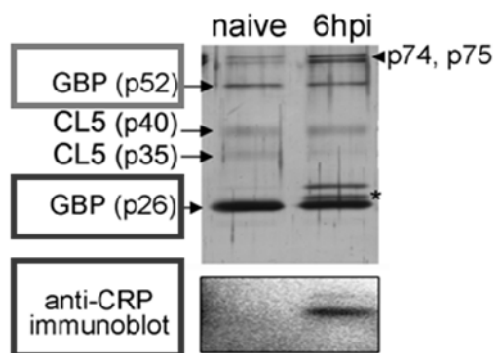


Figure 1.13: CRP interacts with GBP upon host infection. SDS-PAGE analysis show that CRP and CL-5 co-purified with GBP from hemolymph. Western blot using anti-CRP antibody showed that co-purification of CRP and GBP occurred only with 6 hours post infection (hpi) hemolymph. Figure adapted from Ng et. al. (2007) with modifications.

This concept of interactome formation proves to be a recurrent theme, having also been observed in studies involving another horseshoe crab pentraxin, CrOctin (Figure 1.14) (Li et al. 2007). In yeast 2-hybrid interaction studies (Figure 1.15), Le Saux et. al. (2008) also provided strong evidence to suggest that (1) the PRRs act in concert to sense, bind and help eliminate the bacteria and (2) GBP and CRP are recurrent key interactors in the PRR complex formation against pathogen invasion.

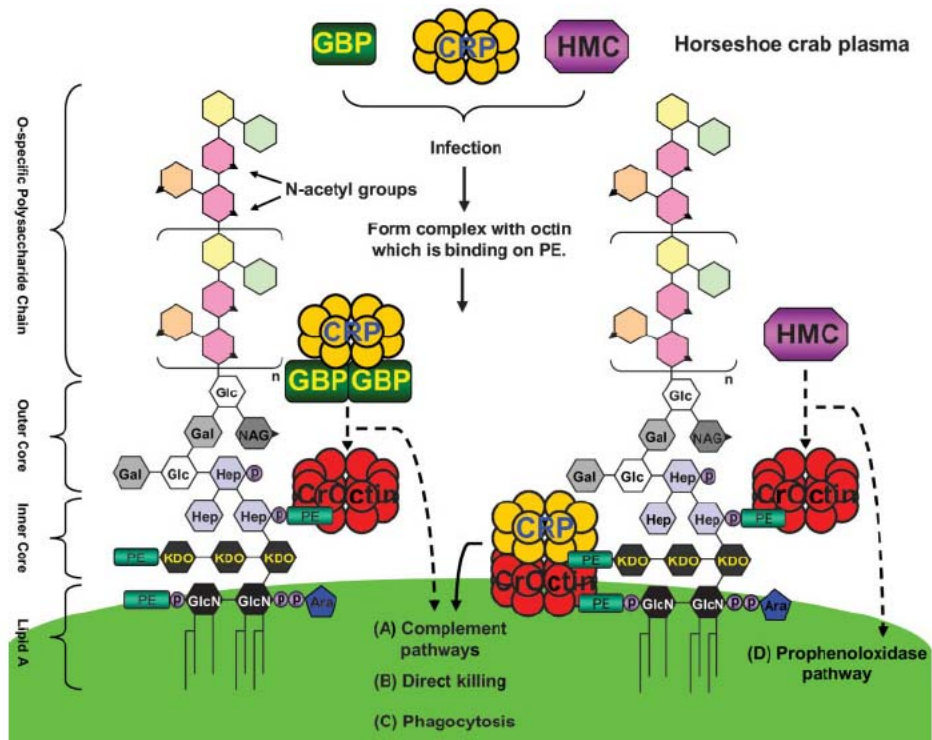


Figure 1.14: A model for pathogen recognition assembly via interaction between CrOctin and other PRRs in the horseshoe crab. Infection triggers CrOctin to anchor on the phosphoethanolamine (PE) groups on the inner core and lipid A (LA) of LPS (bacterial surface in green). Various combinations of PRR interactomes may bind to the invading bacteria via CrOctin's contact with the bacterial PE moieties. Figure adapted from Li et. al. (2007).

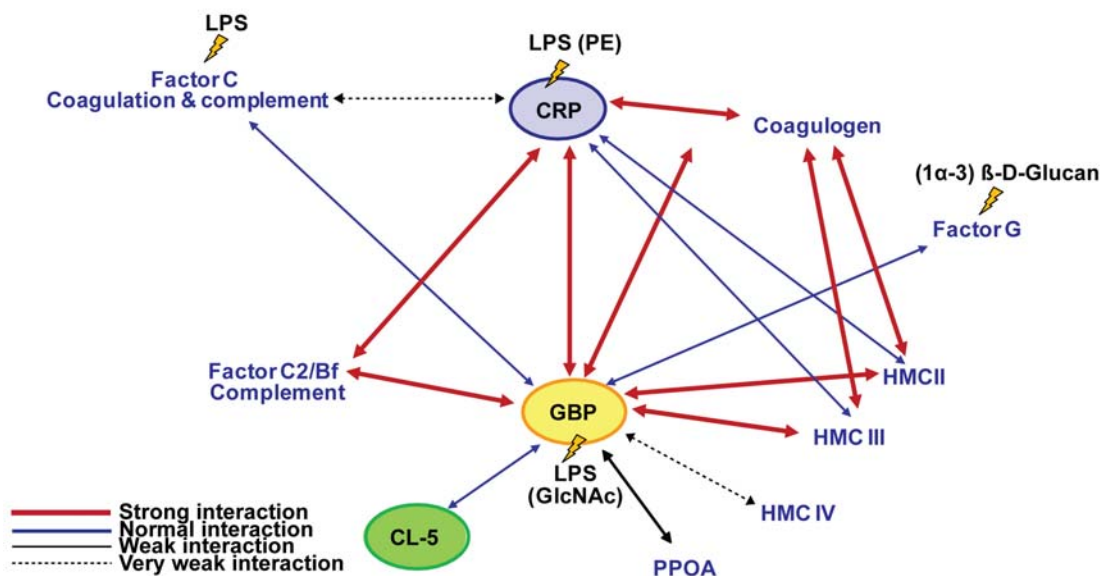


Figure 1.15: Protein interaction network of GBP and CRP. Yeast 2-hybrid analyses showed that GBP and CRP are major interacting proteins in the horseshoe crab hemolymph. Based on this and findings that both GBP and CRP are capable of binding bacteria, it was postulated that both these proteins form a core of PRR complex to recruit other hemolymph PRRs to form a "PRR-interactome" to result in an enhanced host response against the invading pathogen. Figure adapted from Le Saux et. al. (2008).

1.2.4 Pathogen-associated molecular patterns

Earlier in Section 1.1, we defined that the innate immune system detects pathogens via recognition of molecules termed PAMPs that are conserved over a broad range of microbes, and which are not found in the host. The chemical structures of PAMPs are conserved within a particular group of microbes, and the detection of PAMPs can thus indicate the type of microbe to which the immune system activates its most appropriate response. PAMPs of the Gram-positive bacteria include the peptidoglycan and lipoteichoic acid (LTA). LPS from the Gram-negative bacteria is considered to be the prototypical PAMP. PAMPs found in the fungi include carbohydrates such as zymosan, mannan and β -glucan (Roeder et al. 2004). Other molecules that have been reported as PAMPs include the flagellin protein, double-stranded RNA and CpG DNA (Figure 1.16).

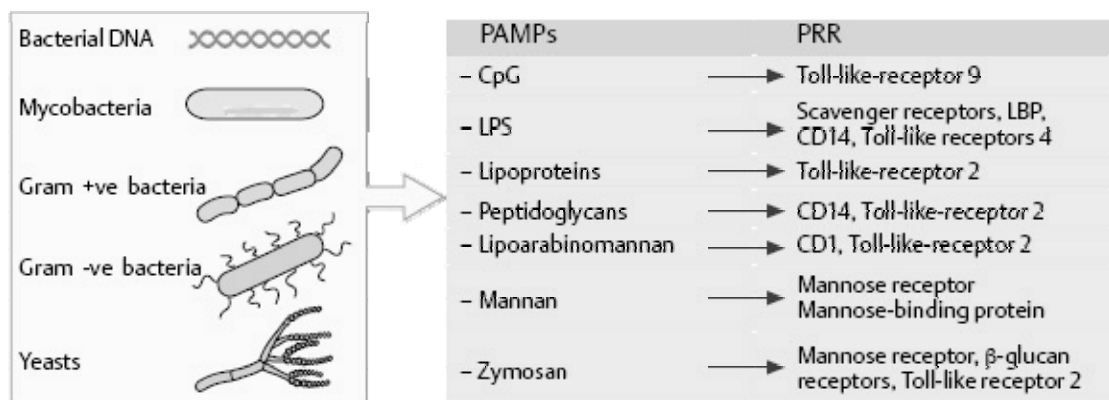


Figure 1.16 : Bacterial PAMPs and their recognition by various PRRs. Bacterial DNA, mycobacteria, Gram-positive and –negative bacteria and yeast are some examples of pathogens whose surfaces display PAMPs that will be detected by host PRRs.

The selective activation of various TLRs by different PAMPs (Roeder et al. 2004; Takeda and Akira 2005) is a good indication that the innate immune is capable of differentiating groups of microbes. One important consideration is that while PAMPs are supposed to be “conserved” across pathogens of one class, PAMPs such as LTA and LPS are not exactly conserved in their complete chemical formula. Rather, they

share a common architecture upon which chemical variations are commonly seen. Such a characteristic may bear important implications for PRRs which target very specific motifs of PAMPs rather than the entire architecture or scaffold of the molecule. This then explains the ability of the invertebrate immune system to use combinations of PRRs like lectin isoforms to recognize and differentiate specific patterns formed by moieties of PAMPs displayed on the surface of different species of pathogens (Beutler 2003; Zhu et al. 2006). In the following section, the LPS and LA will be discussed in greater detail to illustrate the conserved and variable nature of PAMPs.

1.2.4.1 Lipopolysaccharide (LPS)

LPS is a major component of the outer membrane of Gram-negative bacteria (Figure 1.17) such as *Escherichia*, *Salmonella*, and *Pseudomonas*. It is important to the structural integrity of the bacteria and contributes to their ability to cause disease (Tzeng et al. 2002). Deprivation of LPS will cause its death, especially in *E. coli*, due to lowered stability and increased permeability to host factors (Raetz 1990; Onishi et al. 1996). However, there has been preliminary evidence in *Neisseria meningitidis* where mutants lacking LPS appear to be still viable (van der Ley and Steeghs 2003). The mechanism and reason for their viability is still unknown.

LPS is also known as endotoxin and induces a strong response from normal animal immune systems (Hinshaw 1984; Rietschel 1984). Unlike exotoxin, it is not secreted in a soluble form by live bacteria, but it is a structural component in the outer membrane of the Gram-negative bacteria (Figure 1.17) and it is released when bacteria are lysed.

LPS acts as the prototypical endotoxin because it binds the CD14/TLR4/MD2 receptor complex which is present in several immune system cells, including macrophages and dendritic cells (Aderem and Ulevitch 2000; Medzhitov and Janeway 2000). The binding of LPS to the complex triggers the signaling cascade for macrophage/endothelial cells to secrete pro-inflammatory cytokines like tumor necrosis factor (TNF) (Dinarello 1986; Beutler and Cerami 1988; Kiener et al. 1988; Old 1988; Loppnow et al. 1989) and interleukin-1 (IL-1) that could lead to what is called an "endotoxic shock" (Tracey et al. 1986; Morrison and Ryan 1987; Natanson et al. 1989) when there is an exaggerated, systemic response to infection in immune-compromised individuals.

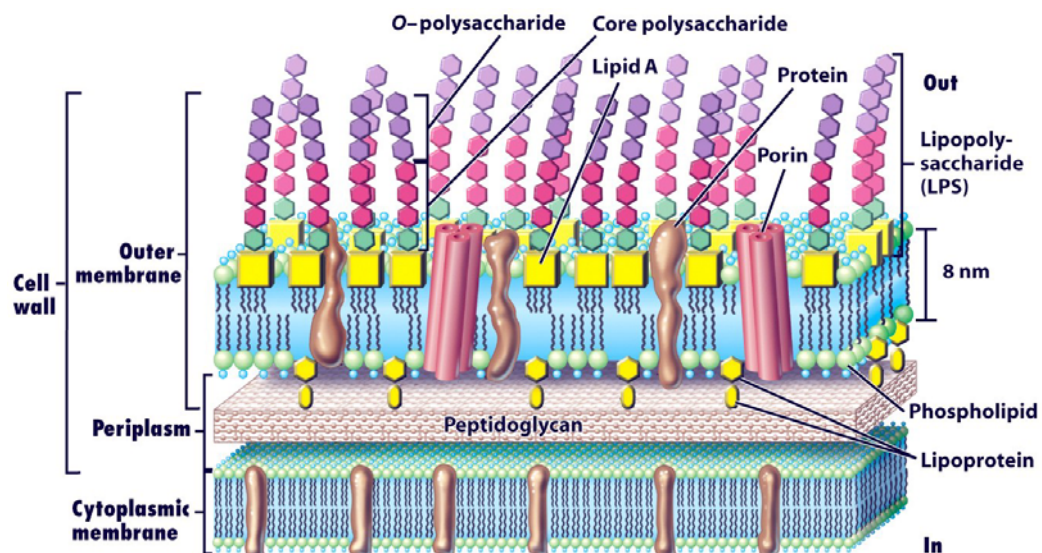


Figure 1.17: Schematic diagram of the cell wall of Gram-negative bacteria. Hexagons in pink and purple depict the O- and core polysaccharide respectively. Green hexagons represent KDO. Yellow squares indicate LA. Figure adapted from Brock Biology of Microorganisms (2006).

LPS is generally constituted by the lipid A moiety (LA) and sugar repeats (the polysaccharide chain) (Nikaido and Vaara 1987) (Figure 1.18). The polysaccharide chain is further divided into the core antigen (R polysaccharide) and the O antigen (O-polysaccharide). Toxicity is associated with the LA and immunogenicity is associated with the polysaccharide chain.

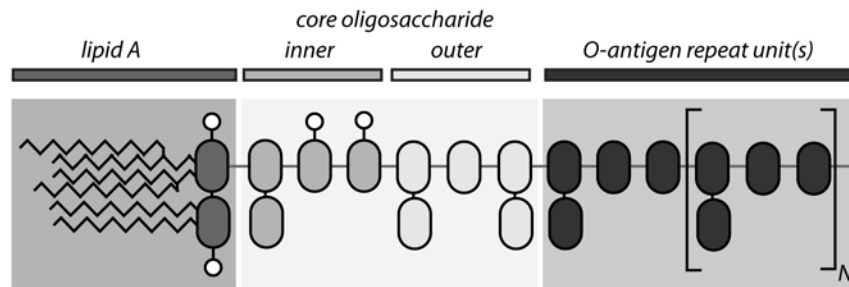


Figure 1.18: Structural organization of LPSs in the Enterobacteriaceae. LPS is usually divided into (1) the highly conserved LA (2) slightly variable core oligosaccharide chain and (3) the highly variable O-antigen saccharide repeat units.

The O-polysaccharide chain is highly variable amongst different bacteria. Among enterobacteriaceae, the LA is virtually constant. With minor variations, the core polysaccharide is common to all members of a bacterial genus (e.g. *Salmonella*), but it is structurally distinct in other Gram-negative bacteria (Reeves and Wang 2002; Patil and Sonti 2004). *Salmonella* and *Escherichia* however, have been shown to have similar cores (Kato et al. 1990).

LPS is widely considered to be the principal component responsible for the induction of septic shock that often accompanies severe infection with gram-negative bacteria. In pharmaceutical production, it is necessary to remove all traces of endotoxin from drugs and drug containers as even small amounts, 0.1 EU/ml of endotoxin, will cause septic shock in humans (Danner et al. 1991; Opal 1995). Components of invading microbes can induce the host to initiate a cascade of events that, if unchecked, can lead to irreversible tissue damage and death (Horn et al. 2000). On the other hand, with the LPS being of crucial importance to the survival of Gram negative bacterial cells, it is therefore a prime target for the design of antimicrobial substances and therapeutic strategies.

1.2.4.2 Lipid A

Lipid A is the lipid component of LPS. It contains the hydrophobic, membrane-anchoring region of the LPS molecule. LA consists of a phosphorylated N-acetylglucosamine (GlcNAc) dimer with 6 or 7 fatty acid chains attached (Figure 1.19). Usually 6 fatty acids chains are found. All the fatty acids in LA are saturated. Some are attached directly to the GlcNAc dimer and others are esterified to the 3-hydroxy fatty acids that are characteristically present. LA of *E. coli* and *Salmonella* is a β ,1-5 linked disaccharide of glucosamine, acylated with R-3-hydroxymyristate at positions 2,3,2' and 3', and phosphorylated at positions 1 and 4'. The envelop of a single *E. coli* cell contains $\sim 2 \times 10^6$ LA residues (Raetz 1986).

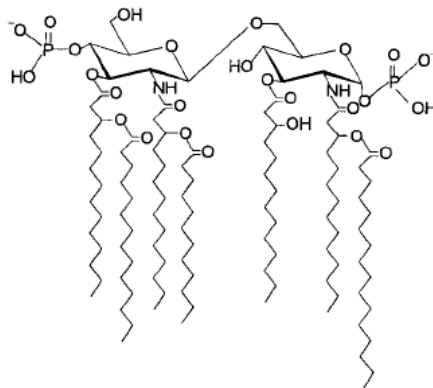


Figure 1.19: Detailed chemical structure of the *Salmonella minnesota* LA. Figure adapted from Tanamoto and Azumi (2000).

The structure of LA is highly conserved among Gram-negative bacteria. While the polysaccharide is non-toxic, LA is the bioactive centre that is responsible for the pathophysiological activities of these bacteria which result in septic shock. The LPS molecule and the bioactive moiety LA, have served as useful targets to design anti-microbial peptides. In insects, an immediate response to LPS by the innate immune system is the production of cationic antibacterial peptides (Lemaitre et al. 1997; Anderson 2000). Such peptides, which can critically detect and eliminate PAMPs, have been proven to be important examples for peptide design. Ideally, the ultimate

goal is to design flexible peptides targeting conserved regions of the PAMP molecule, so that they can act as antibiotics that are resistant to subtle microbial mutations and changes.

1.2.5 Antimicrobial peptides

Antimicrobial peptides (also called host defense peptides) are an evolutionarily conserved component of the innate immune response (Hancock 1999; Gura 2001) and are found among all classes of life. These peptides are potent, broad spectrum antibiotics which demonstrate potential as novel therapeutic agents. Antimicrobial peptides have been demonstrated to kill Gram-positive and Gram-negative bacteria, including strains that are resistant to conventional antibiotics. Examples are mycobacteria (including *Mycobacterium tuberculosis*), enveloped viruses, fungi and even transformed or cancerous cells. Unlike the majority of conventional antibiotics it appears as though antimicrobial peptides also have the ability to enhance immunity by functioning as immunomodulators.

Antimicrobial peptides are generally between 12 and 50 amino acids. They are categorized according to their amino acid composition and structure (Yeaman and Yount 2003). The secondary structures of these molecules follow 4 themes, including i) α -helical, ii) β -stranded, iii) β -hairpin and iv) extended loop (Dhople et al. 2006). Many of these peptides are unstructured in free solution, and fold into their final configuration upon partitioning into biological membranes. Despite differences in structure, most of them share common features like net positive charge and amphiphatic character (Oren and Shai 1998; Hancock 1999) and probably a similar mode of action as well.

Antimicrobial peptides act by permeating the bacterial membrane through several possible mechanisms : ‘carpet-like’ (Huang 2000) , ‘barrel-stave’ or ‘toroidal-pore’ (Oren and Shai 1998) (Figure 1.20) mechanisms or penetration into the cell to bind intracellular molecules which are crucial to bacterial survival (Brogden 2005).

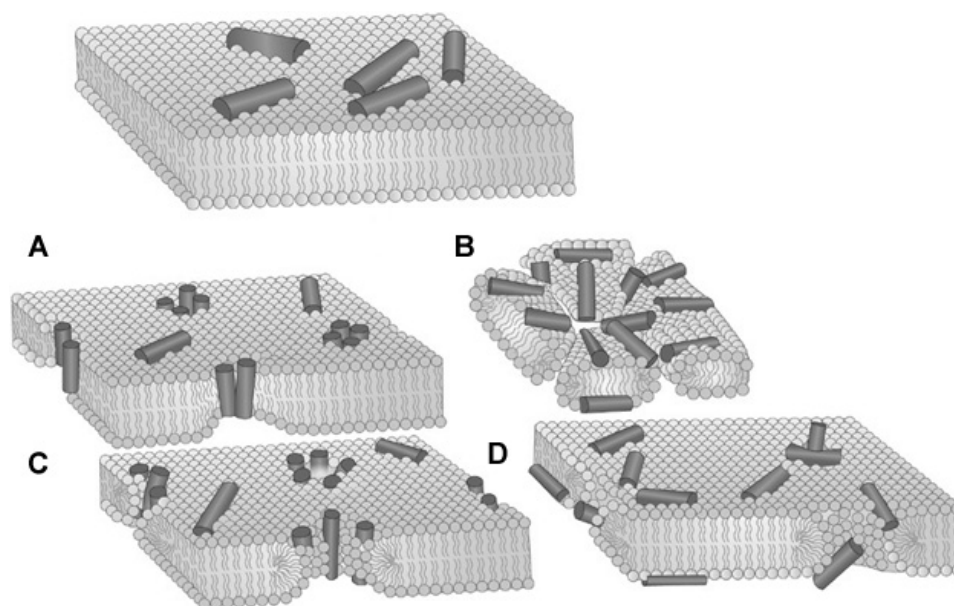


Figure 1.20: The different mechanisms of action of antimicrobial peptides. Antimicrobial peptides (AMPs) have been proposed to disrupt the membrane bilayer during binding by different processes. All these models require, either explicitly or implicitly, that a threshold concentration in the membrane be crossed for disruption to occur. **(A)** Barrel stave. **(B)** Carpet-like mechanism. **(C)** Toroidal pore. **(D)** Disordered toroidal pore. Figure adapted from Melo et. al. (2009).

The horseshoe crab is a rich source of antimicrobial peptides. Tachyplesins I and II, polyphemusins I and II have been found in the hemocytes of the horseshoe crab. They inhibited the growth of not only Gram-negative and Gram-positive bacteria but also fungi, such as *Candida albicans* M9 (Miyata et al. 1989). Another group of horseshoe crab-derived peptides, called the Sushi peptides, were rationally designed based on the core LPS-binding region of Factor C (Ding et al. 1993; Tan et al. 2000; Tan et al. 2000). Because it occurs at the initial step of the coagulation cascade, Factor C

functions as a very sensitive and specific biosensor that is capable of detecting picogram to nanogram levels of LPS (Ho 1983). Two 34-mer Sushi peptides, S1 and S3, display detergent-like properties in disrupting LPS aggregates (Wang et al. 2002; Li et al. 2004; Li et al. 2006).

Based on understanding of the amino acid sequence in the Sushi domains, and comparison of LPS-binding motifs of several other LBPs, Frecer et. al. used computer-aided molecular modeling to show that a predominance of lysine and arginine residues occurs in alternation with hydrophobic residues, *BHB(P)HB* (B=basic; H=hydrophobic; P=polar) (Frecher et al. 2000). Thus, two 34-amino acid sequences containing *BHB(P)HB* residues were found within the Sushi 1 and Sushi 3 domains of Factor C. Interestingly, following such a rational approach, various synthetic peptides containing *BHB(P)HB* motifs have been further tested and showed LPS-binding and neutralizing capabilities similar to those of Sushi peptides (Pristovsek and Kidric 2004).

1.3 Horseshoe crab as an ideal experimental model host for innate immunity study



Figure 1.21: The horseshoe crab

Despite the absence of acquired immune systems, the arthropod hemolymph shows a high degree of specificity against invading pathogens (Iwanaga 2002). Four extant species of the horseshoe crab (Figure 1.21) inhabit the world - *Limulus polyphemus* along the east coast of North America and three species in Southeast Asia including the species in this study, *Carcinoscorpius rotundicauda*. The horseshoe crab possesses a large amount of blood and sizeable tissues compared to other arthropods, which makes it an ideal experimental model for physiological and molecular manipulations.

One prime example of the usefulness of the horseshoe crab is the development of a very sensitive assay for detecting the presence of endotoxin. The Limulus Amebocyte Lysate (LAL) assay, utilizes blood from the horseshoe crab (Bondar et al. 1979) where very low levels of LPS can cause coagulation of the Limulus lysate due to a powerful amplification through an enzymatic cascade. Owing to its extreme sensitivity to endotoxin, LAL has been used widely in the detection of endotoxin in pharmaceuticals, surgical implants, water and food. The LAL test can detect femtogramme levels of endotoxin (Ho 1983).

Furthermore, it is known to harbour a vast variety of plasma proteins which are sensitive to microbial invaders (Ng et al. 2004; Ding et al. 2005; Zhu et al. 2005; Zhu et al. 2006; Jiang et al. 2007; Li et al. 2007; Ng et al. 2007; Le Saux et al. 2008). Literature evidence supports that the horseshoe crab system also possesses anti-viral properties against the Influenza A virus (Murakami et al. 1991), although the active component was not identified. Our lab has also recently observed potent anti-influenza activity in the hemocytes of the horseshoe crab (unpublished data).

Therefore, this makes the horseshoe crab an ideal experimental model to study innate immunity against bacterial and viral infections.

1.4 Overview of thesis

This thesis attempts to elucidate the underlying mechanisms of action of host proteins in the innate immune system; to show how they sense the incoming bacteria, and suggest the various pathways that the PRRs might take to clear the foreign body. This thesis documents the study of two major pathogen recognition receptors in the horseshoe crab hemolymph – GBP and CRP. We characterized their protein-protein and protein-ligand interactions structurally and kinetically, and showed how infection conditions might mediate these interactions.

We modeled the structure of GBP and studied it as a representative member of the Tectonin family of proteins. We uncovered that GBP was able to distinctly interact with pathogens and host proteins. Furthermore, it was interesting that GBP's six seemingly iterative Tectonin domains were able to simultaneously distinguish self/non-self molecules. This prompted us to further study the prevalence of the Tectonin-domain containing proteins in the vertebrates, and especially so in the human. We found widespread occurrence of Tectonin domain-containing proteins in the human, and amongst many other species. We identified and studied a likely distant GBP Tectonin homolog in the human genome, and named it the hTectonin. Interestingly, we found that hTectonin, in immune-related cell lines, and similar to GBP, might interact with proteins important in the innate immune system. These studies led us to conclude that the Tectonin protein family could be an important and integral part of the immune system which remains to be fully characterized.

CHAPTER 2

MATERIALS AND METHODS

2.1 Computational Analysis

2.1.1 Bioinformatics analysis

Identification of Tectonin proteins

The Tectonin domain containing proteins were identified using domain search on the SMART (Simple Modular Architecture Research Tool) database. A position-specific iterated search using the protein sequence on PSI-Search on the Embl server was performed using GBP as the query sequence. Related sequences were chosen after 2 iterations of PSI-Search. Hits were then put through the SMART prediction server to verify their propensity to form Tectonin domains. Multiple sequence alignment was carried out on the curated list of proteins using Promals3D. A phylogenetic tree was then constructed from sequences showing strong domain alignments using PHYLIP with a bootstrap value of 1000.

2.1.2 Protein homology modeling

This section of work was done in collaboration with and co-supervised by Dr. Vladimir Frecer during a 3 week study in the AREA Science Park, Trieste, Italy under the financial support of the Singapore-MIT Alliance.

In the absence of suitable multiple template proteins in the Protein Data Bank (PDB), the 3D model for GBP was homology modeled using the crystal structure of TL-1 (Beisel, H.G.; personal communications with Dr. Vladimir Frecer). Based on a threshold of >40%, the sequence identity of 66.7% between GBP and TL-1 is deemed

sufficient to build a reliable homology model (Petsko and Dagmar 2004). The Molecular Operating Environment software (MOE, Chemical Computing Group, Canada) was used to build the model. The sequences of GBP and TL-1 were aligned using the Blosum62 substitution matrix. The stochastic model building algorithm of MOE uses an adapted Boltzmann-weighted randomized modeling procedure combined with specialized logic for the proper handling of insertions and deletions. The side chain data was assembled from an extensive rotamer library generated by systematic clustering of high-resolution PDB data. MOE created a collection of 10 independent intermediate homology models, which contain different loop candidates and side chain rotamers. Loops were modeled first, in random order. For each loop, a contact energy function analyzes the list of candidates collected in the segment searching stage, taking into account all atoms already modeled. The intermediate models were scored by the electrostatic solvation energy scoring function, calculated via Generalized Born/Volume Integral (GB/VI) methodology. Using molecular mechanics and AMBER99 (Wang et al. 2000) forcefield with electrostatic reaction field correction, the best scoring model was subjected to energy minimization. The final model was inspected using MOE's Protein Geometry stereochemical quality evaluation tools to confirm that the stereochemistry of the model is reasonably consistent with typical values found in crystal structures. The Ramachandran plot generated by RAMPAGE (Lovell et al. 2003) shows that the outlier residues listed remain close to the boundaries of the permitted Psi-Phi values, which are indicated by the light blue contours (refer to Figures 3.21 & 3.28).

2.1.3 Construction of lipid A and saccharide structures

Saccharides and LA molecules (Freder et al. 2000) were constructed using the Biopolymer module in Insight II (Accelrys, CA). Using molecular mechanics and AMBER99 force field with electrostatic reaction field correction, the structures were subjected to energy minimization.

2.1.4 Protein-protein and protein-ligand docking

Preliminary flexible ligand and rigid receptor docking was performed using MOE Site Finder and Docking modules. The top 30 docking poses generated were ranked, and further energy refinement was performed by the Docking module of Insight II using the AMBER99 forcefield. Receptor residues within a 5 Å radius of the docking center were considered as flexible in the refinement process, during which the lowest energy combination of side chain rotamers was determined in several iterative cycles followed by careful energy minimization of the resulting ligand-receptor complex. The ligand-receptor interaction energy E_{int} was computed as the sum of electrostatic and Van der Waals terms as defined in the AMBER99 forcefield.

The HADDOCK2.0 program (Dominguez et al. 2003; de Vries et al. 2007) was used to generate the 3D models of GBP-CRP heterodimer. Peptide sequences from the HDMS analysis involved in protein dimerization which display more than 30% solvent accessible surface area per residue in the homology 3D models of GBP and CRP calculated by NACCESS (Hubbard and Thornton 1993), were defined as the active residues in the guided docking procedure. A total of 400 rigid body docking structures were generated and 40 highest scoring structures were refined by semi-flexible simulated annealing involving active and passive residues. Passive residues

surround the active ones and are located in a layer defined by 3Å interatomic distance from the active residues. Total molecular mechanics energy of the dimer using forcefield formulations from CNS program (Nilges 1995; Nilges 1996), protein-protein interaction energy and protein desolvation energy were considered as a weighted sum of parameters in an empirical scoring function that was used for the rank-ordering of the generated GBP-CRP dimers. In a reference random docking run all residues of GBP and CRP, which displayed >30% solvent accessible surface area, were considered as active residues and amino acids within 3Å inter-atomic distance were considered as passive residues. Generation, refinement and scoring of the GBP-CRP heterodimer models was computed in the same way as described above.

2.1.5 Identification of LPS-binding motifs

Based on Freceer et. al. (2000) and the definition of the LPS-binding motif, an algorithm was written in MATLAB (MathWorksTM, Inc) to conduct large scale and multiple screens for these motifs in proteins. The algorithm will scan for all possible lengths and permutations of the standard *BHB(P)HB* (B-basic, H-hydrophobic, P-polar) residue motifs. The classification of residue properties is based on those listed in Kyte and Doolittle (1982).

2.1.6 Design and synthesis of LPS-binding peptides

After identifying the protein sequence that contains the *BHB(P)HB* motif, a window of 20-25 residues flanking and inclusive of the LPS-motif were studied based on hydrophilicity and secondary structure to assess its suitability for peptide synthesis. Most importantly, the peptide chosen from this region must be hydrophilic for ease of synthesis and is also a confirmation that this peptide exists in a solvent-accessible

region of the protein. The final peptide sequence was sent for synthesis to Genemed Synthesis Inc. (California, USA), with final purity levels >90% (see Appendix A).

2.2 Preparative Methods

2.2.1 Organisms

Horseshoe crabs, *Carcinoscorpius rotundicauda*, were collected from the Kranji estuary of Singapore. Before treatment, horseshoe crabs were washed to remove mud and barnacles. The animals were kept in tanks with a minimal level of 30% (v/v) sea water/fresh water. For experiments in which hemolymph from only one time point of infection (6 hpi, as studied in Ng et. al. 2007) was compared against the naïve hemolymph, infection was performed on male horseshoe crabs of body mass between 80 g to 120 g so as to minimize weight and gender-bias in variations in the bacterial clearance. Bacteria strains used were *Escherichia coli* TOP10 and *Salmonella minnesota* strain R595 and *Pseudomonas aeruginosa* ATCC 27853.

2.2.2 Biochemicals and enzymes

LPS purified from *Salmonella minnesota* R595 (Re mutant) was from List Biological Laboratories, Inc. (Campbell, CA, USA). HiTrap H-hydroxylsuccinimide-(NHS) activated Sepharose High Performance columns, glutathione Sepharose™ 4B, CNBr-activated Sepharose® 4 Fast Flow slurry, Protein-A Sepharose 4 Fast Flow slurry, were products of GE Healthcare (previously Amersham). Deoxynucleotide triphosphates (dNTPs) were from Promega. The concentration stated for dNTP refers to the concentration of each of the different nucleotide, deoxyadenosine triphosphate (dATP), deoxyguanosine triphosphate (dGTP), deoxycytosine triphosphate (dCTP) and deoxythymidine triphosphate (dTTP). X- α -gal was from BD Biosciences

Clontech. Isopropyl-1-thio- β -D-galactopyranoside (IPTG) was from Invitrogen. Sequencing grade-modified trypsin (Porcine) was from Promega. All restriction enzymes were from New England Biolabs or Fermentas. The horseradish peroxidase substrate ABTS (2,2'-azino-bis[3-ethylbenzthiazoline-6-sulfonic acid) was from Roche. Phenylmethylsulfonylfluoride (PMSF), purpald (4-amino-3-hydrazino-5-mercapto-1,2,4-triazole), bovine serum albumin fraction V were from Sigma.

2.2.3 Medium and agar

Luria Bertani (LB) broth was made with 1 % (w/v) Tryptone (Difco), 0.5 % (w/v) Yeast extract (Difco) and 0.5 % sodium chloride at pH 7.0. For LB agar, 1.5 % (w/v) agar (Difco) was added. Synthetic defined (SD) medium was made with 0.67 % (w/v) Yeast nitrogen base without amino acid (Difco), 2 % (w/v) glucose and appropriate amino acid supplements, at pH 5.8. Depending on the nutritional selection that was used, one of the following supplement was added to the SD medium: 0.074 % (w/v) Trp Drop Out (DO) supplement, 0.069 % (w/v) Leu DO supplement, 0.064 % (w/v) Trp/Leu DO supplement, or 0.06 % (w/v) Trp/Leu/His/Ade DO supplement. For medium whereby adenine was not the drop-out supplement, an additional 0.003 % (w/v) adenine hemisulfate was also included. SD plates contained 2 % (w/v) agar. YPD medium was made with 2 % (w/v) peptone, 1 % (w/v) Yeast extract, 2 % (w/v) glucose at pH 6.5. YPDA consist of the same nutrients as above but with 0.003 % (w/v) adenine hemisulfate. For YPD and YPDA plates, 2 % (w/v) agar was added. For use of antibiotics, ampicillin was added at 50 mg/l medium or agar, and kanamycin was added at 50 mg/l medium or agar, unless otherwise stated. IPTG/X-gal selection plates contained 0.05 % (w/v) IPTG and 0.04 % (w/v) X-gal. For X- α -gal plates, 100 μ l of a 2mg/ml stock was spread per plate.

2.2.4 Collection of cell-free hemolymph

The horseshoe crab hemolymph was collected by partial cardiac puncture under pyrogen free conditions. To prevent contamination of the hemolymph, the site of puncture was swabbed with 70 % ethanol. The hemolymph was clarified from amebocytes by centrifugation at 150 g for 15 min at 4°C. The hemolymph was aliquoted, half of which was added with 1 mM PMSF. PMSF irreversibly inhibits serine proteases by sulfonylation of the serine residue in the active site of the proteases, but does not inhibit metallo-, aspartic- and most cysteine proteases with the exception of papain. The aliquots of hemolymph were then quick-frozen in liquid nitrogen and stored at -80 °C until further use.

2.2.5 Depyrogenation of equipment and buffers

Contamination by LPS was minimized by depyrogenation of all equipment that were used. Glasswares were baked at 200 °C for 2 h. Materials that could not be baked, such as tubings for chromatography were treated with 3 % hydrogen peroxide overnight before rinsing with pyrogen-free water (Baxter Inc.).

2.2.6 Purification of GBP from cell-free hemolymph

The cell-free plasma was incubated overnight at 4 °C with Sepharose CL-6B (GE Healthcare). The columns were pre-equilibrated with initial buffer (10 mM Tris, pH 8.8, 150 mM NaCl) and washed with at least 10 column volumes of the initial buffer until a steady base line of <0.01 at $A_{280\text{nm}}$ was consistently obtained. The column was then eluted with the initial buffer at pH 7.4 containing 0.4 M GlcNAc (Sigma). GlcNAc was removed from the eluted protein by ultrafiltration through 3 kDa

MWCO micropore filters (Amicon) and consequently buffer-exchanged into TBS pH7.4.

2.3 Analytical Methods

2.3.1 SDS-PAGE analysis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was carried out to resolve the purified proteins according to their molecular sizes. Vertical minislab gel containing 12 % separating gel (12 % 29:1 Acrylamide:Bisacrylamide, 375 mM Tris-HCl pH 8.8, 0.1 % (w/v) SDS, 0.1 % (w/v) APS, 0.08 (v/v) TEMED) and 5 % stacking gel (5% (w/v) of 29:1 acrylamide :bisacrylamide mix, 125 mM Tris-HCl pH 6.8, 0.1 % (w/v) SDS, 0.1 % (w/v) APS, 0.125 % (v/v) TEMED) were routinely cast using the Mini-protean II system (BioRad). Protein samples were reduced with SDS-PAGE sample loading buffer (50 mM Tris-HCl, pH 6.8, 2 % SDS, 6 % glycerol, 1 % β -mercaptoethanol and 0.004% bromophenol blue) with 3 min boiling before loading into the gel. Electrophoresis was performed at room temperature in Tris-glycine buffer (25 mM Tris, pH 8.3, 250 mM glycine, and 0.1 % SDS) at a constant current of 25 mA per gel. After electrophoresis, the gel was fixed and stained in Coomassie Blue staining solution (0.25 % Coomassie Brilliant Blue R-250, 40 % methanol, 10 % acetic acid) for 30 min. Background staining was then removed by rinsing the gel in destaining solution (50 % (v/v) methanol, 5 % (v/v) acetic acid).

2.3.2 Western blot

In western blot, the electrophoretic transfer was performed in the Mini TransBlot Electrophoretic Transfer Cell (Biorad). The transfer of proteins was carried out at a

constant voltage of 80V for 2 h at 4°C, after which the membrane was incubated for 2 h at room temperature with 20 ml of blocking buffer containing 5% (w/v) skimmed milk in TBST (TBS containing 0.05 % Tween20, v/v). After the incubation, the membrane was rinsed 3 times in 20 ml TBST. The membrane was then incubated overnight at 4°C with primary antibody at the indicated dilution in TBST containing 2.5 % BSA. Subsequently the membrane was washed 3 times with 20ml TBST to remove unbound antibody. Secondary antibody in TBST with 2.5 % BSA was then added to the membrane and incubated at room temperature with gentle shaking. The membrane was washed 3 times with 20 ml TBST to remove unbound antibody. The antigen-antibody complex was developed by incubating the blot with Supersignal West Pico (Pierce) chemiluminescent substrate.

2.3.3 Mass spectrometry

To identify the protein of interest, the protein bands were excised from SDS-PAGE. Care was taken to prevent keratin contamination by wearing a face mask. The gel pieces were carefully rinsed with milliQ water and further cut into 2mm small to facilitate permeation of solutions during the in-gel digestion. Coomassie stain was removed from the gel pieces by a two-step process of rinsing in 50 mM ammonium bicarbonate/ 50 % (v/v) acetonitrile and dehydrating it with acetonitrile. The two cycles of “rinse and dry” treatment was repeated until the gel appears white and opaque. The gel was dried in a speed-vac to completely remove the acetonitrile before subjecting the in-gel protein to disulfide-reduction using 10 mM dithiothreitol (DTT) in 100 mM ammonium bicarbonate at 57 °C for 1 h followed by S-alkylation with 55 mM iodoacetamide in 100 mM ammonium bicarbonate at room temperature for 1 h. The gel was rinsed three times in 100 mM ammonium bicarbonate, dehydrated with 2 rounds of acetonitrile treatment and then re-rinsed once in 100 mM ammonium

bicarbonate and dehydrated again with 2 rounds of acetonitrile treatment. The gel pieces were centrifuged at 5,000 g for 2 min and the acetonitrile was discarded. Any remaining traces of acetonitrile were removed by applying-speed-vacuum, before in-gel digestion was carried out for 12 h with 12.5 ng/ml trypsin (Promega) in 50 mM ammonium bicarbonate at 37 °C. After overnight in-gel digestion, the protein fragments were extracted from the gel with 20 mM ammonium bicarbonate, followed by 5 % formic acid in 50 % aqueous acetonitrile and with 100 % acetonitrile. The extracts were pooled and the solvent was allowed to vaporise in a speed-vacuum before mass spectrometric analysis.

Matrix-assisted laser desorption ionization-Time of flight (MALDI-TOF) analysis was performed at the Protein and Proteomics Centre (Department of Biological Sciences) using Voyager-DE™ STR Biospectrometry™ Workstation (PerSeptive Biosystems) to obtain the peptide mass fingerprint. The peptide mass fingerprint used to search for a protein match in the database using the Mascot website: <http://www.matrixscience.com>. A match occurs if the protein has a significant number of peptide fragments with the same molecular weight after an *in silico* chemical modification (reduction and alkylation, in our case) and enzyme treatment (trypsin). The probability that a match is random or significant is determined using the molecular weight search score (Mowse score) (Pappin et al., 1993).

2.3.4 ELISA

Salmonella minnesota LPS, ReLPS, LA, LTA (Sigma) and GlcNAc-BSA (Dextra Labs, UK) were incubated overnight in binding buffer on 96-well Polysorp™ (Nunc) microplates. After washing off excess ligands, the unbound sites were blocked with

1% BSA and incubated at 25 °C for 2 h. Serially diluted GBP samples (with or without pre-incubation with GlcNAc) were then added to each well and incubated at 25 °C for 2 h. The rabbit antiserum against GBP was added after washing each well. Subsequently, horseradish peroxidase-linked anti-rabbit IgG antibody was added and incubated for 1 h. Peroxidase enzyme activity was determined after adding ABTS (Amersham, GE Healthcare) as substrate and the OD_{405 nm} was measured.

2.3.5 Yeast 2-hybrid co-transformation assay

Co-transformations of the different bait and prey plasmids into *S. cerevisiae* AH109 strain were performed in accordance to standard protocols. All the fragments and full-length CRP and GBP cDNAs (without their signal sequences) were each fused to the DNA-binding domain of Gal4 in the bait plasmid pGBKT7 (BD Biosciences), or to the activation domain of Gal4 in the prey plasmid pGADT7-Rec (BD Biosciences). Figure 2.1 illustrates the concept of cloning the gene of interest into a bacterial vector that contains the Gal4 binding domain (the bait vector) or the Gal4 activation domain (the prey vector).

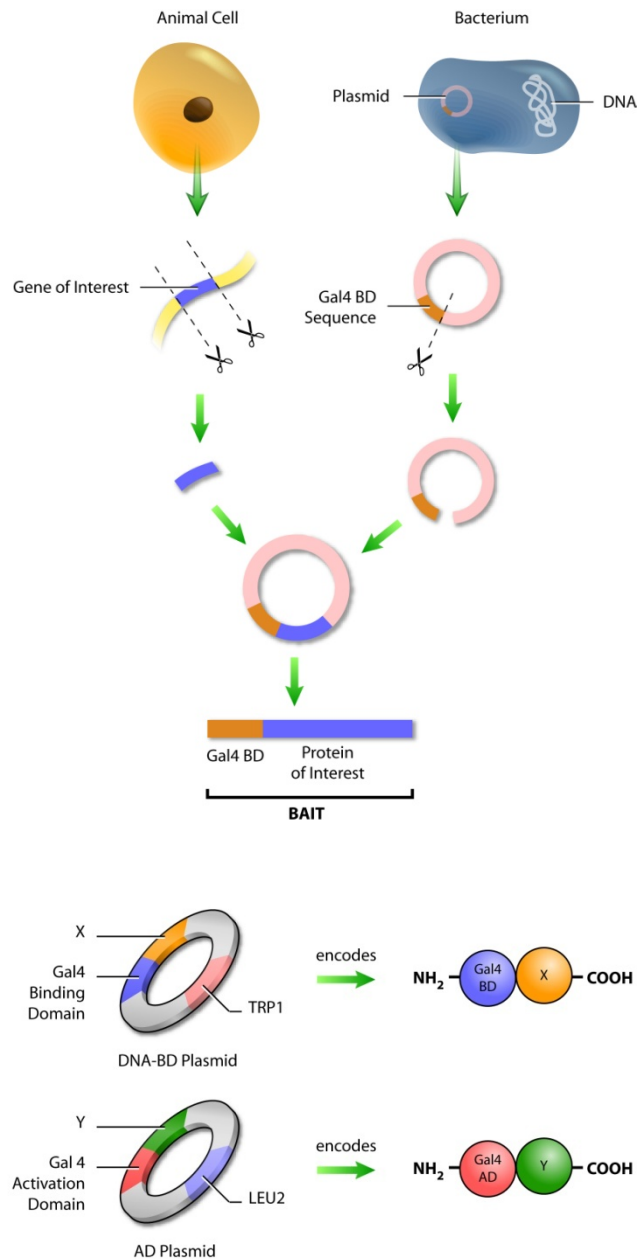


Figure 2.1 : Cloning of genes into yeast 2-hybrid vectors. Fragments and full-length CRP and GBP cDNAs (without their signal sequences) were each fused to the DNA-binding domain of Gal4 in the bait plasmid pGBKT7 or to the activation domain of Gal4 in the prey plasmid pGADT7-Rec. Image adapted from The Science Creative Quarterly (http://www.scq.ubc.ca/?page_id=247).

For selection, synthetic defined (SD) media lacking Leu and Trp (SD-Trp-Leu) or lacking Leu, Trp, His and adenine (QDO medium) were used. Transformants containing bait and prey plasmids were selected on SC-Trp-Leu by incubation for 3.5 days at 30 °C. The resulting colonies were suspended in water and replated on SD-

Trp-Leu and QDO agar at 30 °C for up to a maximum of 7 days. The negative control was co-transformed with a recombinant plasmid and an empty prey or bait plasmid. The positive control was co-transformed with a plasmid expressing the full-length Gal4 transcriptional activator together with the empty pGADT7-Rec vector.

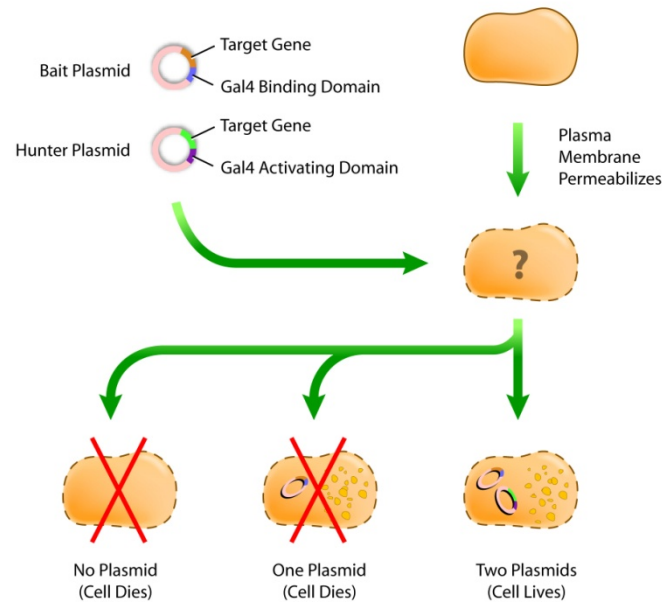


Figure 2.2 : Colony selection of succesful transformants. Only colonies that succesfully incorporated both bait (expressing tryptophan) and prey (hunter, expressing leucine) plasmids will be selected on SD-Leu-Trp media. If there is interaction, the Gal4 binding and activation domain will come together, and histidine and adenine will be expressed as well, leading to growth on QDO selection media. Image adapted from The Science Creative Quarterly (http://www.scq.ubc.ca/?page_id=247).

2.3.6 Yeast 2-hybrid library screening

Amplification of leucocyte cDNA library

Ten ul of the cDNA library was added to 52.5 ml of LB-Ampicilin and 150 ul was plated on each of the 50 LB-Ampicilin plate. The plates were inverted and incubated at 37°C for 18–20 h. Approximately 5 ml of LB/glycerol was added to each plate and colonies were scraped into liquid. All the resuspended colonies were pooled in one flask and mix thoroughly. Subsequent large-scale DNA extraction was done using the Gigaprep[®] kit by Qiagen.

Transformation of hTectonin into AH109 yeast strain

The full-length cDNA of hTectonin in pGBKT7 was small-scale transformed into AH109 yeast strain according to standard protocols. The transformants were then plated on SD-Trp agar plate.

Transformation of cDNA library into AH109 transformed hTectonin

One ml of SD-Trp media was inoculated and shaken at 230rpm overnight at 30°C with several colonies of 2–3 mm each of the yeast strain AH109 transformed with hTectonin. Cells were transferred to a flask containing 150ml SD-Trp and incubated at 30°C for 16–18 h with shaking (250 g) to stationary phase ($OD_{600} > 1.5$). The overnight culture (enough to produce an $OD_{600} = 0.2–0.3$) was transferred into 750 ml of SD-Trp and incubated at 30°C with shaking (230–270 g) until OD was around 0.7. Cells were centrifuged at 1,000 x g for 5 min at room temperature and washed twice in water. Cells were resuspended in 8 ml of TE/LiAc. 0.3 mg of the cDNA library plasmid and 15 mg of herring testes carrier DNA was added to the yeast cells and vortexed to mix. Sixty ml of sterile PEG/LiAc solution was added to the mixture in a 150 ml flask and incubated at 30 °C with shaking (200 g) for 30 min. Seven ml of DMSO was added and mixed gently. The mixture was heat-shocked for 15 mins at 42°C with occasional swirling. The cells were then chilled on ice for 1-2 min, centrifuged at 1000 g for 5 min. The supernatant was removed and the cells resuspended in 10 ml TE. One hundred and fifty ul of this mixture was plated on fifty 150 mm QDO-agar plates each. Transformants were allowed to grow for 4-5 days.

Yeast plasmid extraction

Colonies of co-transformed yeast were first restreaked on QDO-agar plates, and the colonies from these plates were inoculated into 2 ml of SD-Leu-Trp media and incubated at 30°C overnight with shaking (220 g). 1.5 ml of the culture was transferred into an Eppendorf tube and centrifuged at 10,000 g for 1 min. The supernatant was removed and the cells resuspended in 50 µl of TE. Fifty units of lyticase was added (5 units/µl) and the solution incubated at 37°C for 1 h. The subsequent method for DNA extraction was continued from solution S2 of the Axygen DNA Miniprep kit following the standard protocol.

Preparation of electro-competent *E. coli* and the electroporation of plasmids

Five ml of LB media was inoculated with TOP10 *E.coli* and shaken at 230 g overnight at 30°C. Two and a half ml of the overnight culture was transferred into 250 ml of LB media and incubated at 37°C with shaking (220 g) until OD reached 0.7. The transformants were centrifuged at 4000 g for 10 min at 4°C. The supernatant was removed and the cells washed twice in ice-cold distilled water with 10 % glycerol (v/v). The transformants were resuspended in 1 ml distilled water with 10 % glycerol and frozen at -80 °C till use. Two µl of plasmid DNA was mixed with 50 µl of TOP10 electro-competent cells into chilled 2 mm Gene Pulser cuvettes. The transformants were electroporated using Biorad GenePulser Xcell™ at 2.5 kV, 25 µF, 200 Ω. The cells were then mixed with 1 ml of LB and shaken at 37 °C for 1 h before plating on LB-Ampicilin plates.

2.3.7 Dynamic Light Scattering analysis

The particle size of GBP was measured using the principle of dynamic light scattering theory where the shift in light frequency is related to the size of the particles causing the shift (Chu 1991). A 10 μM solution of GBP in TBS was used for measurement with Protein Solutions™ DynaPro™ (Wyatt Technologies Corp.) and the reading was obtained using the DYNAMICS® data collection and analysis software (Wyatt Technologies Corp.).

2.3.8 Protein crystallization

The GBP and CRP solutions were screened for crystallizability with standard kits from Hampton Research and CrystalGen. Specific kit solutions that produced promising crystals were then further modified according to their individual components in the final solution in the grid-like manner to optimize crystallization conditions.

2.3.9 Amide hydrogen exchange mass spectrometry (HDMS) and data analysis

This section of work was done in collaboration with Dr. Ganesh S. Anand, with the help and support from the Protein and Proteomics Centre (PPC) at the National University of Singapore.

HDMS (Sinz 2003; Tsutsui and Wintrode 2007) measures the exchange of amide hydrogens with deuterium when proteins are allowed to exchange with deuterated solvent. Mass spectrometry directly measures increases in mass resulting from the deuterium exchange thereby providing a probe for solvent accessibility. A decrease of mass from the control experiment indicates that a particular site is no longer solvent

accessible. The use of pepsin proteolysis subsequent to the exchange reaction further allows localization of the exchange onto proteolytic fragments of the protein of interest and this provides a powerful tool for mapping protein-protein or protein-ligand interactions.

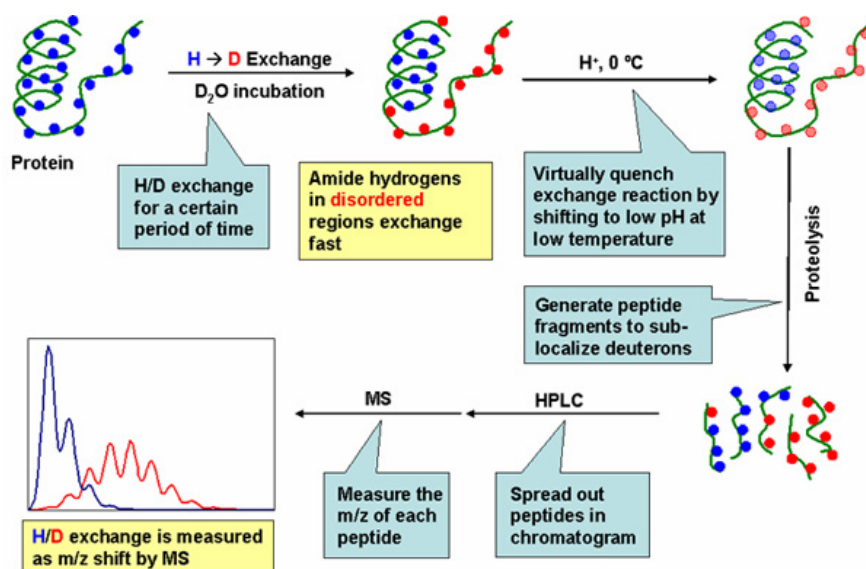


Figure 2.3: Hydrogen exchange mass spectrometry. The concept and experimental procedure of hydrogen-deuterium exchange mass spectrometry. Image adapted from <http://www.exsar.com/Technology/Science.aspx>.

Mixtures containing GBP+GlcNAc, GBP+LA, GBP+CRP, GBP+CRP+LA, and GBP alone as control, were deuterated by diluting 2 µl of protein solution (10 mg/ml, 0.4 mM) in TBS containing 10 mM Tris, 150 mM sodium chloride, pH 7.4 to 18 µl D₂O (Sigma). Solutions were then quenched after timed intervals of 30 s, 1, 2, 5 and 10 min with 180 µl 0.1% trifluoroacetic acid pH 2.5 (TFA, Sigma). Quenched reactions were digested for 5 min with immobilized pepsin (Thermo Scientific). Solutions were spun down for 2 min at 8000 g and 20 µl aliquots of digested solutions were flash-frozen in liquid nitrogen and kept at -80 °C until MALDI TOF MS analysis. Undeuterated samples were included as a negative control. The undeuterated sample was also used to sequence all the pepsin digested fragments by MS/MS sequencing on

the MALDI TOF TOF 4800 instrument (ABI Biosystems). One sample was allowed to exchange with deuterated buffer for 24 h to allow for complete deuteration of solvent exposed regions of the protein. This was used to calculate back exchange under our experimental conditions.

For MALDI TOF MS analysis, the frozen aliquots were thawed and 0.5 μ l was mixed with 0.5 μ l of matrix (15 mg/ml α -cyano-4-hydroxycinnamic acid in 1:1:1 ethanol: acetonitrile: 0.1% TFA, pH 2.5). An aliquot of 0.5 μ l was spotted on the MALDI plate and quickly dried under vacuum and analyzed on a MALDI-TOF mass spectrometer. Spectra were calibrated using Data Explorer (Applied Biosystems) with internal peptide masses 974.51 and 1492.72 and the centroid of the peptide envelopes were measured using Decapp Mass Spec Isotope Analyzer (UCSD, Jeffrey G. Mandell and Elizabeth A. Komives) (Mandell et al. 1998). Back exchange was found to be ~63%, so all centroid values were multiplied by a back exchange factor of 2.67 to calculate the experimental deuterium exchange levels.

A total of 59 pepsin-digested peptide fragments for GBP were generated. These were sequenced by MS/MS sequencing and were found to encompass 92% of the GBP amino acid sequence. Out of these peptides, 20 covering 73% with good signal-to-noise ratio under HDMS experimental conditions were chosen for further analysis, and the extent of deuterium exchange was plotted versus exchange time for GBP, GBP+GlcNAc, GBP+LA, GBP+CRP and GBP+CRP+LA. A total of 70 pepsin-digested peptide fragments for CRP were generated. These were sequenced by MS/MS sequencing and were found to encompass 98% of the CRP amino acid sequence. Out of these peptides, 19 covering 89% with good signal-to-noise ratio

under HDMS experimental conditions were chosen for further analysis, and the extent of deuterium exchange was plotted versus exchange time for CRP, GBP+CRP and GBP+CRP+LA.

2.3.10 Surface plasmon resonance analysis

Real-time biointeraction analyses were performed using a Biacore 2000 instrument (Biacore AB). For GBP-GlcNAc interaction, GlcNAc-BSA (Dextra Labs, UK) was diluted to 10 µg/ml with 10 mM sodium acetate, pH 4.0, and immobilized on the surface of a CM5 sensor chip (Biacore AB) using amine-coupling chemistry according to the manufacturer's specifications. Binding of GBP to immobilized GlcNAc-BSA was measured at a flow rate of 20 µl/min in 10 mM Tris, 150 mM NaCl, pH 7.4. Regeneration of the surfaces was achieved by injecting 20 µl of the running buffer containing 300 mM GlcNAc. For GBP-LPS interaction, LPS, ReLPS and LA from *Salmonella minnesota* (List Biologicals, UK) were diluted to 0.25 mg/ml in 20 mM sodium phosphate, 150 mM NaCl, pH 7.4 and immobilized on the surface of an HPA sensor chip (Biacore AB) according to the manufacturer's specifications. Binding of GBP to the immobilized ligands was measured at a flow rate of 20 µl/min in 10 mM Tris, 150 mM NaCl, pH 7.4. Regeneration of the chip surface was achieved by injection of 20 µl 0.1 M NaOH.

2.3.11 Pyrogene assay for LPS endotoxicity

To test the potential anti-LPS effects of GBP and CRP, the endotoxicity of LPS with and without these proteins were measured using the PyroGene kit (Lonza Inc.). This kit measures the endotoxicity of LPS by using a recombinant Factor C (rFC), which, upon encountering LPS, is activated and hydrolyses a substrate to produce a

fluorescent product, thereby reporting on the endotoxin activity from as low as 0.05 EU/ml. A total volume of 100 μ l, constituted by 50 μ l of LPS at different EU (endotoxin units), 25 μ l of GBP/CRP ranging from 0.01 to 10 μ M and sterile water were added into fluorescence microplate wells (Nunc). The reactions were pre-incubated at 37°C for 10 min. The working reagent was prepared by mixing recombinant Factor C (rFC) enzyme solution, the assay buffer and fluorogenic substrate at a ratio of 1:4:5, respectively. Then, 100 μ l of the working solution was added to the test mixture just before the fluorescence measurement. The reaction was further incubated for 1 h and the fluorescence was measured again. The EU of the test mixture was compared against a standard curve of 0.01-10 EU of LPS.

CHAPTER 3

CHARACTERIZATION OF GBP – A REPRESENTATIVE TECTONIN PROTEIN IN INNATE IMMUNE DEFENSE

In this chapter, GBP is characterized as a representative Tectonin domain-containing protein. GBP has been shown to act as a frontline defense molecule in the Singapore horseshoe crab, *Carcinoscorpius rotundicauda* (Ng et al. 2007; Le Saux et al. 2008). It is able to bind to bacteria and interact with various immune-related proteins. Here, we show how the GBP structure model helped to uncover and explain its significance in forming multiple protein-ligand and protein-protein interactions. We will also show how infection conditions can modify these interactions, and isolate residues and surface contact regions in GBP's host-pathogen recognition. Overall, these structural and kinetic insights help us define the molecular mechanism of action of GBP with other host proteins and with the pathogen. This forms the basis of understanding host-pathogen recognition in Tectonin domain containing proteins.

3.1 Introduction

3.1.1 Tectonin domains in beta-propeller repeats

Despite large variations in their protein sequence, proteins are still able to form the β -propeller fold. Generally, the detection and screening of β -propeller candidates involves a combination of methods : repeat detection, secondary structure prediction and fold recognition (Pons et al. 2003). Therefore, it has been speculated that new propellers can never be recognized from sequence alone, but there are numerous examples that there are certain characteristic sequence repeats that do fold into β -

propellers, for example, the YWTD (Springer 1998), WD40 (Neer et al. 1994) and Kelch domains (Ju et al. 2000) (Figure 3.1).

The YWTD repeats contain the conserved Tyr-Trp-Thr-Asp motif, discovered in over 60 types of extracellular domains with diverse functions. On the basis of theoretical arguments, sequence and secondary structure analysis, threading methods, and experimental data, these repeats were predicted to assume a compact modular structure with a β -propeller fold built by four sheets. Remarkably, the recent structure determination of the low-density lipoprotein receptor, a YWTD protein, indeed showed a six bladed propeller fold (Jeon et al. 2001). A similar argument has been applied to the Kelch motif, identified in several proteins, such as galactose oxidase (Bork and Doolittle 1994). As the sequence repeats match the modular 7-sheet structure of the galactose oxidase propeller, Bork and Doolittle (1994) proposed that Kelch-like proteins have a propeller fold. The WD repeat domain is the most common repeat detected among known human proteins. It contains approximately 40 amino acids and includes well-conserved Trp (W) and Asp (D) amino acids.

More recently, the Tectonin domain was included into this group of β -propeller repeats, where the Tectonin domain seems to be a slight variation of the WD repeats. It also has several characteristic conserved residues inferred from the protein sequences of mainly the slime mold proteins Tectonin I and Tectonin II (Huh et al. 1998) and also from a horseshoe crab protein, TL-1 (Kawabata and Tsuda 2002).

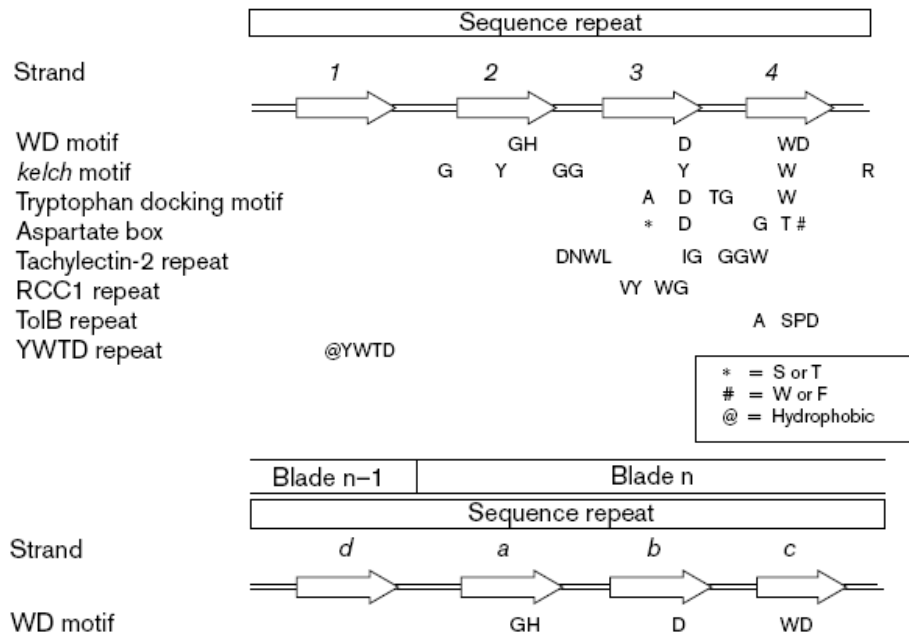


Figure 3.1: Identified conserved sequence motifs in known β -propeller repeats such as the WD-repeat, Kelch, YWTD and others. All are made up of 4 anti-parallel β -strands but have variable amount of conserved residues that defines the particular beta-propeller sequence motif. A single sequence motif repeat can either consist of 4 strands that form a single propeller fold, or 3 strands contributing to a propeller and 1 strand forming the previous propeller fold (velcro strap). For example, the WD repeat has 5 signature conserved residues in its 2nd to 4th beta-strand. The 1st beta strand in its sequence repeat is involved in the formation of the adjoining beta-propeller fold. Adapted from Fulop and Jones (1999).

Examples of β -propeller proteins are the heterotrimeric G-protein complex (Figure 3.2) (Pebay-Peyroula et al. 1997) and the influenza viral protein neuraminidase (Bossart-Whitaker et al. 1993) (Figure 3.3). The G-protein has receptor proteins that act like molecular levers, a trio of 'switch domains' that can adopt 'open' or 'closed' conformations, and a β -subunit that resembles a propeller, these proteins may be thought of as molecular 'nanomachines'. The β -propellers in G-proteins are made up of the WD40 repeats, and one hypothesis suggests that this design provides up to seven different faces on the protein to interact with target ligands while the alpha and gamma subunits of the G-protein complex function as anchors. They also act as "molecular switches," alternating between an inactive and active state, ultimately going on to regulate downstream cell processes (Lodish et al. 1999).

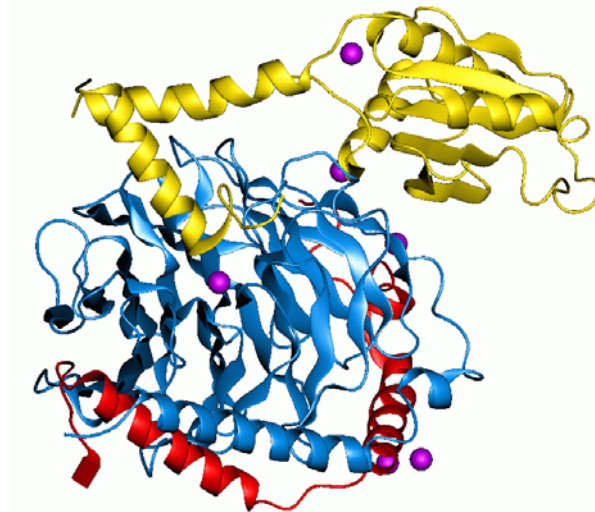


Figure 3.2: The 3D structure of the heterotrimeric G-protein complex. Regions in yellow, blue and red act as different molecular switches. Adapted from Protein Structure and Function, New Science Press Ltd (2009).

The neuraminidase influenza was the first protein structure to reveal a β -propeller fold, with 6-fold pseudosymmetry (Varghese et al. 1983) whose active form is a tetramer. Eukaryotic, bacterial and viral neuraminidases share highly conserved regions of β -sheet motifs. The 3D structure of the β -propeller domain, most used for drug design, is the influenza virus neuraminidase. Using available X-ray crystal structures of sialic acid analogs (Moscona 2005) bound to the active site of the influenza virus neuraminidase has led to the discovery of a series of potent carbocyclic influenza neuraminidase inhibitors like zanamivir and oseltamivir phosphate (Figure 3.4) (Gong et al. 2009; Moscona 2009; Rungrotmongkol et al. 2009). They have emerged as promising antivirals for the treatment and prophylaxis of human influenza infection.



Figure 3.3: The 3D structure of neuraminidase. Neuraminidase here is bound with an inhibitor, 2-deoxy-2,3-dehydro-N-acetylneuraminic acid (DANA), a sialic acid analogue, bound to the central cavity of the β -propeller structure. Adapted from Protein Structure and Function, New Science Press Ltd (2009).

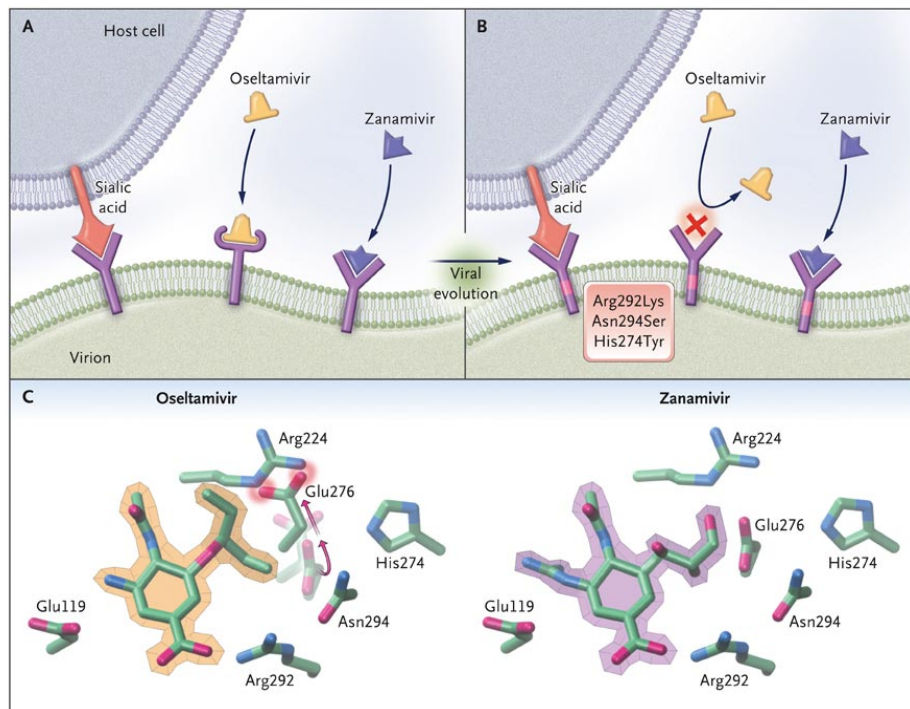


Figure 3.4: Sialic acid structural analogues and their mechanism of action. Structures of oseltamivir and zanamivir, two drugs developed as influenza vaccines. Both are designed as sialic acid analogues, to bind neuraminidase's sialic acid receptors and thus preventing the viral neuraminidase from attaching to host cells which display sialic acid. Figure adapted from Moscona (2009).

3.1.2 Lectins

Lectins are proteins that bind carbohydrates with considerable specificity. They are found in a variety of organisms and are involved in numerous cellular processes, such as host pathogen interactions, targeting of proteins within cells and cell-cell

interactions (Sharon 1993). There are two major groups of lectins – group I, typified by the periplasmic binding proteins and some enzymes have buried ligand binding sites and engulf the ligand upon binding. In contrast, proteins in group II have shallow surface binding sites for their ligands, mostly in the form of a minor depression on the surface of the protein. This group is more versatile than the group I, and is where we find the classical lectin families – legume lectins, C-type lectins and galectins, besides toxins and pentraxins (Figure 3.5)

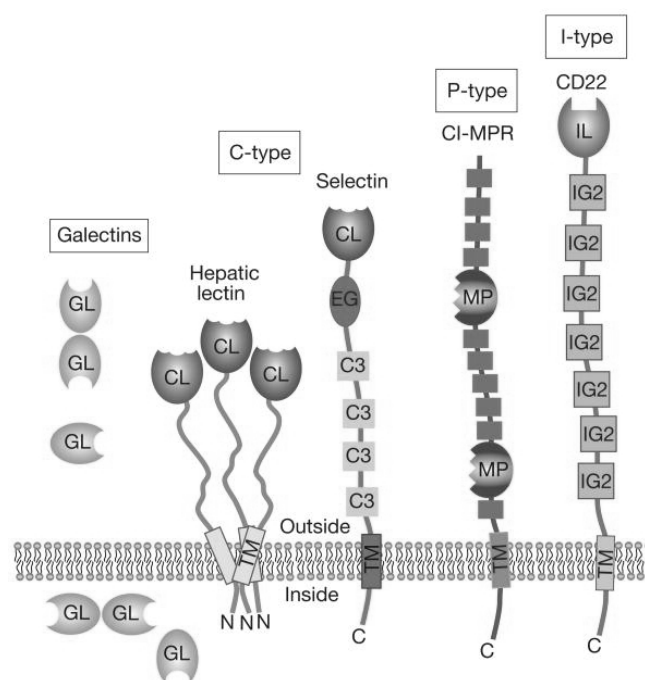


Figure 3.5: Schematic examples of major types of animal lectins, based on protein structure. The emphasis is on the extracellular domain structure and topology. The following are the defined carbohydrate-recognition domains (CRDs) shown: (CL) C-type lectin; (GL) S-type lectin; (MP) P-type lectin; (IL) I-type lectin. Image adapted from *Essentials of Glycobiology* (2009).

Carbohydrates interact with lectins through hydrogen bonds, metal coordination, van der Waals forces and hydrophobic interactions. Despite the overall hydrophilic characteristic of carbohydrates, hydrophobic interactions play a major role in their recognition by lectins. Although, lectins are often multidomain proteins, their sugar-binding activity can usually be ascribed to a single protein module within the lectin

polypeptide designated as a carbohydrate-recognition domain (CRD) (Boraston et al. 2004; Hashimoto 2006).

Analysis of the structural origins of the primary specificity, based on high-resolution structures of protein-carbohydrate complexes representing various families and folds, suggests that although the key interactions responsible for carbohydrate recognition are common, each family has evolved a unique stereochemistry at the principal combining site in order to discriminate between ligands (Rini 1995; Weis and Drickamer 1996). For example, there are many symmetrical β -propeller proteins and lectins that have been shown to bind multiple saccharides, for each of its β -propeller fold (Figure 3.6). This is perhaps attributed to its structural symmetry.

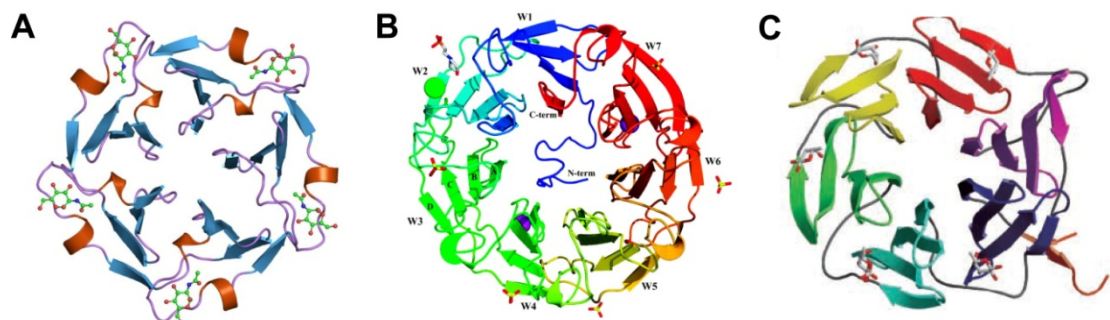


Figure 3.6: Mechanism of specific saccharide binding observed in symmetrical β -propeller fold structures. (A) Tachylectin-2 (TL-2), an immune lectin from *Tachypleus tridentatus*, binds N-acetyl-glucosamine symmetrically on all five of its β -propellers. Figure adapted from Beisel et. al. (1999). (B) *Psathyrella velutina* lectin, an integrin-like fungal protein, a 7 bladed β -propeller protein interacts with GlcNAc in a similarly symmetrical fashion as TL-2. Figure adapted from Cioci et.al. (2006). (C) The 6 propeller *Aleuria aurantia* lectin (AAL) binds fucose within the propeller interface. Figure adapted from Wimmerova et. al. (2003).

It has been noted that the lectin affinity toward monossacharides is low, in the 0.1 to 1.0 mM range (Elgavish and Shaanan 1997). However, lectins often show binding-site dissociation constants in the micro molar range for larger, more complex oligosaccharides. In these cases, the oligosaccharide interacts with secondary sites on

the lectin surface as well as with the primary binding site. Interactions of this type can lead to dissociation constants in the nanomolar range for the appropriate multivalent ligand (Rini 1995; Weis and Drickamer 1996). This is perhaps why we observed that there are many lectins involved in host-pathogen interaction. For example, the outer core of the LPS of Gram-negative bacteria consists of a long chain of sugar molecules which present themselves as multiple targets for the host proteins.

Lectins PRRs – for example, the C-type lectins (CTL) and the TLRs - enable the host to recognize PAMPs, which are mainly via glycolipid structures. C-type lectins contain various CRDs and represent a very heterogeneous group with members such as the macrophage mannose receptor (MMR) and langerin. The MMR is the best-characterized PRR to date (Wileman et al. 1986). Initially identified in macrophages, it is involved in phagocytosis and endocytosis and can recognize mannose, fucose, glucose and N-acetylglucosamine by means of a series of carbohydrate domains. It is able to recognize a wide range of bacteria, fungi and parasites through glycolipid PAMPs. It plays an important role in host defence against fungal pathogens and is involved in glycoprotein clearance.

In addition to CTLs, TLRs do not seem to mediate the uptake of PAMPs but rather stimulate an intracellular signalling cascade. TLRs are expressed on the surfaces of a variety of cells, including epithelial cells, dendritic cells, monocytes and macrophages. They play a major role in innate immunity to microbial pathogens and are the subject of intensive study. Interaction of TLRs with their corresponding PAMPs initiates a rapid cascade of events leading to production of reactive oxygen intermediates, cytokines and chemokines, and promotes the inflammatory response.

3.1.3 The Tectonin domain

The Tectonin domains were first found in the Tectonins I and II proteins of the slime mold, *Physarum polycephalum* (Huh et al. 1998) (Figure 3.7). The *Physarum* Tectonins I and II are expressed on the cell surface of the plasmodium and are involved in the formation of a signaling complex during phagocytosis. The term Tectonin may have come from the oceans' Tectonic plates as the slime mold employs a sliding motion to move on surfaces, similar to the Tectonic plate movement under the sea.



Figure 3.7: Plasmodium stage of *Physarum polycephalum*. During their non-reproductive stages the plasmodial slime molds are thin streaming masses of protoplasm that creep along in an amoeboid fashion sensing food sources, including bacteria. They resemble a moving mass of slime. Figure adapted from <http://www.southernbiological.com/>.

The definition of the Tectonin domain is based on a combination of conserved residues from Tectonin I (25 kDa) and Tectonin II (39 kDa) and further refined using the sequence of TL-1 which is 33% similar to both of the Tectonins (Huh et al. 1998). The deduced amino acid sequences for the Tectonins show that Tectonin I and the C-terminal two-thirds of Tectonin II are 73% identical and are comprised of six similar repeats that vary from 33 to 37 residues in length (Figure 3.8). The Tectonins appear to have diverged from each other after the set of six repeats was established because the sequences are more conserved between corresponding repeats of the two

Tectonins than between repeats within the individual proteins. There have been reports of several other Tectonin-like proteins in the horseshoe crab TL-1 (Kawabata and Tsuda 2002), sponge lectin (Schroder et al. 2003), carp egg glycoprotein (Galliano et al. 2003), and a putative protein in the *Drosophila* (Ponting et al. 2001).



Figure 3.8: Consensus sequence for the definition of a Tectonin domain. Polar residues are represented by p and nonpolar by n; Q represents positively charged residues. The repeats are joined by linkers of 4–7 amino acids. Figure adapted from Huh et. al. (1998).

Because the *Physarum* feeds on bacteria, it has been suggested that the Tectonin domains recognize PAMPs like LPS. A survey on other Tectonin and β -propeller proteins in invertebrates showed that they share certain common features (Schroder et al. 2003). They seem to possess antimicrobial properties demonstrated by the ability to neutralize the LPS of Gram-negative bacteria. However, whether the Tectonin domains can directly bind to PAMPs such as LPS has not been demonstrated experimentally. This opens a wide range of avenues for experimentation: how do the Tectonins detect the pathogen? Do they collaborate with other host proteins for action? Ultimately, do they exist and are conserved in the higher species? We attempt to uncover the potential function of the Tectonins by using GBP as our representative Tectonin protein.

3.2 Results and Discussion

3.2.1 Biochemical properties of GBP

3.2.1.1 Purified GBP shows polymeric forms

The galactose-binding protein was named as such due to its ability to bind Sepharose, which is made up of repeating galactose units. However, it does not only bind

galactose, but is known to bind to several other similar saccharides as well (Chiou et al. 2000; Chen et al. 2001), for example, glucose and GlcNAc. Therefore, GBP is likely to detect structurally similar sugars and associate with them in a similar mechanism.

GBP was isolated by first running the crude hemolymph of the horseshoe crab through a column of Sepharose beads. At the physiological pH of 7.4, with a buffer of 10 mM Tris and 150 mM NaCl, GBP was co-purified with several other key PRRs like HMC and CL-5 (Figure 3.9) which were not particularly known to bind Sepharose. Thus we envisage that at physiological pH, GBP is associated with CL-5 (Zhu et al. 2006; Ng et al. 2007) and HMC (Jiang et al. 2007; Le Saux et al. 2008), and perhaps this is the same phenomenon that exists in the circulating hemolymph *in vivo*. Indeed, there has been evidence via yeast 2-hybrid studies that CL-5 and HMC do interact with GBP (Le Saux et al. 2008).

However, studies have shown that hemocyanin of the *Limulus polyphemus* is able to dissociate into its subunits at pH 8.8 or above (Brenowitz et al. 1983). We speculated that this dissociation into subunits could affect its binding to GBP or the Sepharose beads, and perhaps CL-5 also acts likewise, resulting in a pure elution of GBP after prior removal of the other proteins. This was indeed the case – purified GBP could be obtained with an elution step using GlcNAc at 0.4M in was used to elute GBP from Sepharose at pH7.4, after an initial washing step with 10 mM Tris and 150 mM NaCl at pH8.8 (Figure 3.10). Overall, approximately 8 mg of GBP protein can be purified from 40 ml of cell-free hemolymph (Table 3.1).

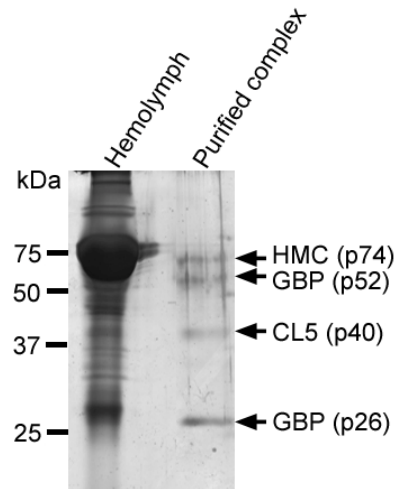


Figure 3.9: GBP is co-purified with partners at pH 7.4. At pH 7.4, GBP was isolated from the horseshoe crab hemolymph as part of a complex together with HMC and CL-5.

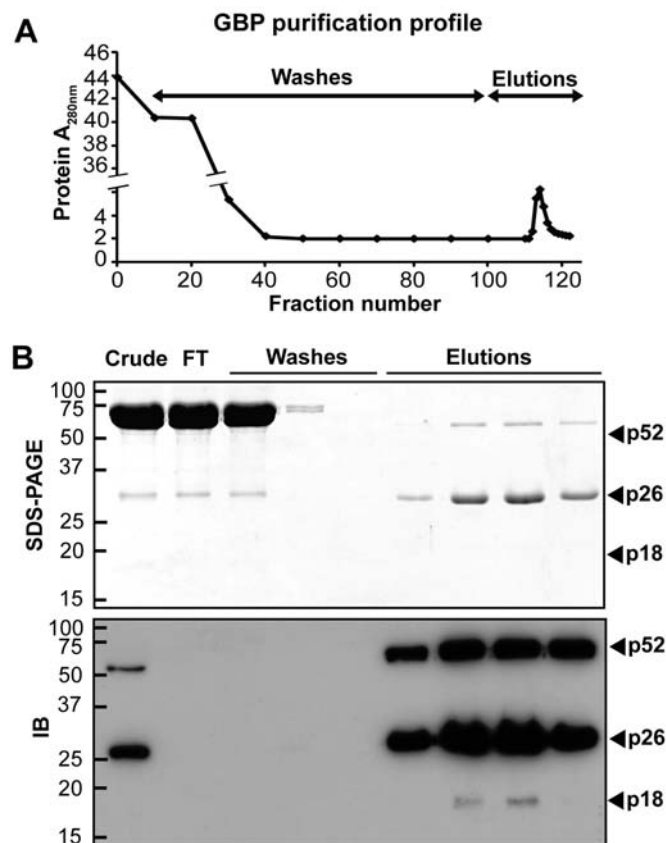


Figure 3.10: GBP is purified at pH 8.8. (A) Horseshoe crab plasma was applied to an affinity column using Sepharose CL-6B as a matrix. The column was equilibrated with 10 mM Tris.Cl pH 8.8, 150 mM NaCl and extensively washed with the same buffer until the washes show no trace of remnant proteins. GBP was eluted with the same buffer containing 0.4 M GlcNAc at pH 7.4. Fractions 112-120 were pooled for ultrafiltration-dialysis to remove GlcNAc and used for further studies. (B) Proteins in the crude extract, flowthrough (FT), washes and representative eluted fractions were resolved by 12% SDS-PAGE and stained with Coomassie Brilliant Blue, and immunoblotted (IB) with GBP antibody.

Table 3.1: Purification of GBP

Step	Volume (ml)	Protein conc. (mg/ml)	Total protein (mg)	Yield (%)
Crude plasma	40.0	48.9	1954.5	100.00
Flowthrough	39.0	45.4	1770.3	-
Sepharose CL-6B	4.20	2.0	8.3	0.4
Microcon (MWCO 3kDa)	0.2	33.0	7.9	0.4

On SDS-PAGE under non-reducing conditions, the purified GBP exists in large molecular sizes in polymeric form (Figure 3.14A, B; lane 'NR'). The higher molecular weight bands are similar to those observed in TPL-1 (Chen et al. 2001). Under reducing conditions, Western blot showed two strong bands and one faint band (Figure 3.11A, B; lane 'R') at 52kDa, 26kDa and 18kDa. These bands were confirmed by mass spectrometry to be the non-reducible dimer, monomer and N-terminal domain of GBP respectively (Figure 3.12). Even though the sequence of GBP has 9 cysteines residues (6 disulphide-linked and 3 free), the 52-kDa GBP dimer was not reducible with β -mercaptoethanol. Neither was it susceptible to boiling. We suggest that a gene encoding for the GBP dimer probably exist in the horseshoe crab.

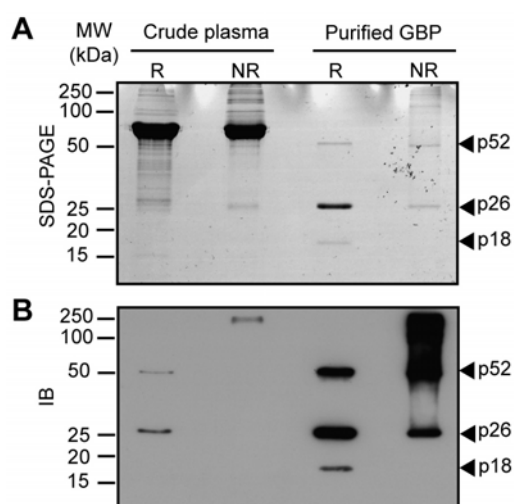


Figure 3.11: Electrophoretic analyses of GBP. Crude plasma and purified GBP were separated by SDS-PAGE with or without reducing agent. The proteins in the gel were either stained with (A) Coomassie Blue or (B) transferred to a membrane, immunoblotted (IB) and detected with anti-GBP antibody. R, reducing condition. NR, non-reducing condition. Molecular weight (MW) markers and different sizes of GBP are shown in kDa.

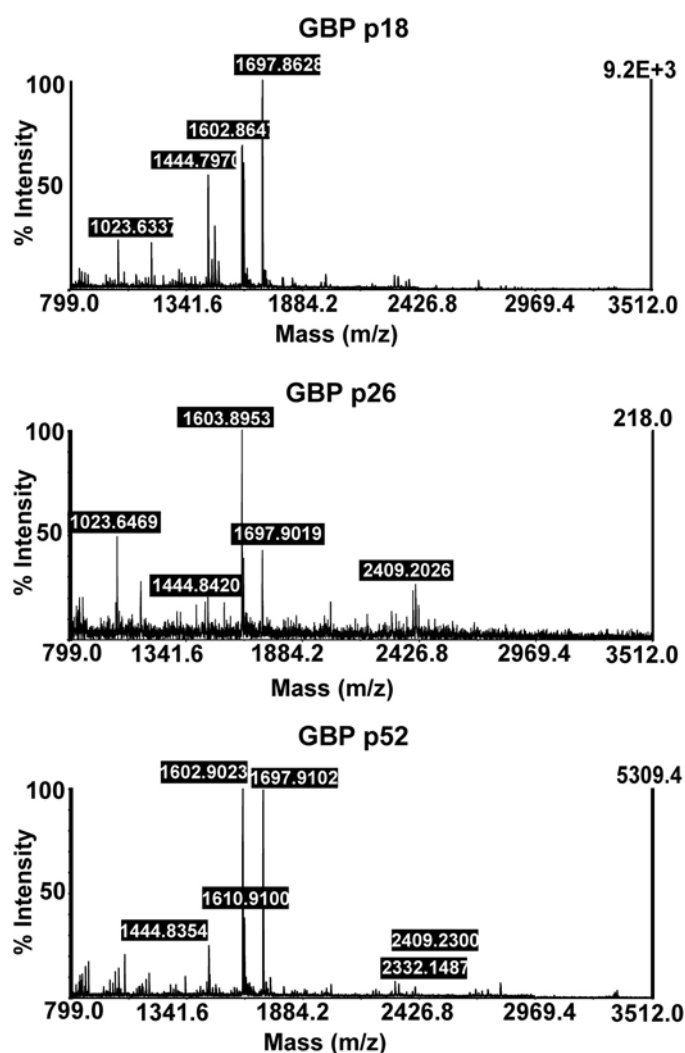


Figure 3.12: Mass spectrometry analysis of purified GBP bands. Matrix-assisted laser desorption ionization-Time of Flight (MALDI-TOF) spectra identified the purified 52 kDa, 26 kDa and 18 kDa protein bands as the dimer, monomer and N-terminal fragment of GBP.

3.2.1.2 GBP is a multimeric complex in solution

In the previous section, we showed that under non-reducing condition, GBP existed predominantly in larger polymeric forms (Figure 3.12, lane ‘NR’). We utilized dynamic light scattering (DLS) analysis to characterize GBP in solution, and revealed that purified GBP in solution exhibited radii of up to 6 nm, corresponding to an average molecular mass of up to 377 kDa, which is equivalent to approximately 14 to 15 GBP monomers (Figure 3.13). This is consistent with observations of β -propeller domains demonstrated to self-assemble (Yadid and Tawfik 2007) indicating their

propensity to polymerize. These results suggest that GBP tends to polymerize or potentially associate with other proteins, forming larger complexes in the hemolymph.

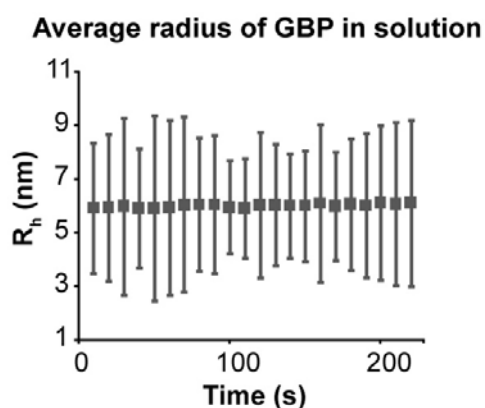


Figure 3.13: GBP exists in polymeric form. DLS analysis of purified GBP indicates prevalence of GBP polymers in solution. The polymers exhibit radii of up to 6 nm, corresponding to an average molecular mass of up to 377 kDa, equivalent to approximately 14 to 15 GBP monomers.

Table 3.2 DLS measurements for GBP of molecular radius, diffusion coefficient and molecular weight in solution over time

Time (s)	Diffusion coefficient ($10^{-9}\text{cm}^2/\text{s}$)	Radius (nm)	Molecular weight (kDa)
10	243.1	5.91	361.5
20	242.6	5.92	363.5
30	240.6	5.97	372.7
40	243.6	5.90	359.1
50	243.5	5.90	359.8
60	242.6	5.93	363.9
70	238.4	6.03	383.1
80	237.9	6.04	385.8
90	238.1	6.04	384.7
100	241.8	5.94	367.4
110	243.9	5.89	357.8
120	238.4	6.03	383.3
130	238.8	6.02	381.2
140	239.8	5.99	376.7
150	239.8	5.99	376.5
160	236.6	6.07	391.9
170	240.6	5.97	372.9
180	237.5	6.05	387.6
190	239.7	6.00	377.3
200	235.5	6.10	397.5
210	237.2	6.06	389.2
220	236.1	6.09	394.6

3.2.1.3 Purified GBP retains saccharide-binding function

Since GBP was eluted with GlcNAc from the Sepharose column, we next determined whether the subsequent removal of GlcNAc from the protein, in order to carry out more functional studies, would retain the protein function. By column centrifugation dialysis with a molecular weight cut-off (MWCO) of 3 kDa, the GlcNAc (approximately 0.2 kDa) containing elution buffer was gradually replaced with the base buffer of 10mM Tris, 150mM NaCl at pH 7.4. The optical density (OD₂₂₅) readings indicating presence of GlcNAc in the column filtrate was monitored until it gave a consistent OD of < 0.01 (Figure 3.14). Once the dialysis was completed, a sample amount of GBP was tested if it could re-bind the Sepharose column. Indeed, the purified and dialysed GBP was able to rebind the Sepharose column (Figure 3.15) hence indicating that the ligands from purification have been removed from GBP and the protein is competent for further structure-function analysis. The ability of GlcNAc to be removed from GBP also shows that GlcNAc associates with GBP not by covalent bonding but more likely through electrostatic interactions.

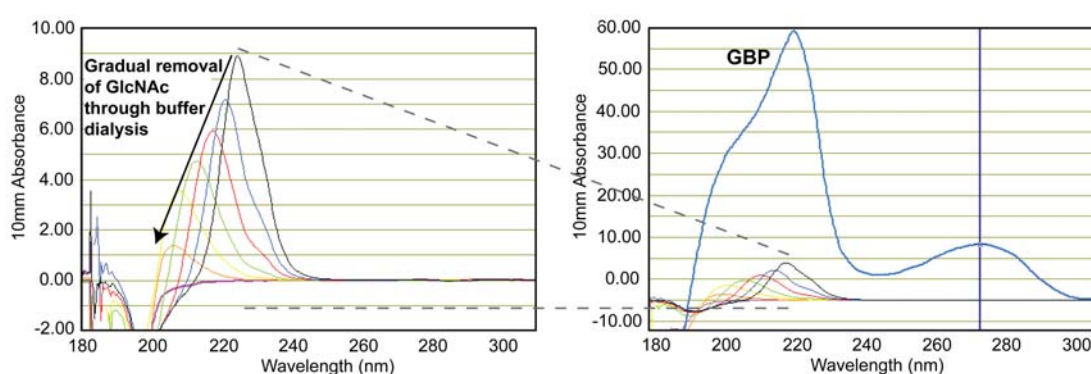


Figure 3.14: OD measurements during GBP purification. (Left panel) GlcNAc in the column centrifugation dialysis filtrate shows an initial profile peak at 225nm which gradually reduces when the buffer is continuously dialysed until baseline GlcNAc OD reading below 0.01 is achieved. This is used as an indicator that GlcNAc has been thoroughly removed from the elution product. **(Right panel)** GBP is then recovered from the retentate of the dialysis column. The GlcNAc peaks from the removal process are overlaid on the graph for comparison.

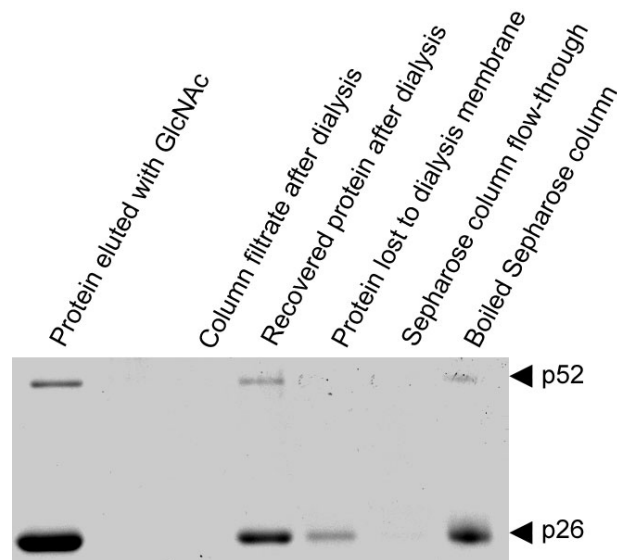


Figure 3.15: GlcNAc-removed GBP is able to re-bind Sepharose. GlcNAc in the GBP elution was removed through dialysis. Majority of the protein could be recovered after the dialysis procedure. A small sample of the dialysed protein was flowed through the Sepharose beads again. The GBP was retained on the Sepharose beads (evidence of the ability to bind Sepharose), and was obtained when beads were boiled.

3.2.2 The GBP structure

Since GBP was observed to interact with CRP under infection conditions (with LPS) (Figure 1.8) (Ng et al. 2007), it was imperative to characterize the structural basis of the interactome formation. There are two possible ways to achieve this: (a) by X-ray crystallographic studies, and (b) through *in silico* predictions coupled with other biophysical methods. Both methods were utilised in our studies, as described in the following sections.

3.2.2.1 Crystallization of GBP and CRP for structure determination

Attempts were made at crystallizing GBP and CRP. The GBP and CRP proteins used were purified from the hemolymph of the horseshoe crab. Kits from Hampton Research and Crystalgen® providing various combinations of crystallization buffer solutions were used (Table 3.3). Both GBP and CRP produced crystals (Figure 3.16) but only those from CRP were viable for X-ray diffraction studies. The CRP crystal

was obtained from buffer #17 (0.2 M Lithium sulfate monohydrate, 0.1 M Tris HCl pH 8.5, 15% w/v PEG 4K) of the Crystal Screen Lite kit. The buffer condition was subsequently optimized to 0.3M LiSO₄, 0.1M, Tris-HCl pH 8.8, 12% PEG 8K to obtain larger crystals. The crystals were then collected and sent for diffraction at the National Synchrotron Light Source, Brookhaven National Laboratory, Upton, New York. Diffraction data was collected at 3.4Å (Table 3.4, Figure 3.17). However, the data was not sufficient for good structure resolution and further optimization of crystallization conditions are required, for example, to (a) grow larger crystals (b) obtain crystals from a different space group by growth under alternate buffer compositions.

Table 3.3 List of crystallization kits used in attempts to crystallize GBP and CRP

Hampton Research	Crystalgen®
Crystal Screen 1	NaMax
Crystal Screen 2	MPDMax
Crystal Screen Lite	MemMax
Crystal Screen Cryo	PhosMax
Index Screen	AsMax
Natrix Screen	CryoMax
Quik Screen	
PEG/Ion Screen	
Grid Screen MPD	
Grid Screen Na Malonate	
Grid Screen NaCl	
Grid Screen PEG	
Grid Screen Ammonium Sulfate	
Lithium Sulfate Screen 1	
Lithium Sulfate Screen 2	
HEPES Screen 1	
HEPES Screen 2	

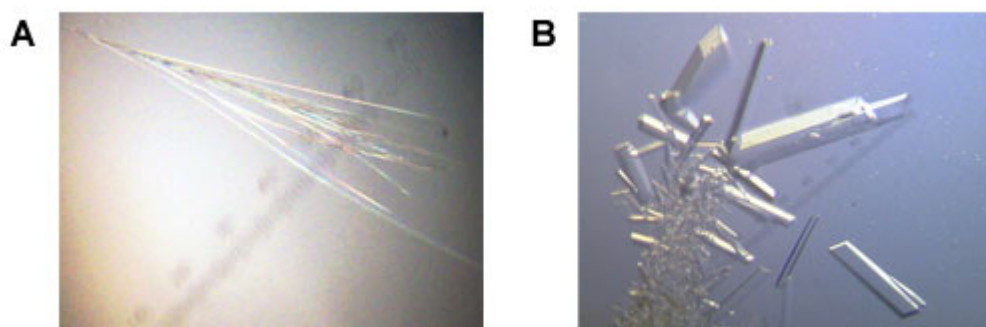


Figure 3.16: Crystals of GBP and CRP. (A) Crystals with needle-like structure were obtained with 5mg/ml GBP using buffer from Crystalgen Phosmax #52 composed of 0.1M Na phosphate pH 7.0, 2M 1,6 hexanediol. Buffer was subsequently optimized to for final crystals. (B) Rectangular cuboid crystals were obtained with 10mg/ml CRP using buffer from Hampton Research Crystal Screen Lite #17 composed of 0.2 M Lithium sulfate monohydrate, 0.1 M Tris HCl pH 8.5, 15% w/v PEG 4K. Buffer was subsequently optimized to 0.3M LiSO₄, 0.1M, Tris-HCl pH 8.8, 12% PEG 8K. Crystals were grown using hanging-drop method with 1:1 ratio of protein solution to crystallization solution.

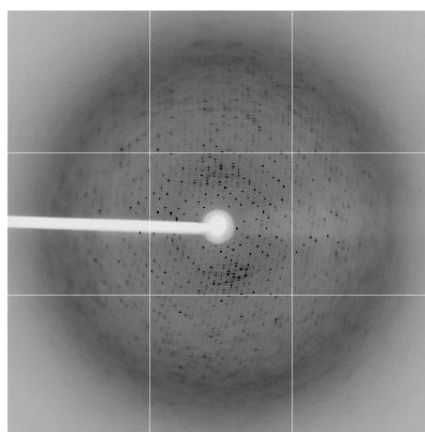


Figure 3.17: X-ray diffraction pattern of CRP crystal. A total of 30 crystals were tested for diffraction. One native data set and two Multiwavelength Anomalous Dispersion (MAD) data sets were collected on the X29 beamline around the Platinum absorption edge at 3.4Å. Diffraction was carried out at the National Synchrotron Light Source (NSLS), Brookhaven National Laboratory, Upton, New York.

Table 3.4 Data obtained from CRP crystal diffraction.

Data collection	
Cell dimensions	Space group P1
a,b,c (Å)	85.145, 121.440, 189.838
α, β, γ (°)	89.311, 81.470, 85.145
Wavelength (Å)	1.1
Resolution (Å)	50~3.5 (3.66~3.5)
R _{sym} (%)	11.1 (46.0)
I/σ (I)	11.2 (2.3)
Completeness (%)	97.3 (93.7)
Redundancy	3.7

3.2.2.2 Computational modeling of GBP and CRP structure

Parallel to crystallization attempts, computational modeling of the GBP and CRP protein structures were carried out. The three dimensional models serve as a basis for further exploring molecular interactions between GBP and CRP, and with their interacting partners or corresponding ligands.

The PSIPRED secondary structure prediction described GBP as a protein made up of β -sheets, in a configuration similar to those observed in both TPL-1 and TL-1 (Figure 3.18). GBP has six tandem repeats (Figure 3.19) and these repetitive sequences are reminiscent of the WD repeats of the β -subunit of G-proteins, suggesting that they fold into a six β -propeller-like structure as well. Therefore, from this information garnered from the amino acid sequence, we can conclude that GBP is likely to form a 6 β -propeller structure.

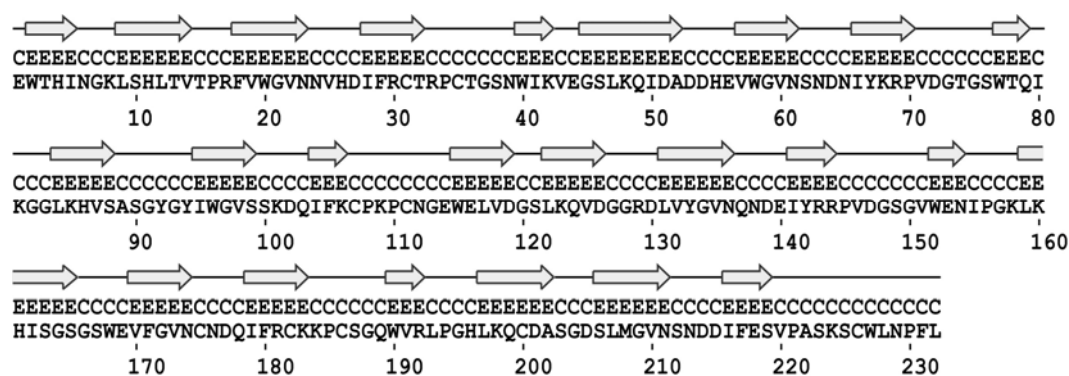


Figure 3.18: Secondary structure prediction of GBP. GBP is predicted via PSIPRED to be of β -strand only conformation.

Domain 1 AE**W**THING**K**LSHLTVTP-RFVW**G**VN NVHDI**F**RC**T**RPCTG
 Domain 2 **S**N**W**IKVE**G**SLKQIDADD-HEVW**G**VN NSNDNIYKR**P**VDTG**T**
 Domain 3 **G**SW**T**Q**I****K**G**G**LKHVSASGYGIW**G**VSSKDQ**I**FK**C**PK**P**CN
 Domain 4 **G**E**W**ELVD**G**SLKQVDGGR-DLVY**G**VN QNDI**Y**RR**P**V**D**GS
 Domain 5 **G**V**W**ENIP**G**K**L**KHISGSGSWEV**F****G**VN**C**NDQ**I**FR**C**KK**P**CS
 Domain 6 **G**Q**W**VRL**P****G**H**L**KQCDASG-D**S**LM**G**VN SNDD**I**FESVP**A**SK

Figure 3.19: The 6 internal tandem repeats in the sequence of GBP. The tandem repeats are similar to those observed in TL-1 (Kawabata and Iwanaga 1999) with several conserved residues at specific locations (in **bold**).

Homology modeling was employed to obtain the GBP model structure. TL-1, which shares 66.7% sequence homology to GBP (Figure 3.20) was taken as the template structure to model GBP (Beisel H-G; personal communications with Dr. Vladimir Frecer, collaborator)). The final model was inspected using stereo-chemical quality evaluation tools (AMBER, RAMPAGE) (Wang et al. 2000; Lovell et al. 2003; Case et al. 2005) to confirm that the model's stereochemistry is reasonably consistent with typical values found in crystal structures. The following Ramachandran plot (Figure 3.21) shows that the outlier residues (red squares) listed in the Table 3.5 remain close to the boundaries of the permitted Psi-Phi values, which are indicated by the lighter contours.

```

Position      10      20      30      40      50      60
TL1           VQWHQIPGKLMHITATPHFLWGVNSNQIYLCRQPCYDGGQWTQISGSLKQVDADDHEVWGVNRRNDIY
GBP           AEWTHINGKLSHLTVTPRFVWGVNNVHDI FRCTR PCTGSNWI KVEGSLKQIDADDHEVWGVNSNDNIY
Conserved    * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Position     70      80      90     100     110     120     130
TL1          KRPVDGSGSWVRVSGKLVKHSASGYGYI WGVNSNDQIYKCPKPCNGAWTQVNGRLKQIDGGQSMVYGV
GBP          KRPVDGTGSWTQIKGGLKHVSASGYGYI WGVSSKDQIFKCPKPCNGEWELVDGSLKQVDGGRDLVYGV
Conserved    * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Position    140      150      160      170      180      190      200
TL1         NSANAIYRRPVDGSGSWQQISGSLKHITGSGLSEVFGVNSNDQIYRCKTKPCSGQWVSLIDGRLKQC DAT
GBP         NQNDEIYRRPVDGSGVWENIPGKLVKHSASGYGYI WGVSSKDQIFRCKKPCSGQWVRLPGHLKQC DAS
Conserved   * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * *

Position    210      220      230
TL1         GNTIVGVNSVDNIYRSG-----
GBP         GDSL MGVNSND DIFESVPASKSCWLNPF L
Conserved   * * * * * * * * *

```

Figure 3.20: Sequence alignment of GBP to TL-1. TL-1 was the protein used as the template for the homology model of GBP. They share 66.7% amino acid sequence similarity.

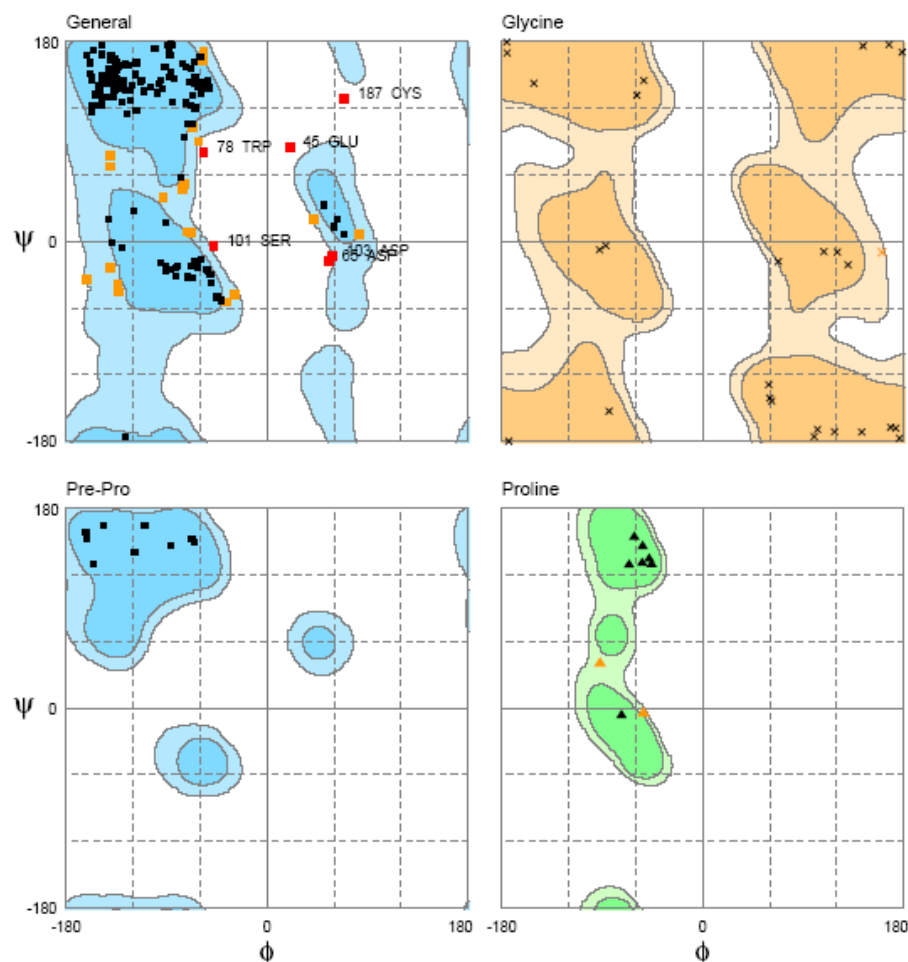


Figure 3.21: The Ramachandran plot of the GBP homology-modeled structure shows that the outlier residues listed (Table 3.5) remain close to the boundaries of the permitted Psi-Phi values, which are highlighted by the light blue contours, indicating that the structure has been reliably modeled. Black, orange and red boxes represent residues in favoured, allowed and outlier regions respectively (refer to Table 3.5 for details).

Table 3.5 List of Phi-Psi outliers in the GBP model.

No.	Residue	Psi	Phi	Score
1	GLU45	84.6	20.2	0
2	ASP65	-17.3	54.9	0.0003
3	TRP78	80.4	-56.8	0.0001
4	SER101	-4.4	-47.9	0.0003
5	ASP103	-13.2	58.8	0.0003
6	ASP131	-34.5	-161.1	0.0005
7	CYS187	128.5	69.3	0.0003

Number of residues in favoured region (~98.0% expected) : 190 (86.8%)
 Number of residues in allowed region (~2.0% expected) : 23 (10.5%)
 Number of residues in outlier region : 6 (2.7%)

We confirmed that GBP is a 6-bladed β -propeller protein, consisting of 6 Tectonin domains (Figure 3.22). Each of the Tectonin domains is made up of 4 β -sheets, which is in agreement with the secondary structure prediction (Figure 3.23, see also Figure 3.18).

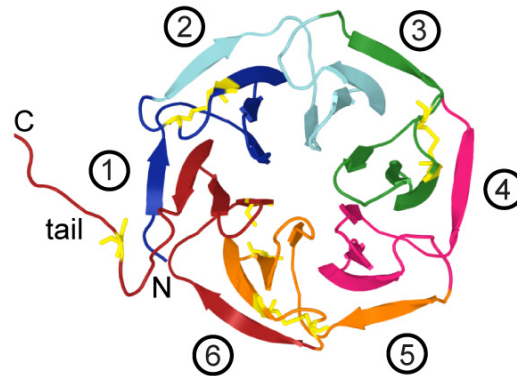


Figure 3.22: Homology model of GBP structure. GBP was predicted to be a 6-bladed β -propeller protein. Numbers in circles represent the 6 Tectonin domains of GBP. Yellow stick structures represent the cysteine residues. The GBP structure contains a “tail” portion, which does not form part of the main β -propeller structure. N=N-terminal, C=C-terminal.

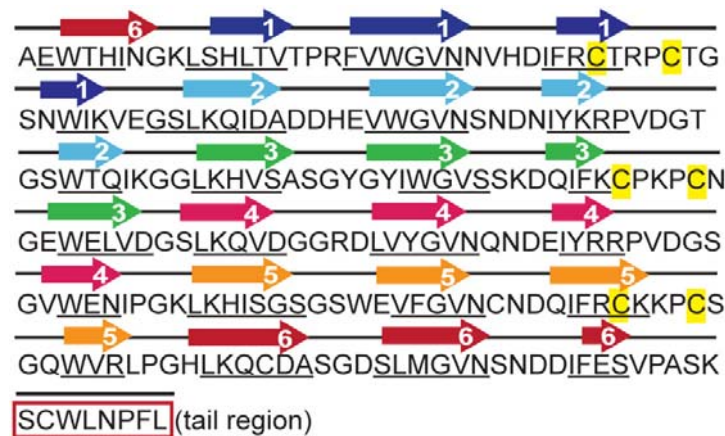


Figure 3.23: The protein sequence of GBP showing 6 Tectonin domain repeats. The modeled β -sheets are displayed to show their positions in the sequence. The coloured arrows correspond to the respective β -strands in the structure shown in Figure 3.22. Underlines amino acids are PSIPRED β -sheet predictions from secondary sequence analysis. Cysteines forming disulphide bonds are highlighted in yellow. The ‘tail’ region that does not make up the main 6 propeller structure is boxed in red.

Analyzing the structural properties of the protein, we observed that out of the 9 cysteine residues in the amino acid sequence of GBP, 6 are likely to be involved in intra-molecular disulphide bridge formation (Figure 3.24) due to their proximity in the

structural space. The other 3 free cysteine residues may be involved in dimerization and formation of higher order homo-oligomers, as observed in Section 3.2.1.2 and Figure 3.13.

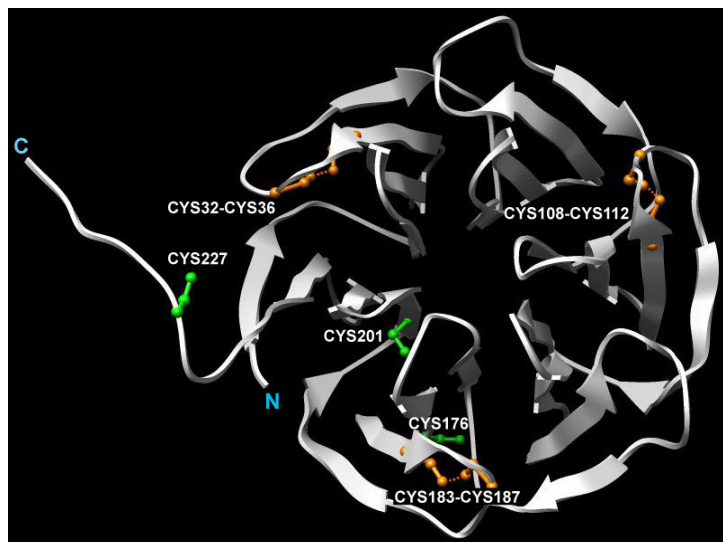


Figure 3.24: Cysteine residues of GBP. GBP contains 9 cysteine residues, of which 6 (orange) are predicted to be involved in intra-molecular disulphide bridge formation and 3 (green) free cysteine residues that may be involved in dimerization and the formation of higher order homo-oligomers.

We observed that the surface of GBP is predominantly hydrophilic (blue, Figure 3.25), with several scattered hydrophobic (red) patches, indicating potential protein-protein or protein-ligand interaction sites. The 6 β -propeller folds of GBP form a hexagonal toroidal-like structure where one end of the central recess is wider and deeper than the other (Figure 3.25A, C). We thus refer to the larger one as a “cavity” the shallower one as a “cavity” (Oubrie et al. 1999; Hata et al. 2008). We also define the cavity end of the tunnel as the top of the molecule and the crevice end as the bottom of the molecule (Figure 3.25B). Predictions by PROFbval (Schlessinger and Rost 2005; Schlessinger et al. 2006) indicate that the flexible regions of GBP occur at the loops between each propeller blade (Figure 3.26). This is consistent with observations from solved β -propeller structures (Fulop and Jones 1999; Paoli 2001).

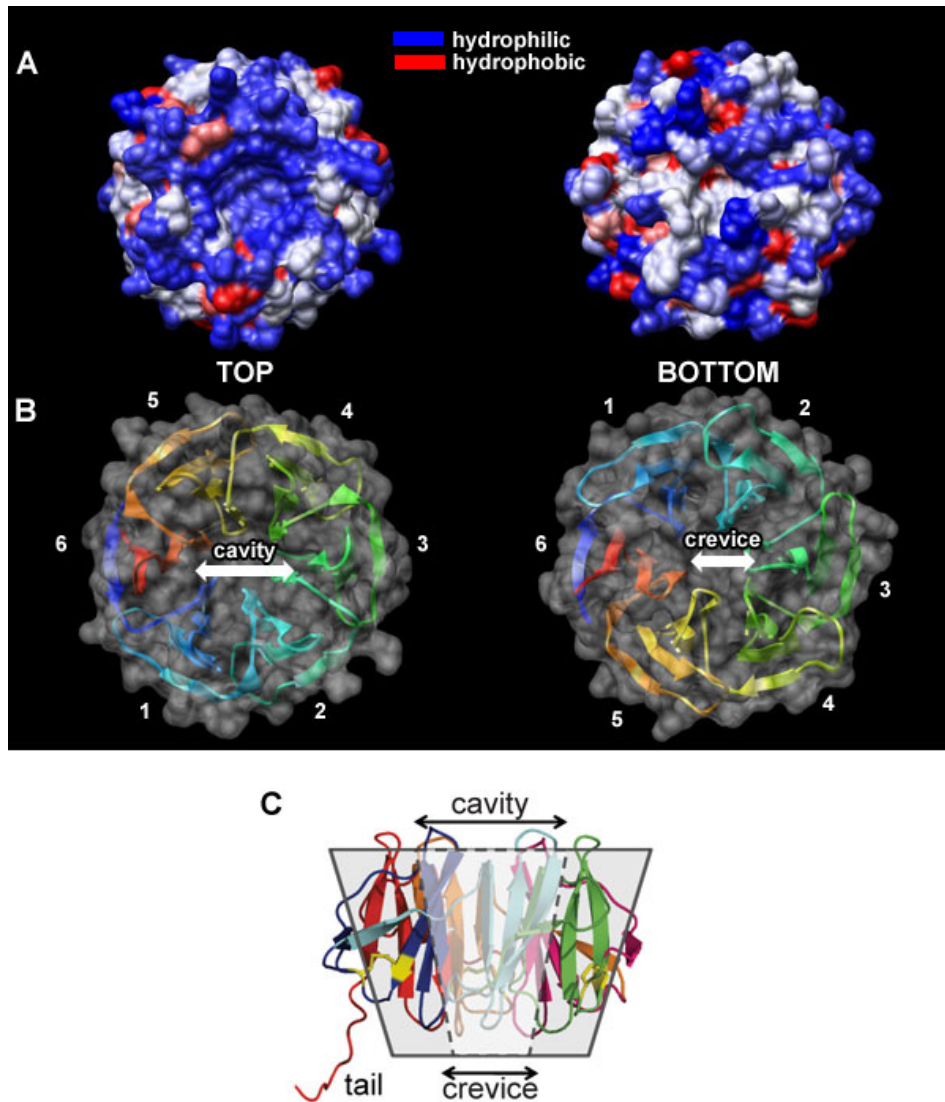


Figure 3.25 : Surface properties of GBP. (A) GBP is predominantly hydrophilic (blue) with scattered hydrophobic patches (red) of possible contact for interaction sites. (B,C) The molecule folds to form a toroidal-like structure with a 6-fold symmetry around the central funnel-shaped molecule displaying a larger “cavity” on the top of the molecule and a smaller “crevice” at the bottom.

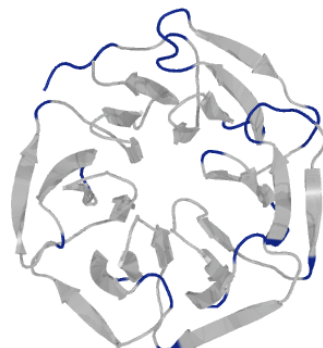


Figure 3.26: Prediction of flexible residues on GBP. The regions in blue indicate flexible residues by PROFbval (Schlessinger et al. 2006). They tend to occur on the loops connecting adjacent β -propellers, similar to those observed in literature (Weis and Drickamer 1996; Paoli 2001; Schlessinger and Rost 2005). Such flexible sites usually indicate ligand binding sites.

In the interest of studying the interaction of GBP with its key interaction partner CRP, the structure of CRP was also modeled (Figure 3.27), through an analogous procedure to GBP as outlined earlier in this section. CRP was homology modeled based on templates from hCRP (PDB:1B09) and human serum amyloid protein (hSAP) (PDB:1SAC), which displays 31% and 30% sequence identity to CRP respectively. The Ramachandran plot (Figure 3.28, Table 3.6) indicates that 94% of the residues are within the allowable range of Phi-Psi values.

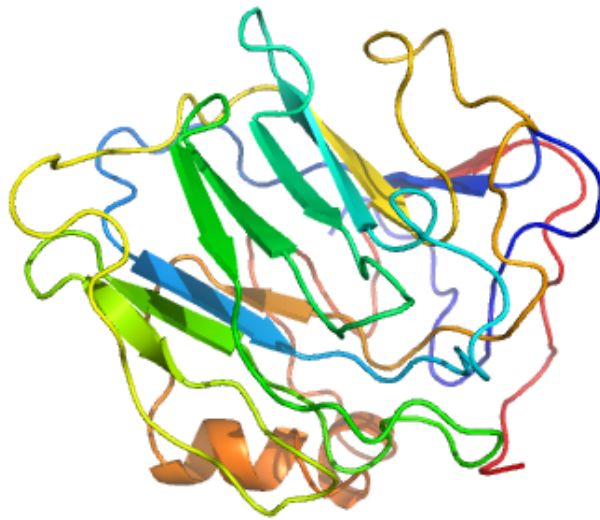


Figure 3.27: The structure of the homology modeled CRP.

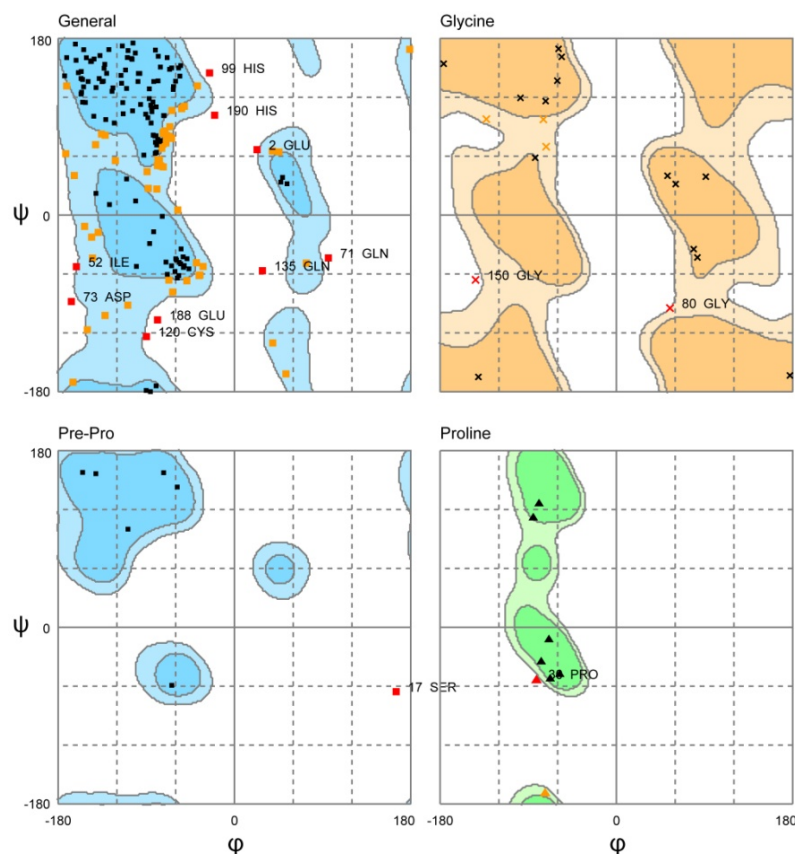


Figure 3.28 : Ramachandran plot for the structure of CRP. The outlier residues listed (Table 3.6) remain close to the boundaries of the permitted Psi-Phi values, which are highlighted by the light blue contours, indicating that the structure has been reliably modeled. Black, orange and red boxes represent residues in favoured, allowed and outlier regions respectively (refer to Table 3.6 for details).

Table 3.6 List of Phi-Psi outliers in the CRP model.

No.	Residue	Psi	Phi
1	GLU2	23.15	66.54
2	SER17	165.11	-65.70
3	PRO30	-81.62	-52.70
4	ILE52	-161.37	-52.69
5	GLN71	95.88	-43.76
6	ASP73	-166.48	-88.13
7	GLY80	54.70	-94.82
8	HIS99	-25.38	144.72
9	CYS120	-89.81	-123.81
10	GLN135	28.58	-56.72
11	GLY150	-143.76	-66.09
12	GLU188	-78.50	-106.83
13	HIS190	-20.21	101.65

Number of residues in favoured region (~98.0% expected) : 148 (68.5%)

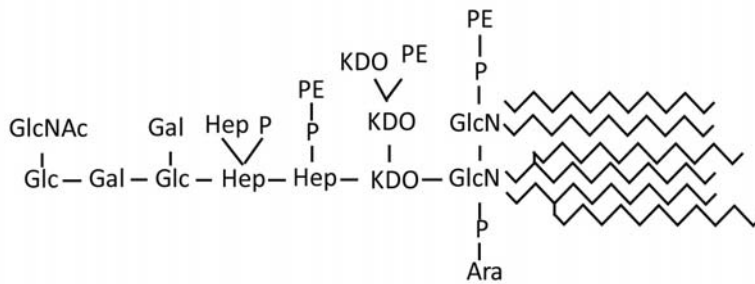
Number of residues in allowed region (~2.0% expected) : 55 (25.5%)

Number of residues in outlier region : 13 (6.0%)

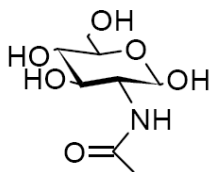
3.2.2.3 Saccharides and LA dock to similar sites in GBP

To define the domains or motifs of GBP that interact with LPS and saccharides, we utilized the computationally-modeled GBP structure for docking studies. Proteins containing β -propeller repeats such as Tachylectin-2 (TL-2) are known to undergo protein-sugar interactions via the backbone atoms of the conserved binding site residues, which are flanked by adjacent β -sheet blades of the β -propeller domains (Beisel et al. 1999; Wimmerova et al. 2003; Cioci et al. 2006). Because GBP has high binding affinity for GlcNAc and possibly for other sugar moieties of the LPS, it is reasonable to expect that the sugar binding sites are also localized between the adjacent β -sheet blades.

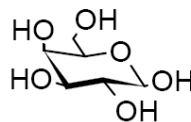
Using computational methods to dock various saccharides found on LPS (Figure 3.29A) onto the model of GBP, galactose (Gal) was predicted to bind GBP with the highest affinity amongst a set of monosaccharides and monosaccharide N-acetylaminines (Gal; GalNAc; glucosamine, Gln; GlcNAc, KDO, and heptose, Hep) (Table 3.7). Among the several potential binding sites for GlcNAc on the GBP surface, the site with the highest affinity was located between the Tectonin domains 1 and 6 (Figure 3.29B). Additional binding sites were also found in between the other adjacent propellers resembling those in TL-2 (Beisel et al. 1999) (see Figure 3.6A), as well as in the central cavity (Figure 3.29B) similar to the predicted KDO-binding site in TL-1 (Kawabata and Tsuda 2002).

ALipopolysaccharide (LPS) of Gram-negative bacteria

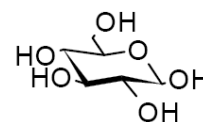
N-acetyl-b-D-glucosamine (GLCNAC)



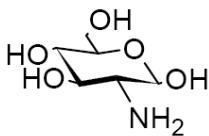
b-D-galactose (GAL)



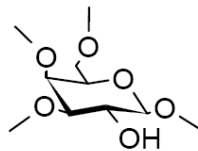
b-D-glucose (GLU)



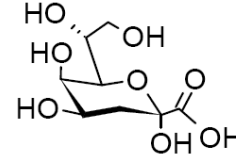
b-D-glucosamine (GLN)



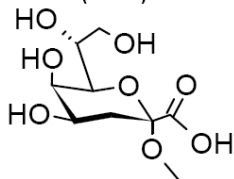
b-D-galactose-CL (GALCL)



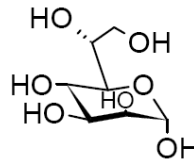
3-deoxy-D-manno-octulosonic acid (KDO)



3-deoxy-D-manno-octulosonic acid (KDO2)



Heptose (HEP3)



Heptose connected (HEP3C)

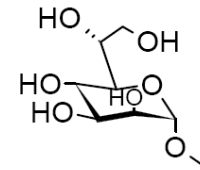
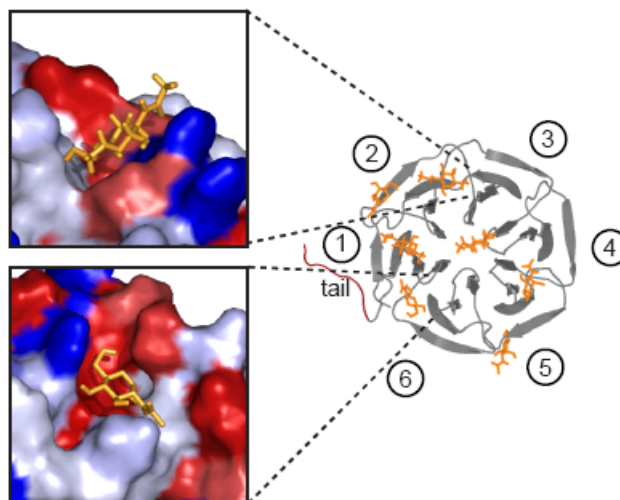
**B**

Figure 3.29: Saccharides dock to the hydrophobic clefts and to the central cavity of GBP. (A) The different sugar structures that are found on the polysaccharide chain of LPS were constructed and their binding to GBP was tested via molecular docking. The energy scores are listed in Table 3.7. **(B)** GlcNAc (orange), the cognate ligand of GBP, was found to dock to either the central cavity or the hydrophobic clefts (inset, red surface) within the individual propellers, similar to that observed in TL-2 (see Figure 3.6A). Numbers 1 to 6 in circles correspond to the 6 Tectonin domains of GBP.

Subsequently, we extended our docking calculations to LA, the core endotoxic region of LPS (Figure 3.30A). 2-N-acetyl-3-O-acetyl- β -D-glucosamine (GlcNAcOAc), the principal component of the disaccharide headgroup of LA, displayed an enhanced affinity towards GBP ($E_{\text{int}} = -66.1 \text{ kcal}\cdot\text{mol}^{-1}$), higher than Gal, the native ligand for GBP (Table 3.7). Phosphate groups at positions 1- or 4- of the GlcNAcOAc did not significantly affect the binding affinity with GBP. The polar disaccharide headgroup of the LA (1,4'-bisphospho- β -(1,6)-2,2'-N-acetyl-3,3'-O-acetyl-D-glucosamine disaccharide) showed significant affinity towards GBP, higher than the sum of the E_{int} of its components (GlcNAcOAc-1-Phos, GlcNAcOAc-4-Phos). The core LA (headgroup together with its fatty acid chains) exhibited similar binding affinity as the headgroup itself. Therefore, we predict that GBP recognizes and preferentially binds the glucosamine disaccharide headgroup of the LA over the non-polar fatty acid chains, consistent with the observation that LA and GlcNAc share similar binding sites in GBP (Figure 3.30B).

Table 3.7: Computed binding energies for top scoring saccharides and LA poses docked to GBP.

Ligand	E_{int} (kcal\cdotmol$^{-1}$)^a
Galactose (Gal)	-58.8
Glucose (Glu)	-51.3
Glucosamine (Gln)	-45.2
N-acetyl-glucosamine (GlcNAc)	-35.0
3-deoxy- α -D-manno-octulosonic acid (KDO)	-52.5
2-N-acetyl-3-O-acetyl glucosamine (GlcNAcOAc)	-66.1
GlcNAcOAc-1-Phosphate	-65.7
GlcNAcOAc-4-Phosphate	-63.3
1,4'-bisPhos-GlcNAcOAc-1,6-disaccharide	-140.0
Core LA	-126.0

^a E_{int} is the sum of electrostatic and Van der Waals ligand-receptor binding energy contributions as defined in the AMBER99 forcefield

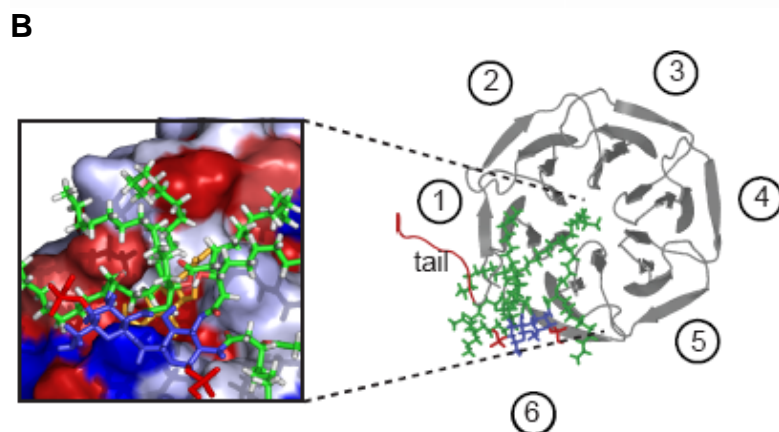
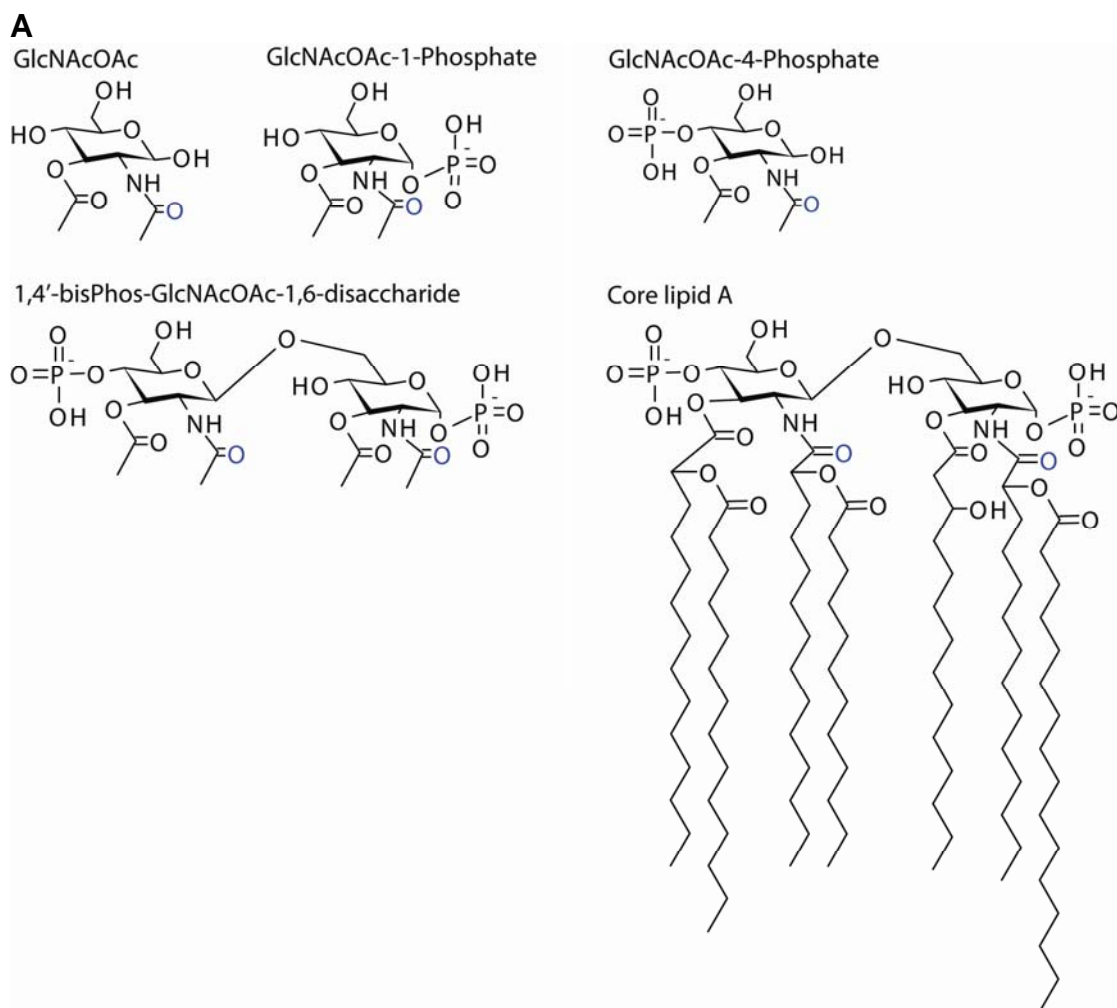


Figure 3.30: LA docks to the hydrophobic cleft of GBP. (A) The different component structures that are found on LA were constructed and their binding to GBP was tested via docking. The energy scores are listed in Table 3.7. (B) The core LA (fatty acid chains – green, phosphate groups – red, glucosamine – blue) was docked to GBP. The highest affinity pose of the core LA to GBP is between the GBP β -propeller Tectonin domains 1 and 6. Numbers 1 to 6 in circles correspond to the 6 Tectonin domains of GBP.

3.2.3 Molecular mechanism of GBP:LPS interaction

After modeling the structures of GBP and CRP, we were prompted to experimentally elucidate the interactions of GBP with LPS and CRP. Previously, our lab showed that infection and LPS challenge triggered serine proteases which affected the level of GBP binding to the pathogen (Le Saux et al. 2008). Furthermore, the binding of GBP to CRP is triggered and the amount of GBP molecules binding to CRP is enhanced upon infection (Ng et al. 2007). Thus, it was imperative to study the molecular mechanisms underlying the interactions, and how they work in concert to form the pathogen recognition interactome. Does GBP first sense the pathogen before being triggered to bind CRP? Or can the Tectonin protein simultaneously interact with both CRP and the pathogen, possibly enhancing the interaction strength through cooperative binding? The following section will attempt to answer these questions.

3.2.3.1 GBP interacts with LPS via sugar groups

Since GBP binds to the galactose residues of Sepharose, and is eluted by GlcNAc, we hypothesized that GBP binds to GlcNAc which is present on the outer core of LPS (Figure 3.31A). Using ELISA with GlcNAc immobilized on the surface showed that purified GBP bound to GlcNAc specifically and dose-dependently (Figure 3.31B). Additionally, the recruitment of GBP to immobilized full length LPS was inhibited by GlcNAc (Figure 3.31C). GBP also showed specific interactions to the truncated fragments of LPS : ReLPS and LA (Figure 3.31D,E). Likewise, GBP is able to bind Gram-positive bacterial LTA (Figure 3.31A), and this binding too was abrogated by the addition of GlcNAc (Figure 3.31F). These results indicate that the GlcNAc moiety is a strong ligand of GBP, and a target molecule for GBP when binding LPS or LTA.

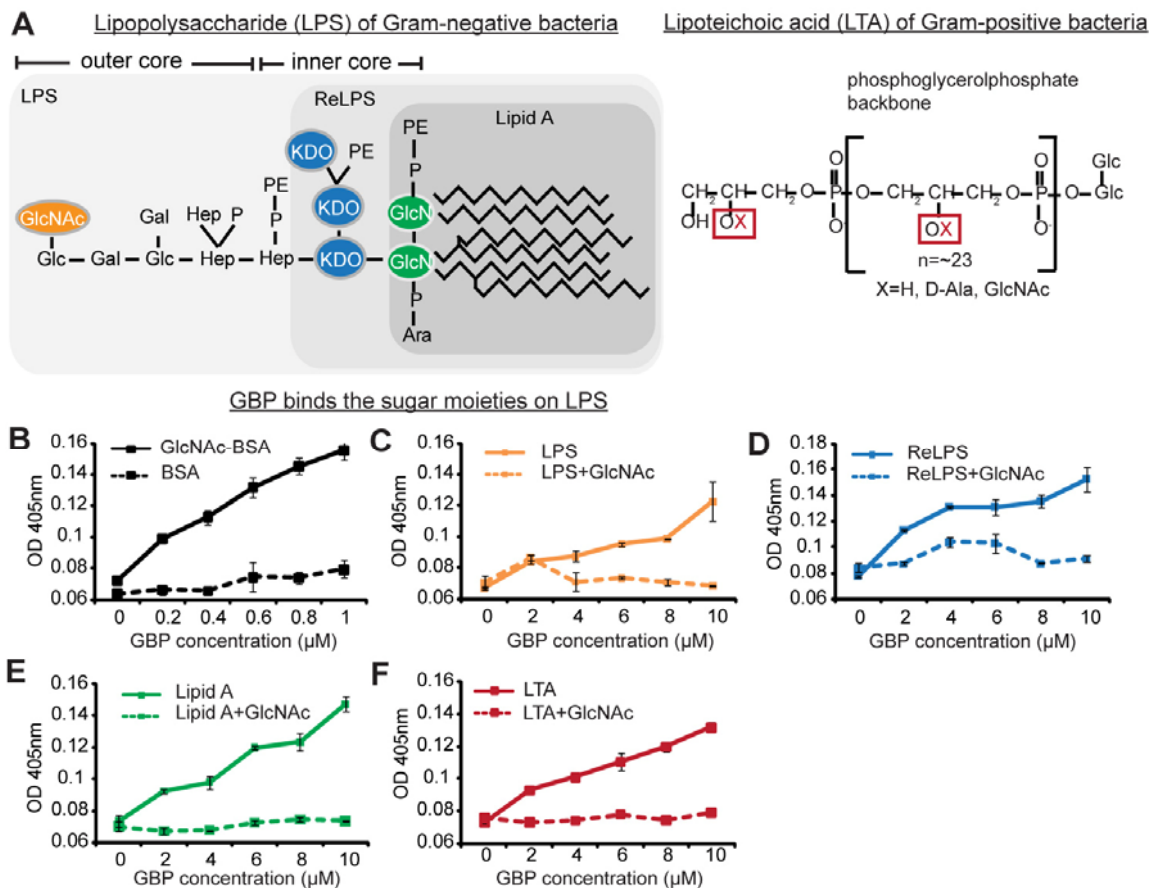


Figure 3.31: The sugar moieties of LPS and LTA bind to GBP. (A) Chemical structures of LPS and LTA. The LPS, ReLPS and LA display different extent of sugar moieties. GlcNAc is located on the outer core of LPS and LTA. (B-F) BSA, GlcNAc-BSA, LPS, ReLPS or LA were first immobilized on ELISA Polysorp™ plates. GBP with or without prior incubation with GlcNAc was then added to the wells containing immobilized ligands. Anti-GBP antibody was used to detect GBP bound on the surface of the ligands. Fluorescence at 405nm was measured to determine the GBP-ligand binding ability. The addition of GlcNAc abrogated the binding of GBP to (B) GlcNAc itself (C) LPS, (D) ReLPS, (E) LA and (F) LTA.

3.2.3.2 The different lengths of LPS bind strongly to GBP

Since GBP interacts specifically with LPS, we examined the binding kinetics of GBP with different regions of LPS. We utilized real-time biointeraction assay by surface plasmon resonance (SPR) to measure the binding affinities. GBP:GlcNAc, GBP:LA and GBP:LPS all displayed similar binding affinities, with K_D values of 1.52×10^{-7} M to 2.52×10^{-7} M (Figure 3.32A-C, Table 3.8). This corroborates our interpretation that GBP binds LPS via its GlcNAc moiety which is commonly present amongst LPS and LA, and thus producing similar levels of binding strength. However, the GBP:LPS

interaction exhibited a slower k_{on}/k_{off} rate compared to GBP:GlcNAc, suggesting that GBP interacts with multiple sugar moieties in the full-length LPS. Interestingly, GBP binds ReLPS with a 10-fold greater affinity (K_D of 2.78×10^{-8} M) (Figure 3.32D) compared to the full length LPS. We suggest that besides having glucose residues (similar to GlcNAc), the ReLPS has several exposed KDO moieties, which may also be available for GBP binding, contributing to the higher overall affinity. Together, these results show that GBP binds the sugar moieties of LPS.

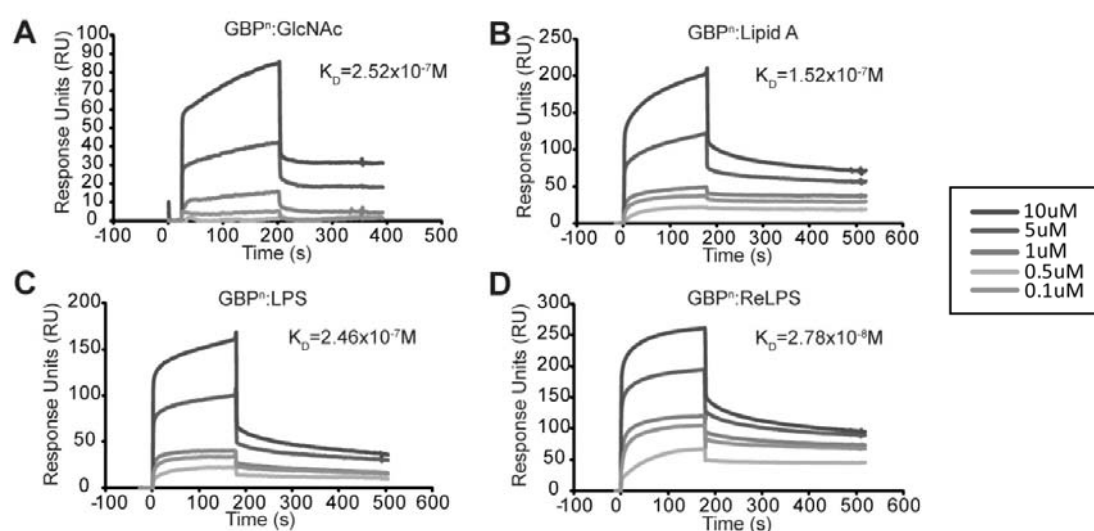


Figure 3.32: GBP binds LPS with high affinity. Surface plasmon resonance analysis of real-time binding of GBP to LPS, LPS-truncates and GlcNAc. GlcNAc or LA or LPS or ReLPS was immobilized on Biacore CM5/HPA chips. GBP at different concentrations was injected over these ligand-bound surfaces and the GBP binding affinity was quantified from the response unit (RU). GBP purified from naïve plasma (GBPn) binds: (A) GlcNAc with a KD of 2.52×10^{-7} M, (B) LA at KD 1.52×10^{-7} M, (C) LPS at KD 2.46×10^{-7} M, (D) ReLPS at KD 2.78×10^{-8} M. The data were analyzed by BIAevaluation Version 3.2. (E) Glutathione Sepharose (GST) protein was used as a non LPS-binding control in this experiment.

Table 3.8: Rate constants and equilibrium dissociation constants of binding kinetics of ligands to GBP.

Ligand	k_a ($\text{mol}\cdot\text{s}^{-1}$) ^a	k_d (s^{-1}) ^b	K_D (mol^{-1}) ^c
GlcNAc	$3.68 \pm 0.13 \times 10^2$	$9.28 \pm 0.73 \times 10^{-5}$	2.52×10^{-7}
LA	$4.98 \pm 0.08 \times 10^3$	$7.58 \pm 0.51 \times 10^{-4}$	1.52×10^{-7}
LPS	$5.45 \pm 0.10 \times 10^3$	$1.34 \pm 0.06 \times 10^{-3}$	2.46×10^{-7}
ReLPS	$3.43 \pm 0.71 \times 10^4$	$9.53 \pm 0.41 \times 10^{-4}$	2.78×10^{-8}

^a rate constant of ligand-GBP association.

^b rate constant of ligand-GBP dissociation.

^c equilibrium dissociation constant of ligand-GBP dissociation: $K_D = k_d/k_a$.

3.2.3.3 The interaction between GBP and LPS is independent of Ca^{2+}

Infection is usually accompanied by local acidosis and hypocalcaemia (Holland et al. 2002; O'Croinin et al. 2008). As a central plasma protein PRR, CRP is known to form a pathogen recognition complex in a calcium dependent manner (Ng et al. 2004; Ng et al. 2007). However, the exact function or effect of the cation-binding on GBP (or the Tectonin lectins) is unknown. Therefore, here, we tested if GBP's binding to LPS could be affected by Ca^{2+} binding. We ran surface plasmon resonance with GBP in 3 conditions: buffer alone, no free calcium in the system; buffer with EGTA, to chelate calcium ions from the system; and buffer supplemented with the physiological calcium concentration of 2.5mM. We find that all three conditions gave a similar binding affinity of 10^{-7}M (Figure 3.33). We thus conclude that calcium binding does not play a role in the GBP-LPS interaction.

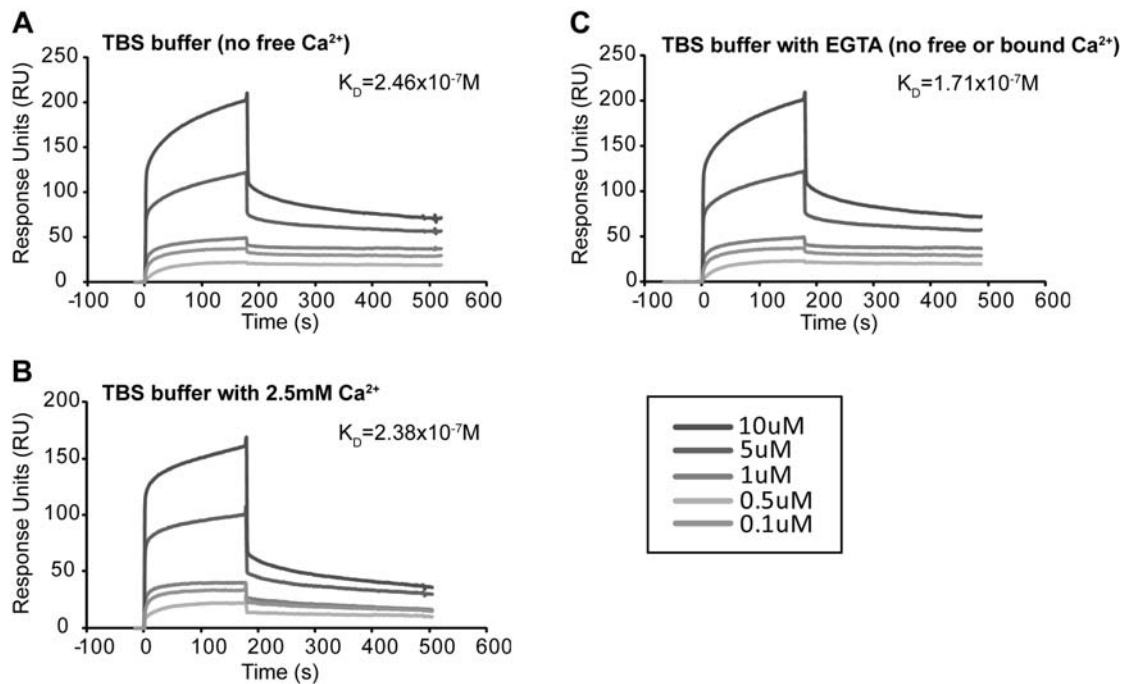


Figure 3.33: The interaction between GBP and LPS is independent of Ca^{2+} . SPR showed that GBP's binding affinity to LPS did not change in (A, B) the presence or (C) absence of Ca^{2+} .

3.2.3.4 GBP interacts with LPS via distinct interaction surfaces

At this juncture, we set out to experimentally map the interaction sites of GBP with LPS, and test the validity of the computationally predicted LPS-binding sites on GBP. We also aimed to localize the Tectonin domains within GBP which confer preferential and distinctive binding to CRP. To achieve this, we used amide hydrogen-deuterium exchange coupled with mass spectrometry (HDMS). The identification of exposed or hidden sites is represented by mass spectrometry-identified peptide sequences. A mass shift resulting in a lower mass value is an indication of specific peptide sequence contained in an interaction surface, and thus is involved in an interaction (Figure 3.34).

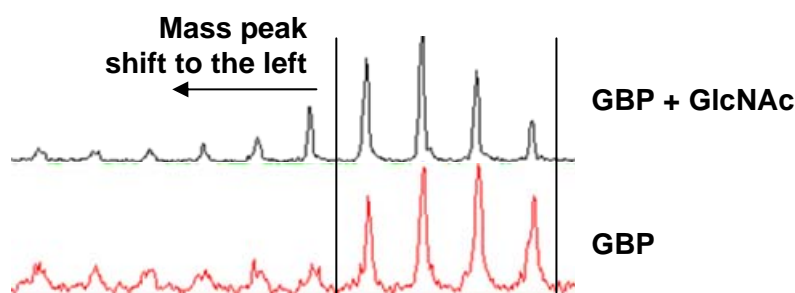


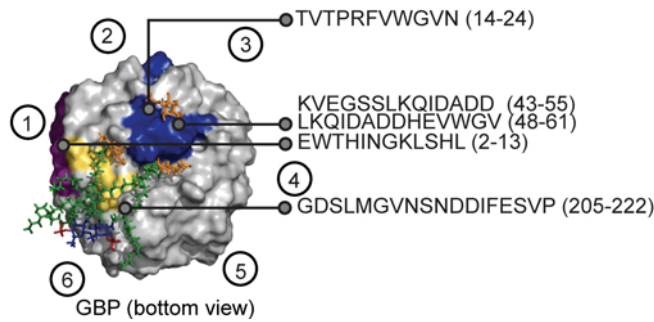
Figure 3.34 : Example of HDMS mass shift in the GBP sample. An example of a mass spectrometry output showing the shift of the mass peaks to the left in a specific peptide sequence of GBP, comparing the state with or without the ligand (GlcNAc). Here, we can see that there is a left-shift of the GBP+GlcNAc peaks (in black), indicating that this particular peptide has a site blocked from deuteration – an indication of the peptide involved in the binding or interaction site of GlcNAc with GBP.

Comparing the extent of deuterium exchange of GBP alone and GBP incubated with GlcNAc showed nine peptides with significant differences in deuterium exchange. Peptides spanning GBP residues 14-24 (TVTPRFVWGVN), 43-55 (KVEGSSLKQIDADD) and 48-61 (LKQIDADDHEVWGV) (Figure 3.35A, blue surfaces) showed decreased deuterium exchange in the presence of GlcNAc (Figure 3.35B, right column; Table 3.9), suggesting these regions in GBP bind GlcNAc. These results are consistent with the *in silico* docking predictions (Figure 3.35A,

GlcNAc molecules in orange, see also Figure 3.25) because the regions coincide with the interface between the β -propeller folds and also the central cavity and crevice which exhibit one of the highest binding affinities for GlcNAc docking. Therefore, these regions of GBP likely contain the GlcNAc recognition site. Peptide 205-222 (GD β SLMGVNSNDDIFESVP) (Figure 3.35A, yellow surface), corresponding to the Tectonin domains 6-to-1, showed a decreased deuterium uptake in the presence of LA (Figure 3.35B, left column; Table 3.9) , suggesting a possible LA binding site within the region which is proximal to the predicted LA-binding site.

Interestingly, when GBP purified from infected HSC hemolymph was used for the experiment, an additional peptide 2-13 (EWTHINGKLSHL) (Figure 3.35A, purple surface), with decreased deuterium uptake (Figure 3.35B, left column; Table 3.9), corresponding to an interaction site was found. The peptide sequence happens to contain the motif “HINGK”, which follows a known LPS-binding pattern of *BHB(P)HB* (B=basic, H=hydrophobic, P=polar amino acid residues), previously reported by Frecer et. al. (2000). They showed the propensity of these motifs in beta structures to recognize and bind the LA moiety of LPS with high affinity. The findings in these HDMS experiments correlated well with previous motif prediction studies by Frecer et. al. and also our work done in computational molecular docking (see Section 3.2.2.3).

A GBP surface peptide interactions with Lipid A and GlcNAc



B GBP - ligand interactions

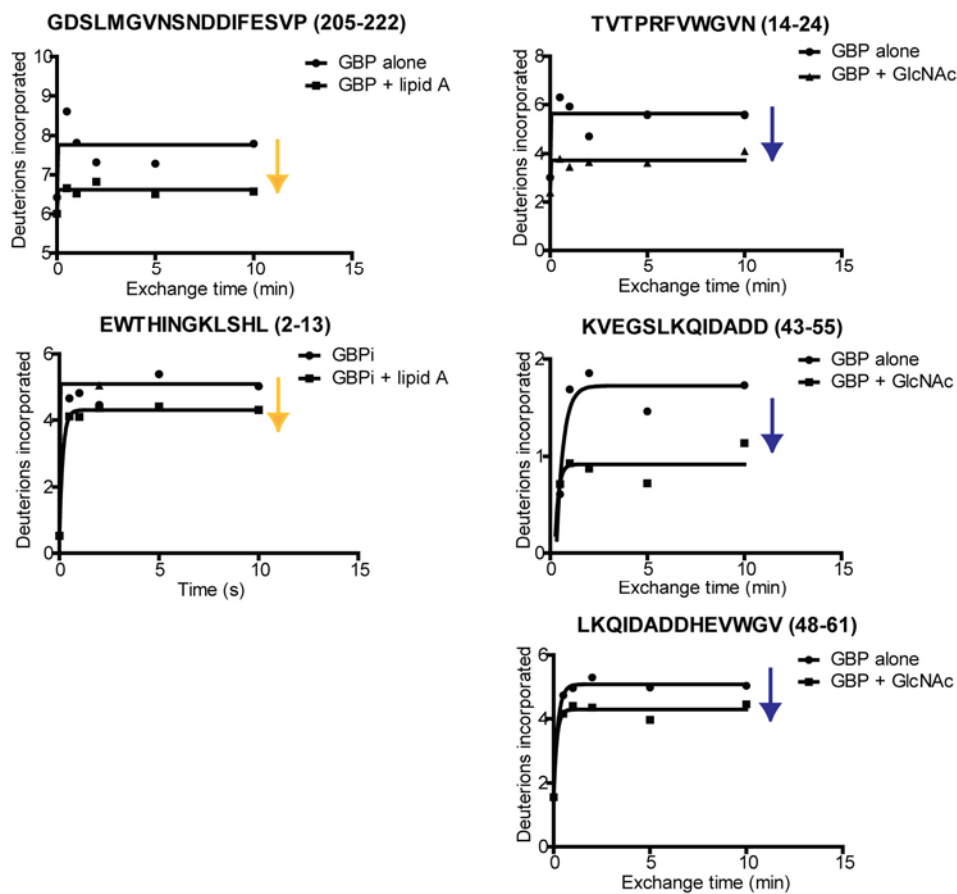


Figure 3.35: HDMS analysis - surface interaction of GBP with either GlcNAc or LA. (A) Peptides showing decreased deuterium uptake in GBP when in interaction with either GlcNAc or lipid A were mapped onto the surface of the structure (GlcNAc, blue surface; LA, yellow & purple surfaces). The top poses from the docking results (GlcNAc, orange; LA, green-blue-red) are included for comparisons with HDMS observations. Numbers represent the corresponding Tectonin domains on GBP. **(B)** Graphs showing the change in deuterium uptake in the respective peptides. Downward arrows indicate the reduction of deuterium uptake. The colour of the arrows correspond to the surface colours in (A).

3.2.4 Mechanism of action of GBP: CRP interaction

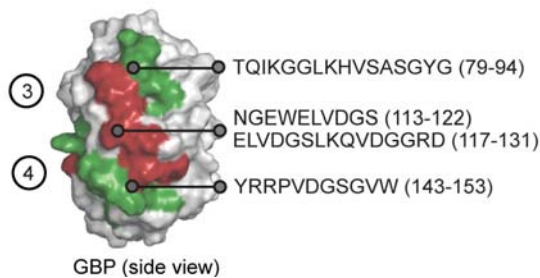
3.2.4.1 HDMS reveals that GBP interacts with CRP through a non-symmetrical protein-protein contact

Using the same HDMS principle of determining the contact interfaces, we found that in the presence of CRP, the GBP peptides 113-122 (NGEWELVDGS) and 117-131 (ELVDGSLKQVDGGRD) (Figure 3.36, red surface), corresponding to the outermost β -strand of the 4th Tectonin domain, showed a decreased deuterium exchange (Figure 3.36B, right column; Table 3.9) suggesting that it contains the GBP-CRP interaction site. Correspondingly, the adjacent peptides spanning residues 79-94 (TQIKGGLKHVSASGYG) and 143-153 (YRRPVDGSGVW) (Figure 3.36, green surfaces) showed an increased deuterium exchange (Figure 3.36B, left column; Table 3.9), indicating greater solvent accessibility, possibly due to induced conformational changes which tallies with the possibility that the red surface is participating in an interaction, hence a slight change in conformation at its neighbouring surfaces.

Conversely, we found that three peptide regions in the CRP molecule - 1-12 (KVKFPPSSSPSF), 8-19 (SSPSFPRLVMVG) and 121-150 (MGVTFRQGGLVVLGQDQDSVGGGFDAKQSL) - showed a decreased deuterium exchange, indicating interactions with GBP (Figure 3.37, Table 3.10). Although these regions are far apart in the primary sequence, when mapped to the homology-modeled CRP structure, they are juxtaposed in their tertiary structure, as expected if they interacted with GBP. Peptide 121-150 includes residues known to be crucial for calcium binding (D136, Q137, D138, Q148), which provides an explanation for the observation that CRP binds GBP only when the Ca^{2+} level is low (Ng et al. 2007). Taken together, these results show that GBP has distinct binding sites for LPS and

CRP; the binding sites for LA and GlcNAc are localized to the hydrophobic clefts interfacing propeller folds, whereas the binding sites for CRP are contained within the 4th Tectonin domain. Therefore, despite the apparent structural symmetry of the 6 Tectonin domains of GBP, it is able to differentiate between 2 different structures – LPS from the pathogen with its Tectonin domain 6-to-1 and CRP, its partner protein in the host system with its Tectonin domain. This unique property of the Tectonin domains makes it a very important bridge in host-pathogen interactions.

A GBP conformational change and interaction with CRP



B GBP interaction sites with CRP

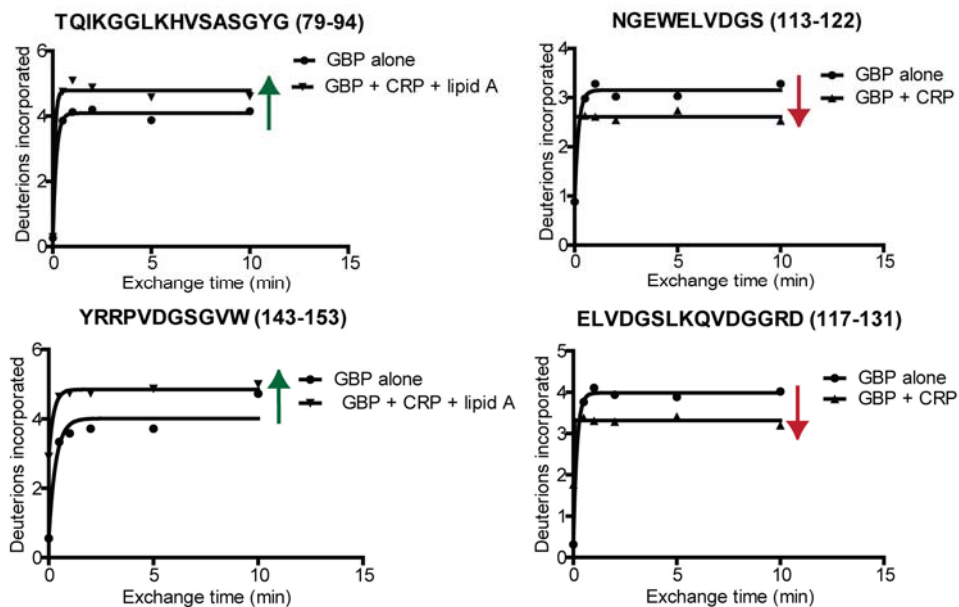
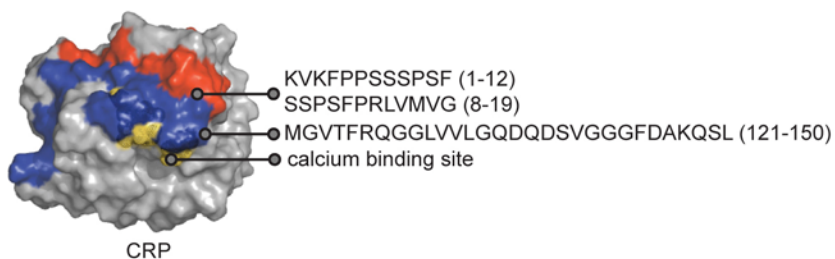


Figure 3.36: HDMS analysis – surface interaction of GBP with CRP. (A) Peptides showing change in deuterium uptake in GBP when in interaction with CRP were mapped onto the surface of the structure (red – decreased deuterium uptake; green – increased deuterium uptake). Numbers represent the corresponding Tectonin domains on GBP. (B) Graphs showing the change in deuterium incorporation in the respective peptides. Downward arrows indicate the reduction of deuterium uptake, and upward arrows indicated increase in deuterium uptake. The colour of the arrows correspond to the surface colours in (A).

A CRP interaction sites with GBP



B CRP interaction sites with GBP

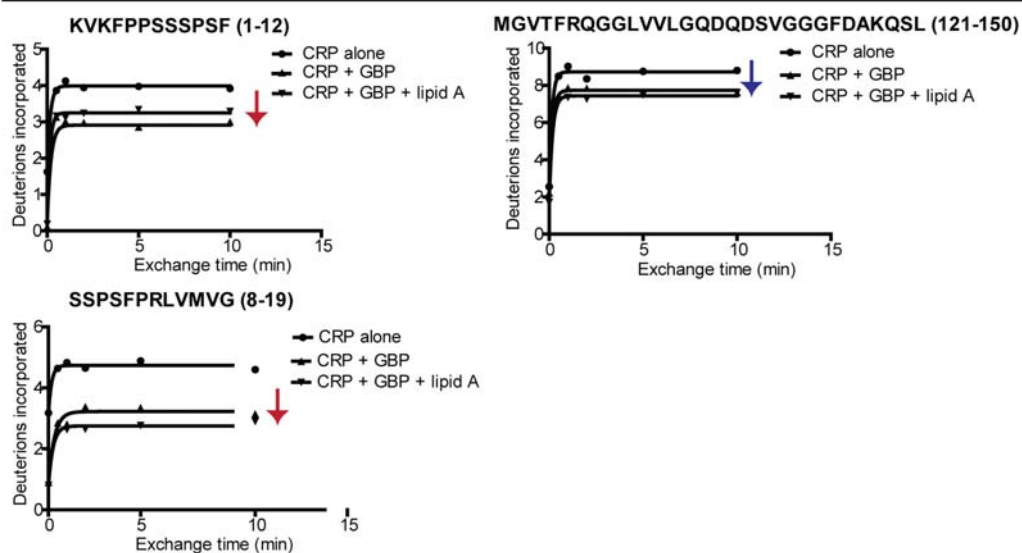


Figure 3.37: HDMS analysis – surface interaction of CRP with GBP. (A) Peptides showing decrease in deuterium uptake in CRP when in interaction with GBP were mapped onto the surface of the structure (blue and red surfaces). The known calcium binding site of CRP based on the important residues for Ca^{2+} binding is labeled in yellow. (B) Graphs showing the change in deuterium incorporation in the respective peptides. Downward arrows indicate the reduction of deuterium uptake. The colour of the arrows correspond to the surface colours in (A).

Table 3.9: Summary of H²H exchange data for GBP.

Fragment of GBP (m/z)	No. of amides	Deuteration (5 min)				
		GBP	GBP with GlcNAc	GBP with LA	GBP with CRP	GBP with LA & CRP
2-13 (1434.74)	11	4.13±0.44	4.40±0.52	4.25±0.21	4.58±0.17	4.60±0.15
14-24 (1275.68)	9	6.04±0.44	2.58±0.02	5.80±0.21	5.15±0.02	5.37±0.31
17-24 (974.51)	6	2.08±0.08	2.21±0.10	2.19±0.03	2.22±ND	2.19±0.01
36-47 (1337.61)	11	4.73±0.16	4.92±0.19	4.74±ND	5.76±0.15	5.46±0.28
43-55 (1417.71)	12	1.57±0.11	1.11±0.39	1.53±0.01	2.01±0.10	4.95±0.24
48-61 (1624.79)	13	4.99±0.01	4.74±0.77	5.01±0.13	5.34±0.04	5.23±0.04
56-67 (1383.62)	11	4.10±0.24	4.34±0.08	4.23±0.18	4.49±0.25	4.45±0.03
66-78 (1492.75)	11	4.12±0.19	4.46±0.28	4.26±ND	4.53±0.01	4.58±0.01
68-78 (1265.62)	9	3.37±0.42	3.38±0.21	3.35±0.03	3.49±0.12	3.48±0.07
79-94 (1602.85)	15	4.09±0.21	4.28±0.16	4.17±0.03	4.20±0.50	4.59±ND
107-121 (1788.8)	12	4.96±0.28	5.21±0.22	5.07±0.01	5.37±0.10	5.22±0.08
113-122 (1105.57)	9	3.21±0.18	3.34±0.18	3.29±0.08	2.74±ND	2.66±0.05
117-131 (1587.79)	14	3.99±0.11	3.33±0.17	4.01±0.02	3.51±0.10	4.11±0.01
127-143 (1926.88)	16	4.37±0.34	4.52±0.19	4.38±0.06	4.61±0.07	4.51±0.04
142-153 (1404.73)	10	4.11±0.27	4.18±0.23	4.50±0.21	3.93±0.42	3.95±0.36
143-153 (1291.65)	9	3.98±0.26	4.71±0.26	4.64±0.10	4.91±ND	4.84±0.03
154-169 (1709.89)	14	4.23±0.32	4.61±0.37	4.32±0.05	4.51±0.11	4.54±0.01
161-177 (1877.85)	16	5.72±0.32	5.78±0.24	5.82±0.05	6.14±0.07	6.14±0.04
198-214 (1795.79)	16	4.16±0.20	4.35±0.08	4.23±ND	4.23±0.01	4.27±0.04
205-222 (1895.83)	16	7.87±0.59	7.26±0.34	6.63±0.12	6.65±0.09	6.61±0.07

ND denotes no significant difference in standard error of mean.

In accordance with similar studies in (Brudler et al. 2006; Hamuro et al. 2006; Horn et al. 2006) changes in deuterium incorporation of more than ±10% were considered to be significant.

Table 3.10: Summary of H²H exchange data for CRP.

Fragment of CRP (m/z)	No. of amides	Deuteration (5 mins)			
		CRP	CRP with LA	CRP with GBP	CRP with GBP & LA
1-12 (1307.69)	8	3.87±0.10	4.94±0.14	3.23±0.38	3.58±0.25
8-19 (1292.66)	9	4.52±0.35	4.74±0.61	3.30±0.04	2.96±0.19
15-32 (2167.07)	16	5.75±0.19	6.59±ND	5.90±0.03	6.17±0.17
32-46 (1867.04)	14	6.47±0.23	7.59±0.07	7.02±0.10	6.98±0.06
34-46 (1533.89)	12	2.98±0.04	4.55±0.13	2.99±0.14	3.05±0.05
48-58 (1206.54)	10	4.24±0.11	4.71±0.23	4.77±0.06	4.84±0.02
68-79 (1398.74)	11	4.40±0.09	5.66±0.28	4.83±0.07	4.97±0.04
84-116 (3820.73)	32	11.58±0.06	13.50±0.09	12.48±0.20	12.84±0.18
90-115 (3029.31)	25	8.69±0.23	10.32±0.06	9.58±0.26	9.65±0.16
104-130 (2816.39)	26	8.31±0.54	9.26±0.21	8.26±0.49	8.40±0.47
121-150 (3082.52)	29	8.41±0.32	10.62±0.19	7.55±0.06	7.58±0.11
122-137 (1673.89)	15	4.55±0.11	5.36±0.32	5.01±0.45	5.44±0.05
127-138 (1228.61)	11	2.68±0.22	2.84±0.34	2.81±0.10	2.90±0.21
144-153 (1093.55)	9	3.54±0.32	4.06±0.09	3.65±0.17	3.58±0.30
157-180 (2928.41)	22	9.09±0.24	9.05±0.03	8.56±0.17	8.46±0.46
161-169 (1114.57)	8	5.18±0.03	5.86±0.36	5.13±0.09	5.14±0.09
181-195 (1913.97)	14	5.33±0.20	6.05±0.07	9.09±0.02	9.44±0.40
192-200 (1057.49)	8	3.65±0.08	4.17±0.14	4.97±0.07	4.87±0.06
194-204 (1198.57)	10	3.70±0.37	4.49±0.06	3.97±0.53	4.33±0.11

ND denotes no significant difference in standard error of mean.

In accordance with similar studies in (Brudler et al. 2006; Hamuro et al. 2006; Horn et al. 2006) changes in deuterium incorporation of more than ±10% were considered to be significant.

3.2.4.2 Yeast 2-hybrid interaction analyses show interaction domains consistent with HDMS observations

Following the HDMS results, we used yeast 2-hybrid co-transformation of GBP with CRP to validate the earlier findings. Guided by the homology-modeled GBP structure, we sub-cloned the 6 Tectonin domains of GBP individually; in duos (domains 1+2, 2+3, 3+4, 4+5, 5+6); and in trios (domains 1+2+3, 4+5+6). Each of these GBP sub-clones was tested for their interactions with the full-length GBP (for homodimerisation) as well as with CRP (heterodimerisation).

We find that each Tectonin domain exhibits different ability to interact with GBP or CRP from the differing strength of the yeast growth with the different constructs (Figure 3.38). We observed that clones with three contiguous Tectonin domains consistently gave strong growth equivalent to those of full-length transformations, implying that 3 Tectonin domains are sufficient to interact as strongly as the full-length GBP with itself and with CRP. Clones consisting of 2 domains in general show stronger growth compared to the single domain clones. We therefore postulated that two or more consecutive Tectonin domains are needed for consistent strong interactions between GBP and CRP. From the interaction yeast growth observations, we summarize that constructs of Tectonin domains 3+4 and 4+5, are likely to be strongly involved in GBP-CRP interaction. While domain 4 (as identified by HDMS, see Figure 3.36A, red surface) individually shows a lower level of interaction, when coupled with its neighbouring Tectonin domains of either domain 3 or 5, strong interactions were achieved. We believe that neighboring domains are important for proper propeller folding to facilitate protein-protein interactions. We surmise that the

Tectonin domain 4 of GBP is the key interaction domain with CRP based on the observations of interaction of domains 3+4 and 4+5.

Yeast 2-hybrid interaction between GBP Tectonin constructs with GBP/CRP

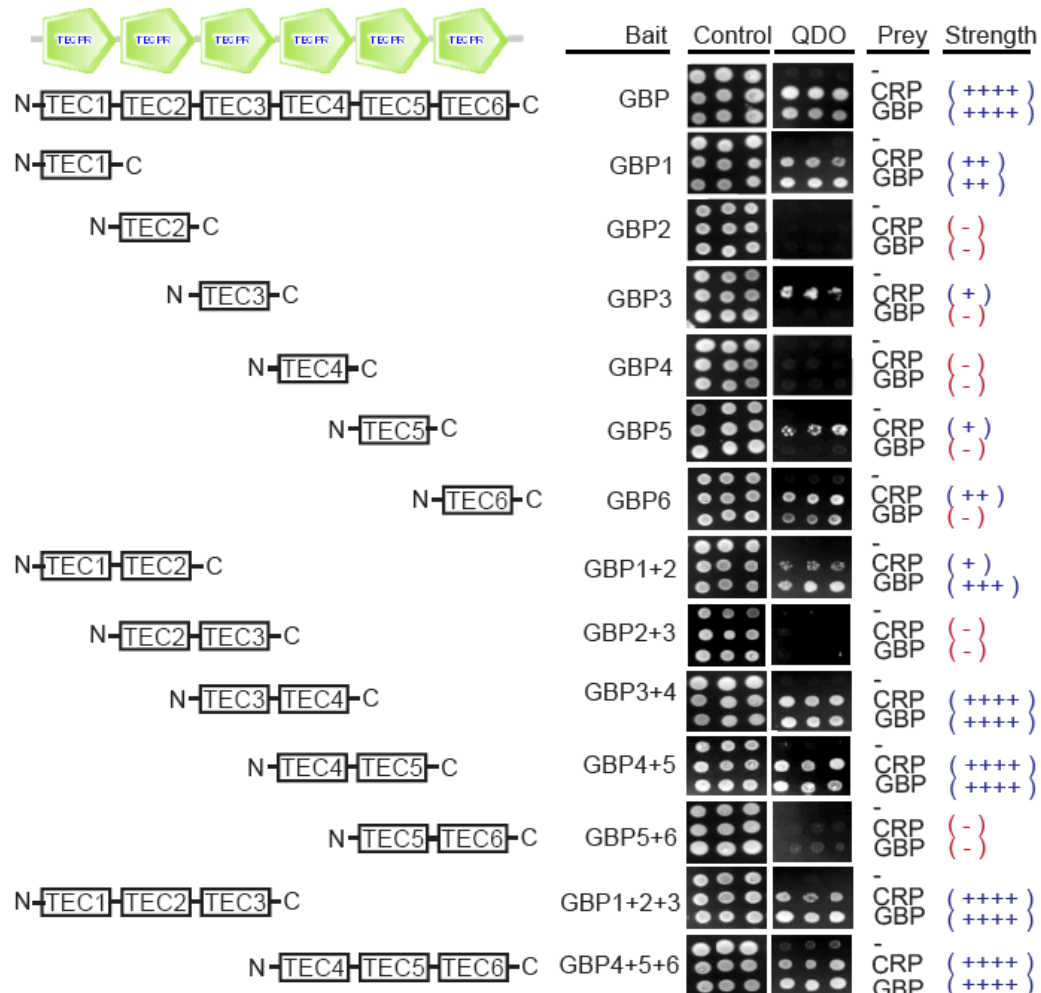


Figure 3.38: Yeast 2-hybrid analysis shows specific Tectonin domains of GBP interact with CRP. Single, double and triple Tectonin domain-constructs were subcloned and tested for their interaction with CRP and with full-length GBP. Tectonin constructs were cloned into pGBKT7 vector and CRP/full-length GBP were cloned into pGADT7 vector. The 2 constructs were co-transformed and spotted on SD-Leu-Trp plates (double-dropout media, as control) and SD-His-Ade-Leu-Trp (quadruple-dropout media, QDO). The strength of interaction is reflected by the intensity of yeast growth.

3.2.4.3 Protein-protein docking reaffirms the feasibility of GBP:CRP binding region

After having identified interaction surfaces by HDMS together with yeast 2-hybrid co-transformation, showing that the Tectonin domain 4 of GBP is vital in the

interaction with CRP, we aimed to model a structural representation and test the feasibility of the resulting complex formation using protein-protein docking. The interacting peptide sequences determined earlier using HDMS (see Figure 3.36 and 3.37) were used to define the active residues in order to generate restraints to drive the protein docking process using the HADDOCK docking application (Dominguez et al. 2003; de Vries et al. 2007). A guided docking run (van Dijk et al. 2005) using data gleaned from the HDMS experiments (see Figures 3.36 and 3.37) was carried out by defining the GBP sequence 113-131 and CRP sequences 1-19 and 121-150, more precisely the solvent accessible GBP residues 113-122 and 130-132 and CRP residues 1-10, 16-19, 119-128 and 138-148 as active residues. A blind docking (Hetenyi and van der Spoel 2002; Hetenyi and van der Spoel 2006) algorithm run which considered all solvent accessible residues of both monomers as potential protein-protein interaction sites was also carried out in comparison to the guided docking run.

We observed that the non-symmetric GBP-CRP heterodimer model obtained by guided docking displayed a higher score, reflecting higher stability, than the symmetric structure generated by blind docking that involves the contact interfaces of all the Tectonin domains of GBP to CRP. This once again indicates the preference of specific Tectonin domains taking part in the GBP-CRP interaction. Figure 3.39 illustrates the highest scoring refined structure of the GBP-CRP model complex out of a pool of 400 generated dimers together with the corresponding energies and scores.

Taken together, these results show that GBP has distinct binding sites for LPS and CRP; the binding sites for LA and GlcNAc moieties of LPS are localized to the hydrophobic clefts interfacing propeller folds, specifically through GBP Tectonin

domains 6-1, whereas the binding sites for CRP are contained within the 4th Tectonin domain of GBP.

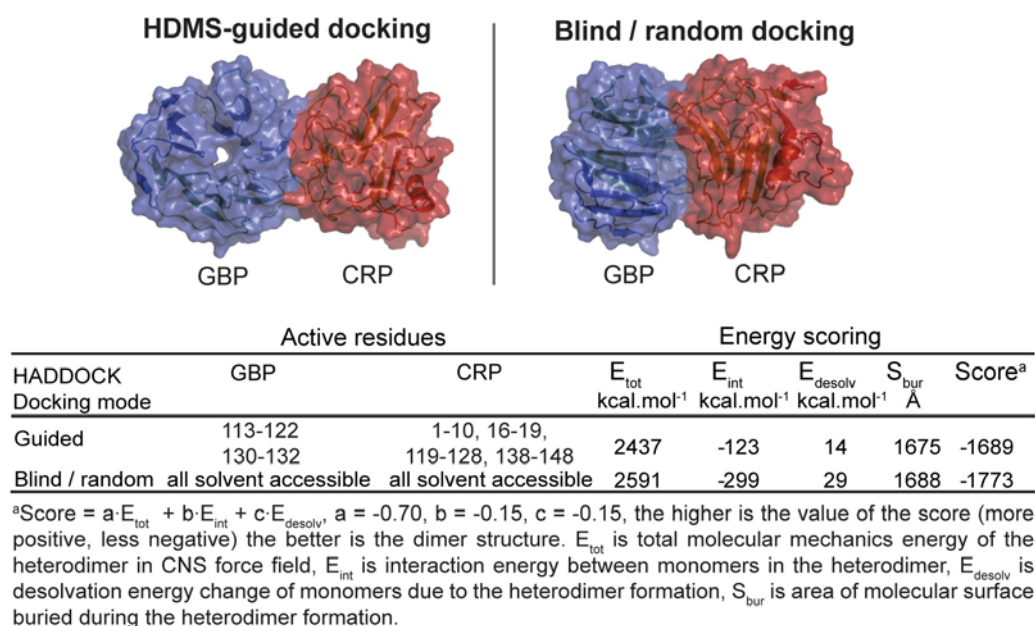


Figure 3.39: Guided docking of the GBP-CRP interaction. The GBP-CRP (blue and red respectively) interaction using HADDOCK was done with HDMS information to define active residues involved. Blind/random docking utilizing all solvent-accessible residues as active residues in the interaction was used for comparison. The accompanying table lists scores obtained for the best models from the guided and blind runs.

3.2.5 Effects of infection condition on the GBP, CRP and LPS interactions

3.2.5.1 Infection enhances interaction between GBP and CRP to LPS

GBP had been shown to interact with CRP only during infection (Ng et al. 2007), suggesting that certain infection conditions prime these molecules for interaction. We first tested if infection *in vivo* alters the affinity of GBP for LPS. Thus, GBP was isolated from animals that were either infected (GBPⁱ) with *P. aeruginosa*, or not infected (GBPⁿ), and used for binding with ReLPS and LA. Compared to GBPⁿ from naïve animals, the GBPⁱ showed increased affinity to ReLPS (K_D of 8.60 x 10⁻⁹ M) and to LA (5.11 x 10⁻⁸ M) (Figure 3.40). Besides having a higher binding affinity in the state of infection, the infection binding curves also exhibits a slower release rate (k_d rate) of LPS from GBPⁱ. This means that after binding to LPS, the GBPⁱ releases

LPS more slowly. These findings suggest that after initial recognition and binding to the sugar moieties, the adjacent chemical groups on the LPS molecule strengthens and enhances the anchorage of GBP onto the bacterium, leading to a slower rate of release from the protein.

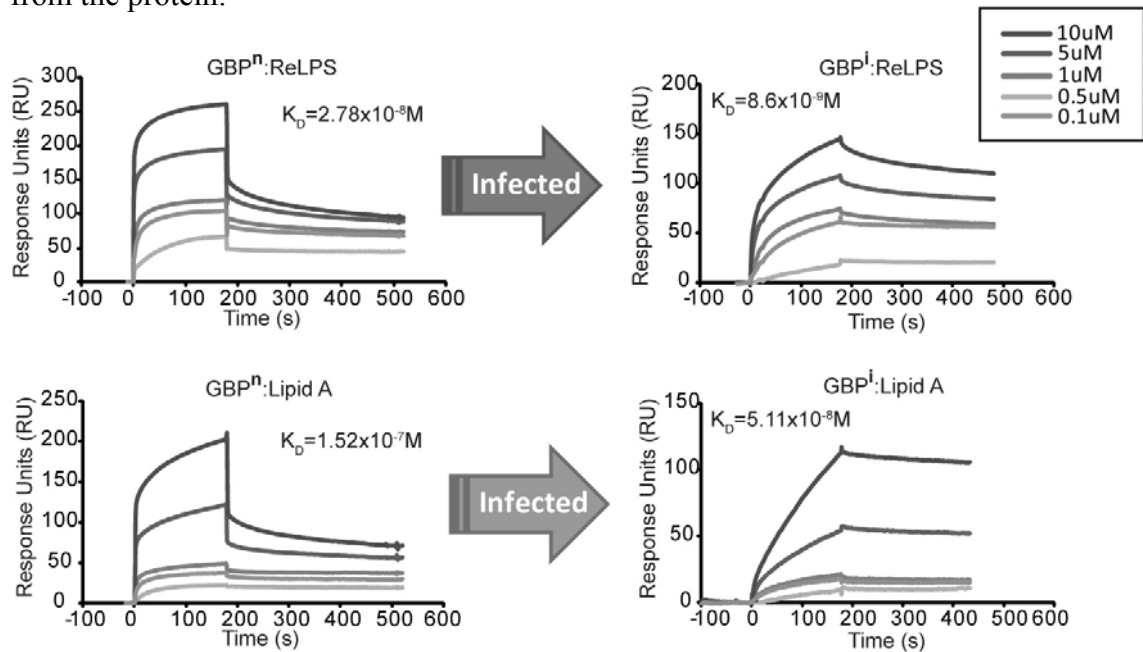


Figure 3.40: Binding affinity of infected GBP. Purified GBP isolated from infected animals were injected over surfaces immobilized with either ReLPS or LA and their binding affinity was quantified. Compared to GBP from naïve plasma (GBPⁿ; see Figure 3.32), GBP from the infected plasma (GBPⁱ) show increases in affinity for both the ReLPS ($K_D 8.6 \times 10^{-9} \text{ M}$) and LA ($K_D 5.11 \times 10^{-8} \text{ M}$).

Next, we determined if infection enhances interaction between GBP and CRP. To quantify the affinities between GBPⁿ:CRPⁿ and GBPⁱ:CRPⁱ, the purified proteins, CRPⁿ or CRPⁱ, were injected separately over HPA chips which have been previously immobilized with GBPⁿ or GBPⁱ bound to LA surfaces. As shown, the K_D between GBPⁿ and CRPⁿ was $2.10 \times 10^{-7} \text{ M}$ (Figure 3.41A) whereas the K_D between GBPⁱ and CRPⁱ was $1.66 \times 10^{-10} \text{ M}$, indicating that infection has enhanced the GBP-CRP affinity by 1000-fold (Figure 3.41B).

In the acute phase of infection, the invading microbes could usurp Ca^{2+} , causing a Ca^{2+} -depleted condition (Aslam et al. 2008). Therefore it was of interest to investigate whether cation depletion in general plays any role in the interaction between GBP and CRP. We measured the K_D between GBP^n and CRP^n in the presence of EDTA which depletes divalent cations from the solution. The K_D was 1.05×10^{-10} M, similar to that between GBP^i and CRP^i (Figure 3.41C). Next, we substituted EDTA for EGTA (a calcium chelator), and obtained similar affinity values (3.10×10^{-10} M) (Figure 3.41D). These results suggest that the depletion of Ca^{2+} ions may be a key factor in representing the state of infection, this condition enhances the GBP:CRP interaction, probably by excluding Ca^{2+} from binding to the same site in CRP that also interacts with GBP.

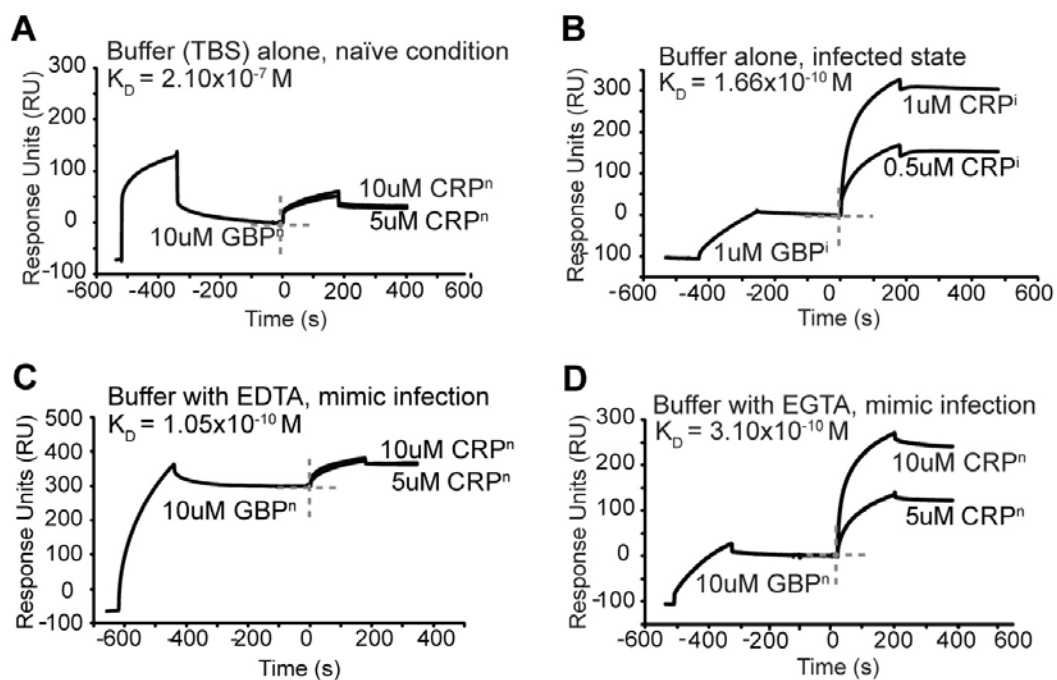


Figure 3.41: Effect of infection upon GBP-CRP interaction with LA. GBP was first bound to the surface immobilized with LA followed by injection of CRP. (A) GBP and CRP purified from naïve plasma (GBP^n , CRP^n) show a basal level of interaction (K_D of 2.1×10^{-7} M). (B) When proteins purified from infected plasma (GBP^i , CRP^i) were injected over the same LA-immobilised surface, the interaction affinity increased 1000-fold (K_D of 1.66×10^{-10} M). (C) When EDTA was added to deplete cations and simulate an infection condition between the two naïve proteins, it also produced a 1000-fold increase in affinity (K_D of 1.05×10^{-10} M) similar to the use of proteins from infected plasma. (D) Using EGTA as a more specific calcium chelator, we observed a similar degree of binding affinity as (C), with a K_D of 3.10×10^{-10} M.

3.2.5.2 Infection causes irreversible conformational change to GBP and CRP

To observe if the subsequent *in vitro* repletion of calcium is able to return the proteins to a naive state, we supplemented the infected proteins with a physiological level of 2.5 mM Ca^{2+} . There was no reduction in binding affinity, with the K_D remaining at 4.95×10^{-10} M (Figure 3.42) (condition without Ca^{2+} displayed K_D of 1.66×10^{-10} M). When the Ca^{2+} level was raised further to 10 mM, the K_D still remained at 2.58×10^{-10} M. Therefore, the replenishment of calcium does not alter the 1000-fold increased binding affinity between the two proteins. It is conceivable that infection has triggered an irreversible and dramatic conformational change in the proteins, positively inducing their interaction towards the formation of a pathogen-recognition interactome.

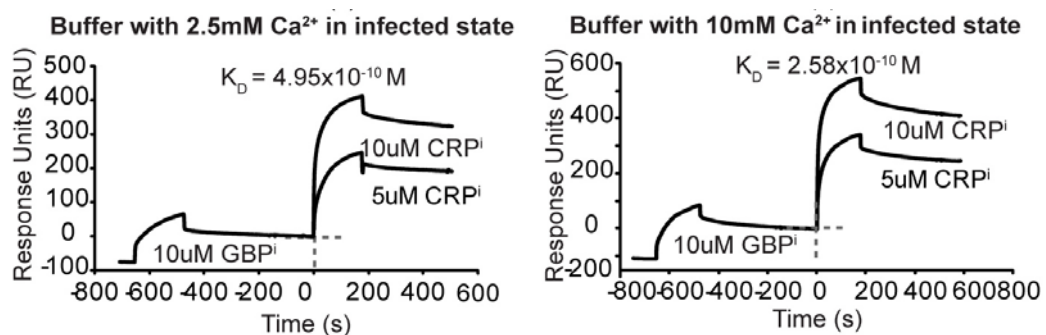


Figure 3.42: Effect of calcium on infected proteins binding to LA. GBP was first bound to the surface immobilized with LA followed by injection of CRP. GBP and CRP purified from infected plasma (GBP_i, CRP_i) show an interaction affinity of 4.95×10^{-10} M when the running buffer conditions was supplemented with the physiological level of Ca^{2+} , producing a similar affinity to that when Ca^{2+} was not supplemented. Increasing the Ca^{2+} concentration to 10mM did not alter the binding affinity significantly.

In earlier studies with other species, fluctuations in cation levels observed during infection were reported (Blackwell et al. 2000; Maguire 2006; Ong et al. 2006; Papp-Wallace and Maguire 2006) to affect protein-protein interactions and consequently, regulate the immune response. However, the work of Ng et al. (Ng et al. 2007) on the horseshoe crab suggested that plasma factors other than Ca^{2+} ions may enhance the interaction of GBP and CRP. We showed that the re-introduction of Ca^{2+} did not

return the binding affinity between GBPⁱ and CRPⁱ to the basal level of 10⁻⁷ M. This led us to postulate that while calcium depletion seems to represent the state of infection in the invertebrates or mammalian systems, the conformational changes in these plasma PRR proteins could be irreversible and that the resulting protein that has been structurally altered by binding to PAMPs (in this case the bacterial LPS) would likely recruit other PRRs (Le Saux et al. 2008) in the formation of the pathogen-recognition interactome, and trigger other downstream effectors for opsonisation by macrophages.

3.2.5.3 GBP and CRP binds and disrupts LPS micelles and exposes its

endotoxicity

Since GBP and CRP have been shown to bind LPS with high avidity, it was of interest to understand how they affect the endotoxicity of LPS. We utilized the PyroGene test kit (Lonza Inc.) to measure the endotoxicity of LPS. This kit uses recombinant Factor C (rFC), which is activated upon encountering LPS and hydrolyses a substrate to produce a fluorescent product, thereby reporting on as low as 0.01 EU/ml of LPS. Here, we firstly showed that individually, the purified GBP and CRP proteins were pyrogen-free (Figure 3.43). But when reacted with increasing doses of LPS (0.5-2.0 EU), we showed that the GBP and CRP bind LPS to progressively increase the endotoxicity of LPS (Figure 3.43D). This could be due to the disruption of the potential micellar form of LPS in solution (Aurell and Wistrom 1998; Li et al. 2004; Yu et al. 2006), exposing the endotoxic potency of the LPS and thus yielding a higher endotoxic activity. This phenomenon of increased endotoxicity has been observed before (Berger et al. 1991; Kaca and Roth 1995; Howe et al. 2008), where the binding of immune-related proteins to bacterial

structures was an initiating step to recruit other plasma factors or to trigger immune pathways such as the complement cascade and cytokines which eventually leads to the neutralization and elimination of the LPS. Increasing the protein concentration within the same EU produced only a small increase in the endotoxicity. We suspect that the amount of protein added initially was already sufficient in dispersing the LPS and the further addition of protein will not further increase endotoxicity.

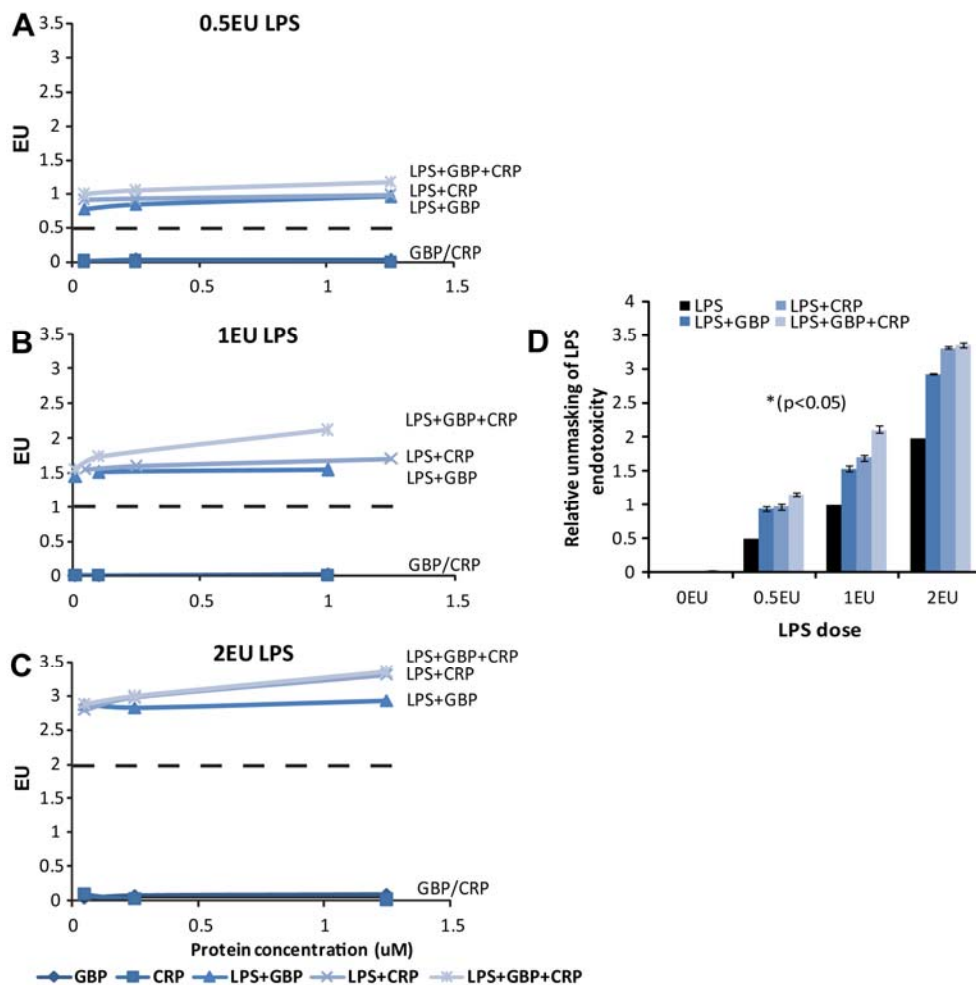


Figure 3.43: GBP and CRP exposes LPS endotoxicity. GBP cooperates with CRP with concentrations of LPS from (A) 0.5EU (B) 1EU (C) 2EU. (D) The unmasking effect becomes increasingly significant at higher concentrations of LPS due to the higher probability of micelle formation.

Furthermore, DLS experiments showed that upon the addition of LPS, GBP aggregates to form even larger molecular size structures. Correspondingly, its diffusion coefficient reduced, indicating that the rate of movement of the aggregates

has reduced (Figure 3.44). This phenomenon could represent an example of the aggregation of proteins to form a pathogen recognition complex upon bacterial detection in the host. We therefore suggest that *in vivo* GBP, together with CRP, are able to bind to the surface of the invading bacteria and disrupt the LPS layer on the outer membrane of the invading Gram-negative bacterium (Figure 3.45). This would lead to the exposure of the endotoxic effects of LPS to other immune factors in the host to stimulate downstream effectors for pathogen clearance.

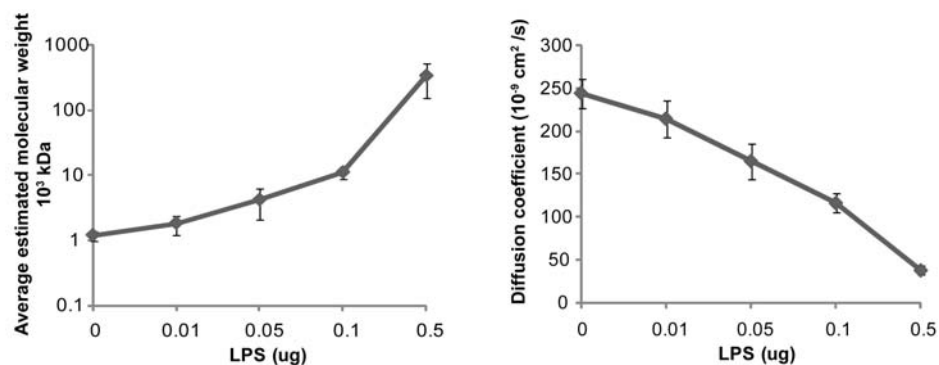


Figure 3.44: Molecular size of GBP polymers increase with addition of LPS. DLS showed that upon the addition of LPS, the molecular size in solution of GBP increases, as its diffusion coefficient decreases – an indication that a larger molecular size structure is formed and the rate of movement of this larger complex has reduced.

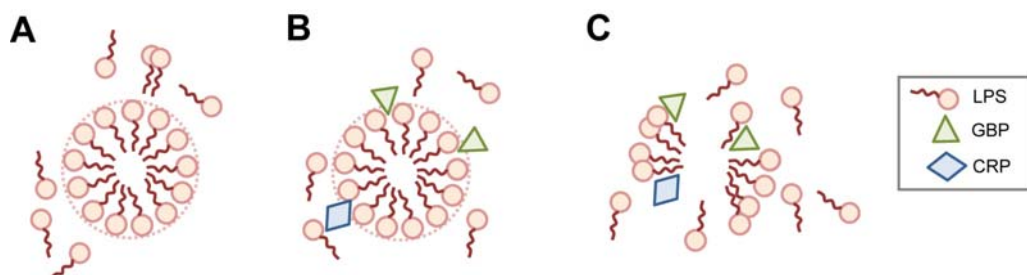


Figure 3.45: Model of micelle disruption to increase LPS endotoxicity. (A) Lipid A (LPS) tends to aggregate to form micelles in solution and could lower the levels of endotoxicity detection. (B) The introduction of GBP and CRP could disrupt the micellar structure. (C) The disruption of micelles exposes more individual LPS molecules which increases endotoxicity levels.

3.3 Summary

In summary, in view of the presence of binding sites for saccharide components of LPS located between the Tectonin domains and the preferential interaction of GBP

Tectonin domain 4 for CRP, it is conceivable that GBP plays a dual role in protein:PAMP (GBP:LPS) and protein:protein (GBP:CRP) interactions during infection. Even though the GBP molecule appears to be constituted by six iterative and apparently similar β -propeller Tectonin domains, these β -propellers are able to distinguish cognate PRRs and bacterial LPS. Together with our findings that GBP is also capable of disrupting LPS micelles and the increasing endotoxicity of LPS (most likely by micelle disruption) it is evident that Tectonin proteins like GBP, in collaboration with other PRRs, have (1) a part to play in pathogen (LPS) detection and (2) takes part in initiating the innate immune defense mechanism. Figure 3.46 summarizes the key findings of this chapter.

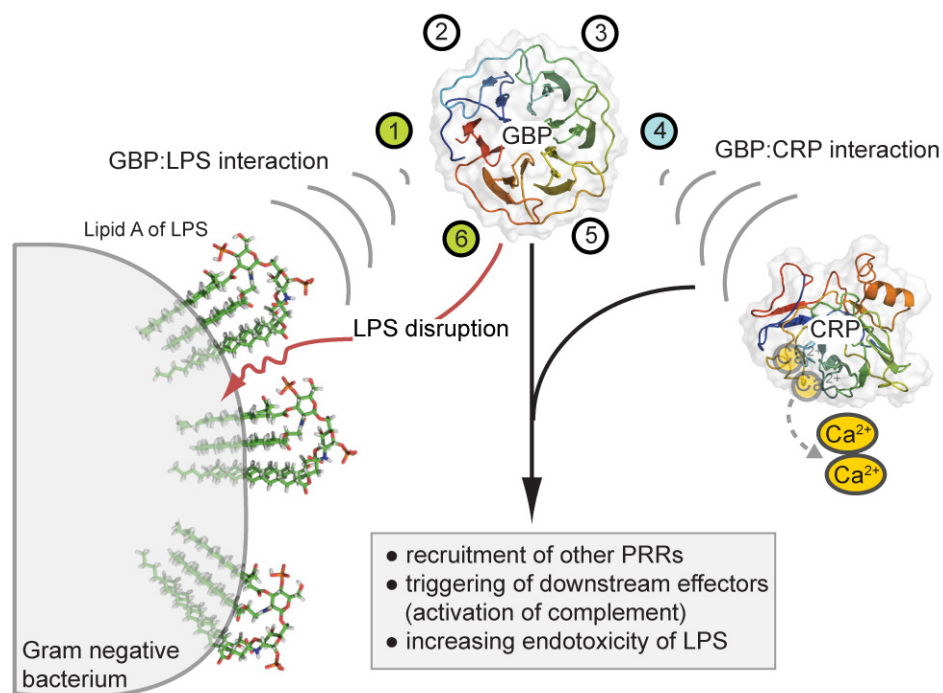


Figure 3.46: Proposed mechanism of GBP interactions and formation of the pathogen recognition complex for innate immune response. The GBP Tectonin domains 6-to-1 (green circles) bind the LA displayed on the Gram-negative bacterium, resulting in the disruption of the LPS micelles, increasing the endotoxicity of LPS, triggers other immune factors to neutralize and eliminate the bacteria. Tectonin domain 4 (blue circle) interacts with CRP, as determined by SPR, yeast 2-hybrid and HDMS experiments. The ensemble recruits other PRRs like carcinolectins (see Li et al, 2007; Ng et al, 2007) to further stabilize and form the antimicrobial complex to drive downstream effectors and complement activation pathways. The removal of Ca²⁺, possibly from CRP, greatly enhanced the GBP-CRP interaction affinity, positively inducing their interaction towards the formation of a pathogen recognition complex.

CHAPTER 4

hTECTONIN – DISCOVERY OF A NOVEL TECTONIN PROTEIN IN THE HUMAN

After characterizing and showing that the GBP Tectonin is an important member of the pathogen recognition interactome in the lower species, the next logical step was to examine if such Tectonins are structurally and functionally conserved in the mammalian species, in particular the human. Since the important partners of GBP, CRP and CL5, have homologous proteins in the human system, it is also likely that a homologous Tectonin protein exists.

4.1 Introduction

4.1.1 Are the Tectonin proteins evolutionary conserved? Do PRR:PRR interactomes exist in the human system?

Advances in sequence genomics have resulted in the accumulation of a large number of protein sequences derived from genome sequences. Although the human genome database has been completed a decade ago, about 50% of the human proteome still remain hypothetical as their functions are unknown. The elucidation of the functions of these hypothetical proteins can lead to additional protein pathways and revelation of new cascades completing our fragmentary knowledge on the proteome complex. Information on the network of protein–protein interactions will increase logarithmically. New hypothetical proteins may serve as disease markers and pharmacological targets.

The prime targets for the discovery of functional proteins are those which show relationship to lower species by way of sequence similarities, domain conservation or other inferences. Another approach is to look at lower species with no discovered homologs in the mammalian system. One example is this group of Tectonin domain-containing proteins. Tectonin domain-containing proteins, which belong to a subclass of proteins of the larger β -propeller family, have only thus far been studied in the invertebrates, namely the horseshoe crab, the slime mold and the sponge.

An exhaustive search in the databases for vertebrate proteins revealed no amino acid sequence homologs of Tectonin domain-containing proteins, prompting us to postulate that perhaps these Tectonin domain-containing proteins (henceforth referred to as Tectonin proteins) have evolved through the species. There are many examples of other family of proteins in meiosis-related proteins, kinetochores, cell gap contacts and nuclear pore complexes which show no homology at amino acid sequence level. However, they are hinted at conservation of the domain architecture organization and three-dimensional structure of functionally important domains in proteins in the budding yeast, nematode, *Drosophila*, *Arabidopsis*, and human. Several databases like SCOP, CATH, SMART, which also employ domain and secondary structure classification for structure sorting and function prediction were used to search for β -propeller structures and possibly distance relationships by domain conservation. This is especially useful when searching for related proteins with low sequence homology or when sequences have adapted through evolution from the invertebrates to the mammals. We thus sought to identify Tectonins in the vertebrates and examine their potential functional conservation in host-pathogen interactions.

In this chapter, we present evidence for the identification and characterization of a novel human Tectonin-domain containing protein which we named the hTectonin. We discovered that although no Tectonin proteins showing close sequence homology were identified in the vertebrates, they are distantly related phylogenetically through their Tectonin domain architecture. hTectonin was found to be expressed in immune-related cell lines. By screening a human leukocyte library for hTectonin interaction partners, we found some interesting proteins which function in immunity. We also showed that the hTectonin retained features shown in its invertebrate counterparts by having LPS-binding motifs of similar pattern. We then compared peptides derived from hTectonin and GBP and showed that indeed, the Tectonin domains within the protein harbouring the LPS-binding sequence motifs are able to bind LPS with similarly high affinity, suggesting a conserved function and potential role for Tectonins in innate immunity.

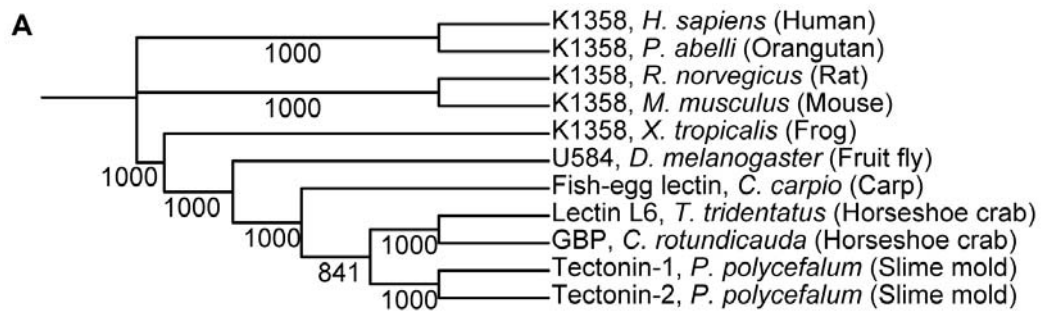
4.2 Results and Discussion

4.2.1 hTectonin – a distantly related homolog of the invertebrate Tectonins

4.2.1.1 hTectonin is widely present across the various species

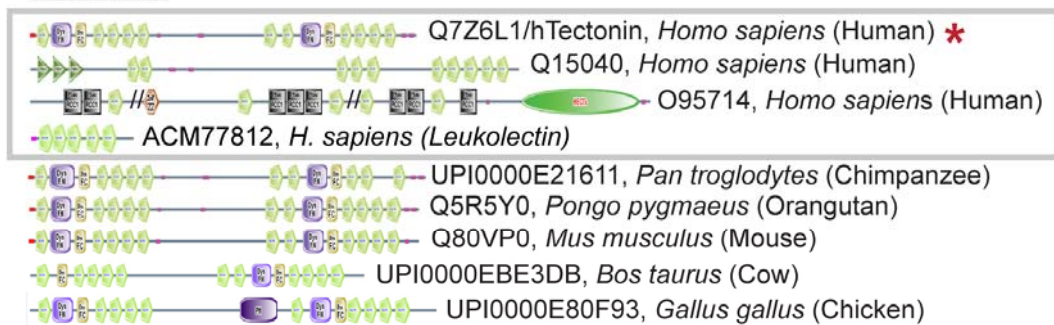
In mammals, the identity and role of proteins with Tectonin domains are unknown. Those identified or studied in the invertebrates exhibit immune defense properties. Here, we sought to examine whether the Tectonin domains are structurally and functionally conserved in the mammals. A position-iterated search using known Tectonin proteins in the invertebrates revealed a group of vertebrate Tectonin proteins to be distantly related (Figure 4.1A). Interestingly, all the hits that were considered related came from the same family of proteins, coding for an unknown protein named KIAA1358. A further check using SMART domain analysis revealed that the human protein in this family, assigned Q7Z6L1, is one of only 3 human proteins (Q7Z6L1, Q15040 and O95714) that contains the Tectonin domain architecture (Figure 4.1B). Q7Z6L1 codes for a predominantly Tectonin domain-containing protein, suggesting that the domains probably form a significant part of the molecular structure and play an essential role in these proteins.

A more in-depth search using Q7Z6L1 as the query sequence resulted in the discovery that it exists in numerous other species with significantly high homology (Figure 4.2). This high homology of Q7Z6L1, from the worm to the human, prompted us to suggest its evolutionary conservation and functional significance. Large scale mRNA study on cancer tissues have also shown that the Q7Z6L1 gene (DKFZp434B0335) is up-regulated in prostate cancer (Bull et al. 2001), and is downregulated when the cancer growth is suppressed by antigen inhibitors (Nickols and Dervan 2007). These data led us to suggest that Q7Z6L1 might have an immuno-regulatory function.



B Tectonin domain-containing proteins

Vertebrates



Invertebrates

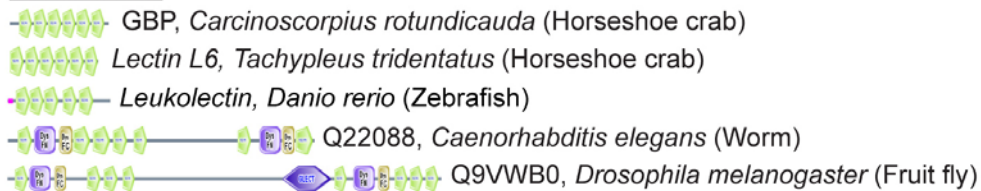


Figure 4.1: hTectonin is distantly related to the invertebrate Tectonins. (A) The phylogenetic tree constructed after a PSI-Search query using the invertebrate Tectonins revealed K1358 family of proteins as closely related Tectonin domain containing proteins in the mammals and also in lower species like the frog. The numbers at the nodes are an indication of the level of confidence for the branches as determined by bootstrap analysis (1000 bootstrap replicates). (B) Bioinformatics domain analysis utilizing SMART (Schultz et al. 1998; Letunic et al. 2009) shows existence of Tectonin domain-containing proteins both in invertebrates and vertebrates from the horseshoe crab lectins, worm, up to humans. Of interest in this study is the protein hTectonin (red asterisk) which appear to have homologues in other species as well, for example in *P. troglodytes* (chimpanzee), *P. pygmaeus* (orangutan), *M. musculus* (mouse), *G. gallus* (chicken), *C. elegans* (worm) and *D. melanogaster* (fruitfly).

Furthermore, Miftari and Walther (2009) recently discovered a Tectonin domain-containing protein similar to the ones studied in the vertebrates in the human leukocyte which they have named the leukolectin. All these findings imply that Tectonin domain-containing proteins play key roles in immunity. We thus dubbed

Q7Z6L1 as ‘hTectonin’ and we selected this protein for further structural and functional analyses.

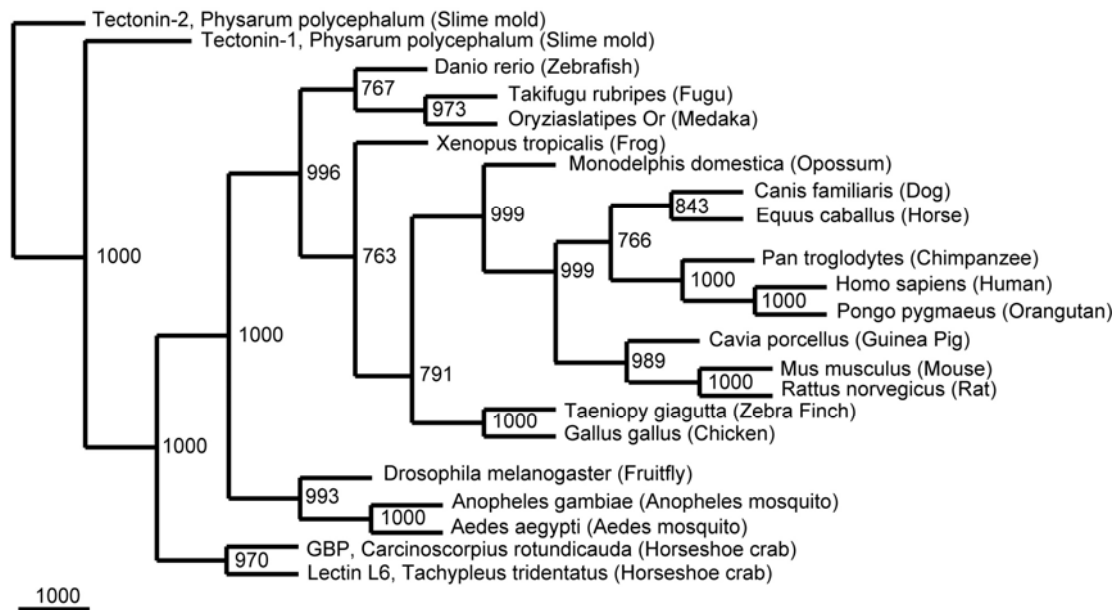


Figure 4.2: hTectonin gene is widespread across many species. The phylogenetic tree of hTectonin homologues constructed by primary sequence similarity shows its prevalence and conservation among a vast number of different species, right down to the worm, *C. elegans*. The human hTectonin protein was used as a query sequence in BLAST. Top hits were then compiled and multiple sequence alignment based on a guide tree was done using CLUSTALW (Thompson et al. 2002) and the alignment was edited with Jalview (Waterhouse et al. 2009). The tree was constructed using the neighbour joining algorithm of the PHYLIP package (Retief 2000). The numbers at the nodes are an indication of the level of confidence for the branches as determined by bootstrap analysis (1000 bootstrap replicates).

4.2.1.2 hTectonin β -sheets are highly conserved

From the multiple sequence alignment of the Tectonin domains, we confirmed a pattern of sequence repeats of about 40 to 50 residues each, which is a unique characteristic of β -propellers (see Figure 3.19). In addition, secondary structure prediction of hTectonin by PSIPRED predicted these conserved repeats to form the β -strands of a β -sheet topology, consistent with β -propeller architecture (Figure 4.3). This further supports that the hTectonin is highly likely to be a bona fide Tectonin domain protein.

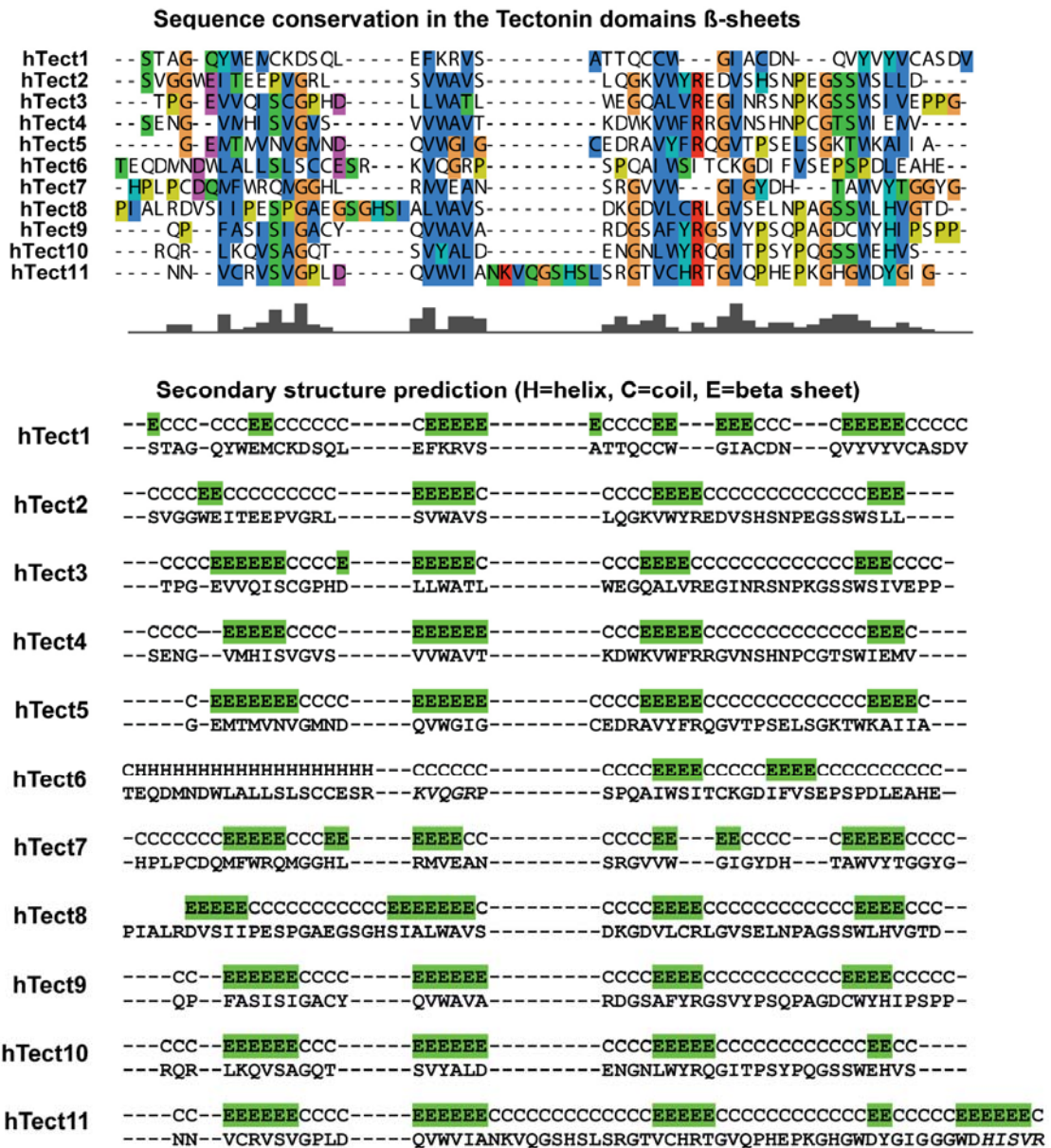


Figure 4.3: hTectonin forms β -sheets in its Tectonin repeats. CLUSTALW alignment of all the individual Tectonin domains in hTectonin and PSIPRED secondary structure prediction indicates that the 11 Tectonin domains of hTectonin contain 4 highly conserved repeats that form β -strands (highlighted in green), a motif that is characteristic of the β -propeller fold. E, β -sheet; C, Coil; H, Helix.

4.2.2 In search for interaction partners of hTectonin

4.2.2.1 hTectonin gene is expressed in immune cell lines

Since there are some evidence suggesting an immune role for hTectonin or its homologues, coupled with the discovery of leukolectin in the human leukocyte (Miftari and Walther 2009), we sought to screen immune cell lines and the human

leukocyte cDNA library to search for hTectonin gene expression. Firstly, we checked and confirmed that the hTectonin gene is indeed expressed in the A549 (lung epithelial cells), U937 (monocytes) cell lines and the human leukocytes. Figure 4.4 shows the expression competency of hTectonin thus corroborating its immune relevance.

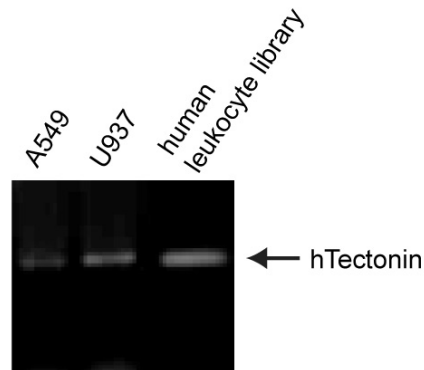


Figure 4.4: hTectonin cDNA is found in the human T-cells (A549), monocytes (U937) and leukocytes. For A549 and U937 cells, RNA was extracted using TRIzol reagent followed by RT-PCR using hTectonin-specific primers. For the human leukocyte cDNA library, PCR on the library cDNA using hTectonin-specific primers was done.

4.2.2.2 hTectonin interacts with immune-related molecules in the leukocyte

library

Earlier, we have shown that the hTectonin gene is expressed in the leukocyte cDNA library. Since Miftari and Walther (2009) discovered the leukolectin, we sought to screen the leukocyte library to search for potential partners of hTectonin. By yeast 2-hybrid library screening, we searched for other proteins in the leukocyte cDNA library that interacted with hTectonin. A total of 32 independent clones were isolated and sequenced (Table 4.1).

Several clones with potentially interesting functions related to the immune system - neutrophil cytosol factor 1 (NCF1), Src-like adaptor 2 (SLAP-2), Ubiquitin-specific-processing protease CYLD and the EDAR (Ectodysplasin A receptor)-associated

death domain protein (EDAR-ADD) - were selected and their interactions confirmed (Figure 4.5). NCF1 is involved in superoxide release and cellular defense response (Volpp et al. 1989). Defects in NCF1 are often characterized by the inability of neutrophils and phagocytes to kill microbes they have ingested (Casimir et al. 1991; Noack et al. 2001). SLAP-2 is expressed in immune-related tissues and is involved in B-cell mediated immunity, T-cell activation and intracellular receptor mediated signaling pathway (Holland et al. 2001). CYLD is a negative regulator of TRAF-2 and NF-kappa-B signaling pathway, and interacts with NEMO, TRAF-2 and TRIP (Brummelkamp et al. 2003; Kovalenko et al. 2003; Regamey et al. 2003; Trompouki et al. 2003). EDAR-ADD, through its interaction with EDAR, acts as an adaptor, and links the receptor to downstream signaling pathways (Kumar et al. 2001).

Together with literature findings, these evidence point to the probable function of hTectonin as an immune-related protein. Moreover with leukolectin being found expressed in the blood (human leukocyte), it is not hard to postulate that Tectonin proteins in general are immuno-regulatory proteins.

Table 4.1 Putative interaction partners of hTectonin identified via yeast 2-hybrid screening of the human leukocyte cDNA library
(continued on next page).

Accession	Name	Function	Tissue
P04899	Guanine nucleotide-binding protein G(i), alpha-2 subunit	Guanine nucleotide-binding proteins (G proteins) are involved as modulators or transducers in various transmembrane signaling systems. The G(i) proteins are involved in hormonal regulation of adenylate cyclase: they inhibit the cyclase in response to beta-adrenergic stimuli.	Kidney, and mammary gland
Q7Z419	p53-inducible RING finger protein; E3 ubiquitin-protein ligase RNF144B	E3 ubiquitin-protein ligase which accepts ubiquitin from E2 ubiquitin-conjugating enzymes UBE2L3 and UBE2L6 in the form of a thioester and then directly transfers the ubiquitin to targeted substrates such as LCMT2, thereby promoting their degradation. Induces apoptosis via a TP53/p53-dependent but caspase-independent mechanism.	Skeletal muscle, placenta
NP_060554	Hook-related protein	Bind to microtubules and organelles through their N- and C-terminal domains, respectively. The encoded protein localizes to discrete punctuate subcellular structures, and interacts with several members of the Rab GTPase family involved in endocytosis. It is thought to link endocytic membrane trafficking to the microtubule cytoskeleton.	Brain
Q15560	Transcription elongation factor A protein 2	Necessary for efficient RNA polymerase II transcription elongation past template-encoded arresting sites. The arresting sites in DNA have the property of trapping a certain fraction of elongating RNA polymerases that pass through, resulting in locked ternary complexes. Cleavage of the nascent transcript by S-II allows the resumption of elongation from the new 3'-terminus.	Testis, ovary
NP_054883	FXYD domain-containing ion transport regulator 5	Uncharacterized. Mouse FXYD5 termed RIC (related to ion channel). FXYD1 (phospholemman), FXYD2 (gamma), FXYD3 (MAT-8), FXYD4 (CHIF), and FXYD5 (RIC) have been shown to induce channel activity in experimental expression systems.	
P21912	Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial	Iron-sulfur protein (IP) subunit of succinate dehydrogenase (SDH) that is involved in complex II of the mitochondrial electron transport chain and is responsible for transferring electrons from succinate to ubiquinone (coenzyme Q).	Brain, liver

P19634	Sodium/hydrogen exchanger 1	Involved in pH regulation to eliminate acids generated by active metabolism or to counter adverse environmental conditions. Major proton extruding system driven by the inward sodium ion chemical gradient. Plays an important role in signal transduction.	Kidney and intestine.
Q9H6Q3	Src-like-adaptor 2 isoform a	Adapter protein, which negatively regulates T-cell receptor (TCR) signaling. Inhibits T-cell antigen-receptor induced activation of nuclear factor of activated T-cells. May act by linking signaling proteins such as ZAP70 with CBL, leading to a CBL dependent degradation of signaling proteins.	Thymus, Hapatoma, Prostate
P10253	Lysosomal alpha-glucosidase	Essential for the degradation of glygogen to glucose in lysosomes.	Placenta, Testis, Platelet
P05771	Protein kinase C beta	This is a calcium-activated, phospholipid-dependent, serine- and threonine-specific enzyme. PKC is activated by diacylglycerol which in turn phosphorylates a range of cellular proteins. PKC also serves as the receptor for phorbol esters, a class of tumor promoters. May be considered as a novel component of the NF-kappa-B signaling axis responsible for the survival and activation of B-cells after BCR cross-linking.	Hippocampus, Platelet, Fetal brain
Q9NTJ4	Alpha-mannosidase 2C1	Hydrolysis of terminal, non-reducing alpha-D-mannose residues in alpha-D-mannosides.	Testis, uterus, tonsils, epithelium
P43490	Nicotinamide phosphoribosyltransferase (Pre-B-cell colony-enhancing factor 1)	Catalyzes the condensation of nicotinamide with 5-phosphoribosyl-1-pyrophosphate to yield nicotinamide mononucleotide, an intermediate in the biosynthesis of NAD. It is the rate limiting component in the mammalian NAD biosynthesis pathway	Expressed in large amounts in bone marrow, liver tissue, and muscle. Also present in heart, placenta, lung, and kidney tissues.
Q9BSJ2	Gamma-tubulin complex component 2	Gamma-tubulin complex is necessary for microtubule nucleation at the centrosome.	Ubiquitously expressed

Q92982	Ninjurin-1; Nerve injury-induced protein 1	Homophilic cell adhesion molecule that promotes axonal growth. May play a role in nerve regeneration and in the formation and function of other tissues. Cell adhesion requires divalent cations.	Widely expressed in both adult and embryonic tissues, primarily those of epithelial origin.
P54762	Ephrin type-B receptor 1	Receptor for members of the ephrin-B family. Binds to ephrin-B1, -B2 and -B3. May be involved in cell-cell interactions in the nervous system.	Brain
P14598	Neutrophil cytosol factor 1	NCF2, NCF1, and a membrane bound cytochrome b558 are required for activation of the latent NADPH oxidase (necessary for superoxide production). Defects in NCF1 are the cause of chronic granulomatous disease autosomal recessive cytochrome-b-positive type 1 (CGD1) [MIM:233700]. Chronic granulomatous disease is a genetically heterogeneous disorder characterized by the inability of neutrophils and phagocytes to kill microbes that they have ingested. Patients suffer from life-threatening bacterial/fungal infections.	Widely expressed.
P47914	60S ribosomal protein L29	Cell surface heparin-binding protein HIP	Brain
P51511	Matrix metalloproteinase-15	Endopeptidase that degrades various components of the extracellular matrix. May activate progelatinase A.	Placenta
O95793	Double-stranded RNA-binding protein Staufen homolog 1	Binds double-stranded RNA (regardless of the sequence) and tubulin. May play a role in specific positioning of mRNAs at given sites in the cell by cross-linking cytoskeletal and RNA components, and in stimulating their translation at the site. Binds tubulin. Binds with low affinity single-stranded RNA or DNA homopolymers. Interacts with CASC3 in an RNA-dependent manner. Interacts with the influenza virus nonstructural protein NS1.	Widely expressed. Brain and placenta included.
P49023	Paxillin	Cytoskeletal protein involved in actin-membrane attachment at sites of cell adhesion to the extracellular matrix (focal adhesion). Binds in vitro to vinculin as well as to the SH3 domain of c-SRC and, when tyrosine phosphorylated, to the SH2 domain of V-CRK.	Placenta, Epithelium, T-cell

Q16236	Nuclear factor erythroid 2-related factor 2 (NFE2)	Transcription activator that binds to antioxidant response (ARE) elements in the promoter regions of target genes. Important for the coordinated up-regulation of genes in response to oxidative stress. Heterodimer. May bind DNA with an unknown protein. Interacts with KEAP1. Interacts via its leucine-zipper domain with the coiled-coil domain of PMF1.	Widely expressed.
Q9NQC7	Ubiquitin-specific-processing protease CYLD	Negative regulator of TRAF2 and NF-kappa-B signaling pathway. Has deubiquitinating activity that is directed towards non-'Lys-48'-linked polyubiquitin chains. The inhibition of NF-kappa-B activation is mediated at least in part, by the deubiquitination and inactivation of TRAF2 and, to a lesser extent, TRAF6. Interacts with NEMO, TRAF2 and TRIP.	
Q15269	Periodic tryptophan protein 2 homolog (PWP2)	Unknown function, contains 14 WD propeller repeats	Muscle
NP_542776	EDAR-associated death domain	Interacts with EDAR, a death domain receptor known to be required for the development of hair, teeth and other ectodermal derivatives. Through its interaction with EDAR, this protein acts as an adaptor, and links the receptor to downstream signaling pathways. Two alternatively spliced transcript variants of this gene encoding distinct isoforms have been reported.	
P62328	Thymosin beta-4 (THYB4)	Binds to and sequesters actin monomers (G actin) and therefore inhibits actin polymerization. Expressed in several hemopoietic cell lines and lymphoid malignant cells. Decreased levels in myeloma cells. Decreased levels in THP-1 cells after treatment with recombinant interferon-lambda.	Expressed in several hemopoietic cell lines and lymphoid malignant cells.
O75348	V-type proton ATPase subunit G 1 (ATP6G)	Catalytic subunit of the peripheral V1 complex of vacuolar ATPase (V-ATPase). V-ATPase is responsible for acidifying a variety of intracellular compartments in eukaryotic cells.	Ubiquitous.
P61916	Niemann-Pick disease type C2 (NPC2)	Defects in NPC2 are the cause of Niemann-Pick disease type C2 (NPC2). NPC2 is a fatal autosomal recessive hereditary disease characterized by the accumulation of low-density lipoprotein-derived cholesterol in lysosomes.	Epididymis

Selected hTectonin library screened partners for interaction confirmation

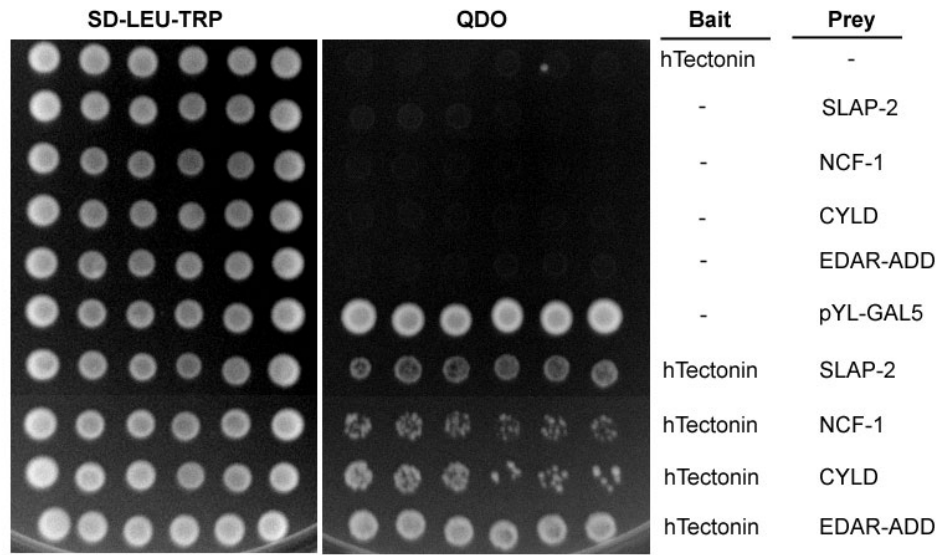


Figure 4.5: Selected hTectonin partners for interaction confirmation. Several immune-related partners of hTectonin were selected and their interaction confirmed using yeast 2-hybrid co-transformation.

4.2.2.3 hTectonin interacts with ficolin

Based on the rationale that (i) hTectonin is an architectural homolog of GBP and (ii) GBP interacts with CL5 which is a homolog of the human ficolin, we had expected that the initial library screen would have picked up ficolin as an interaction partner. However, we did not manage to do so. This could be attributed to the incompleteness of a library screen where statistically, any one pool of interactions might not contain all interaction partners. Nevertheless, based on the homologs in the horseshoe crab, we decided to perform yeast 2-hybrid analysis using full-length hTectonin and its subclones as the bait and M-ficolin, which is also found in the monocytes, as the prey. Results showed that indeed, the hTectonin interacts specifically with M-ficolin (Figure 4.6). M-ficolin has in turn been shown to interact with CRP. Since both CRP and M-ficolin are key proteins of the complement classical and lectin pathways, respectively, this is the first evidence for the potential function of a human Tectonin domain-containing protein in frontline immune defense.

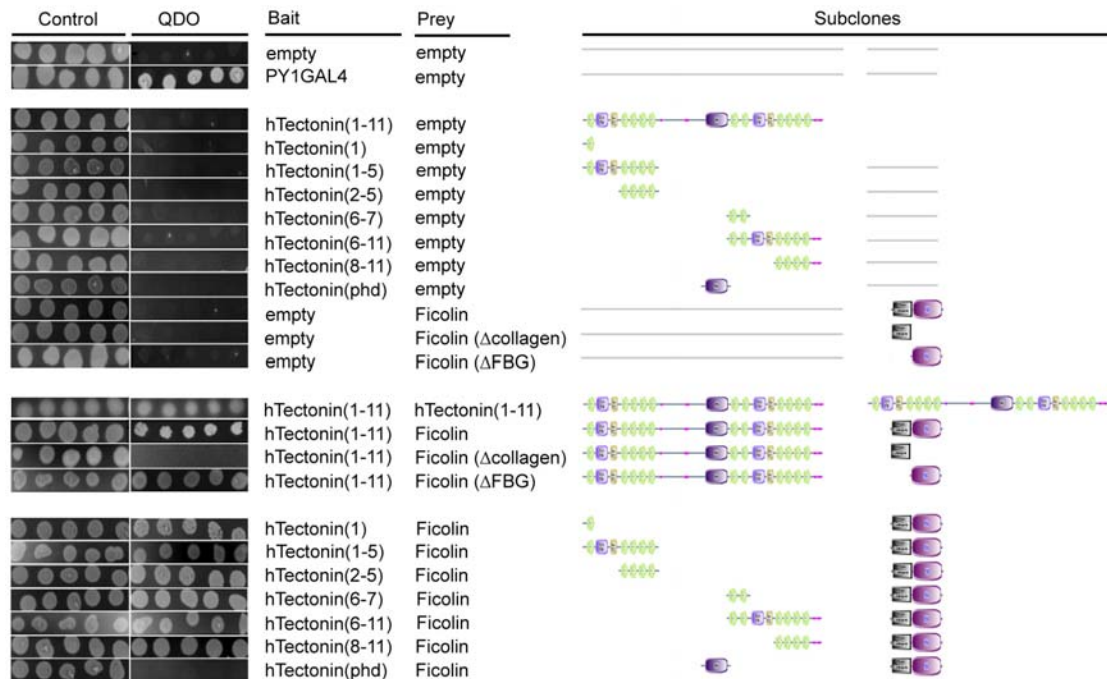


Figure 4.6: hTectonin interacts with ficolin. Yeast 2-hybrid shows that hTectonin interacts (i) with itself, suggesting the possibility of oligomerization, as observed in other beta-propeller proteins; and (ii) with ficolin, a human complement protein. Furthermore, interaction with ficolin specifically occurs through the Tectonin domains of the hTectonins. This demonstrates a possible functional conservation of Tectonin domains since the tectonin domains of GBP was shown to interact with CL-5, a homolog of ficolin.

Further delineation of hTectonin to isolate its functional domains showed that only the sub-clones expressing the predicted Tectonin domains interacted with M-ficolin. Furthermore, only the fibrinogen-like (FBG) domain of M-ficolin was shown to interact with the hTectonin, concurring with recent findings that the FBG domain is responsible for ligand-binding. These results suggest that the protein-protein interaction between the hypothetical hTectonin and M-ficolin is not random, but structurally and positionally specific, and that the hTectonin is potentially involved in immune regulation, acting through its Tectonin domains.

4.2.2.4 hTectonin protein expression increases in response to LPS stimulation

In order to if hTectonin had any immune response activity, we test both its mRNA and protein expression under LPS stimulation. Two immunoregulatory cell lines –

Jurkat (T lymphocytes) and U937 (monocytes) were used in this study, and cells were stimulated with *S. minnesota* LPS. mRNA and protein profiles were observed at 6 hpi. We noticed that the mRNA expression of hTectonin did not change post infection (Figure 4.7). However, the hTectonin protein expression increased post-infection. This increase takes place not in Jurkat cells but in U937 cells which contains TLR4, a known LPS receptor. This lead us to predict that hTectonin might possibly be responding to LPS signals through a TLR4 signalling mechanism. Also, NCF-1, a potential interaction partner of hTectonin discovered via yeast 2-hybrid screening showed expression in U937 cells only under infection condition. This again shows immune-responsive behavior of these proteins.

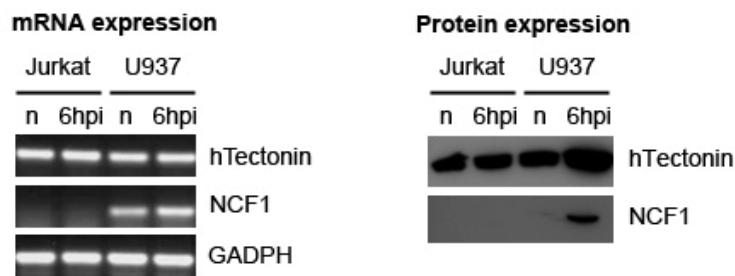


Figure 4.7 : hTectonin responds to LPS stimulation. (Left) mRNA profile of hTectonin remained constant even in infection in both Jurkat and U937 cells. NCF-1, expressed only in U937, showed a slight increase in mRNA expression 6hpi. GADPH is used as a control. (Right) hTectonin protein expression increased 6hpi in U937 cells, which display the TLR4 receptor. NCF-1 was observed to be expressed only when infected.

4.2.3 LPS-binding peptides in Tectonins

Apart from being able to interact with host immune proteins, we were also interested to know if hTectonin, like GBP, is able to bind PAMPs. This rationale stems from our earlier HDMS experiments (Section 3.2.2) where we identified the regions of GBP which interacted with LA. We also observed in GBP that the particular group of residues forming the surface has the LPS-binding pattern of *BHB(P)HB*, which was earlier defined and studied by Freceer et. al. 2000, as the minimal motif for LPS-binding. Since GBP and hTectonin share similar Tectonin domains, we postulate that

hTectonin might also have such motif in its protein sequence that can confer LPS-binding property.

4.2.3.1 An algorithm was developed for large-scale screening of LPS-binding motifs in proteins

Because hTectonin is 1165 residues long, it would be more efficient to design an algorithm to screen sequences instead of manually searching the protein sequence for the LPS-binding motif. Furthermore, it will be able to do large-scale screen of multiple protein sequences simultaneously.

The program LPSMotif (Figure 4.8) was designed to accept a FASTA formatted file with a list of protein sequences from the user, and checks if it contains the LPS-binding motifs of *BHBHB*, *BHPHB*, or the extended *BHBHBHB*. The protein sequence is read, and each residue converted into its property group (either B=basic, H=hydrophobic, P=polar or X=all other groupings). Once the sequence has been converted into a string of B, H, P and Xs, the entire sequence is scanned to search for the defined pattern. The algorithm also included an option for user-defined patterns if slight deviations in motifs were required. If a motif was found, the corresponding amino acid sequence will be displayed output together with the residue position in protein sequence and average hydrophobicity value (Figure 4.9).

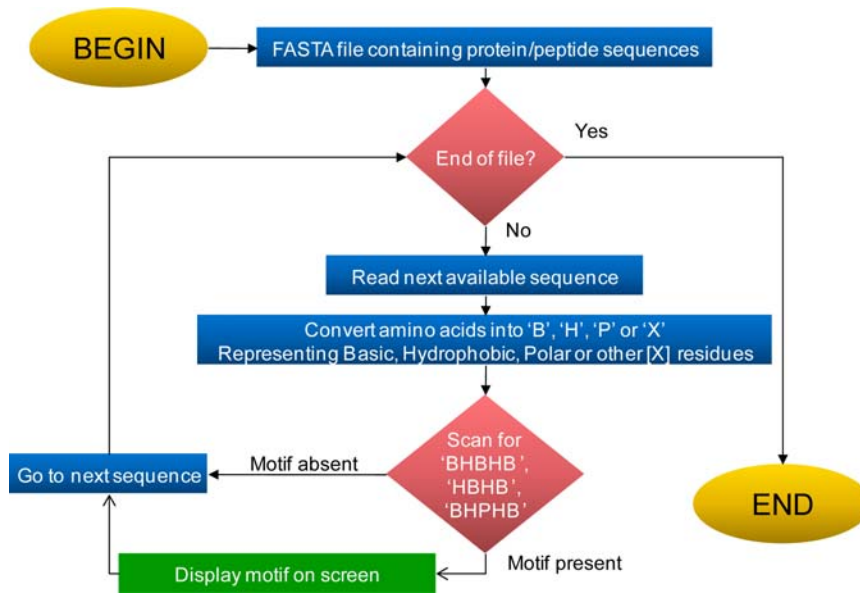


Figure 4.8: Algorithm workflow of the LPSMotif search application. The program works by first checking if the file is empty. If the input file is empty, the program will terminate. If not, it will read the next available sequence and starts converting the protein sequence according to its property (either B=basic, H=hydrophobic, P=polar, or X=for others). Next, it will scan along this converted sequence for the desired LPS motif. If there are no such motifs in the sequence, the program will go to the next protein sequence available. If the motif is present, it will be displayed on the program screen along with the information on the detected motif.

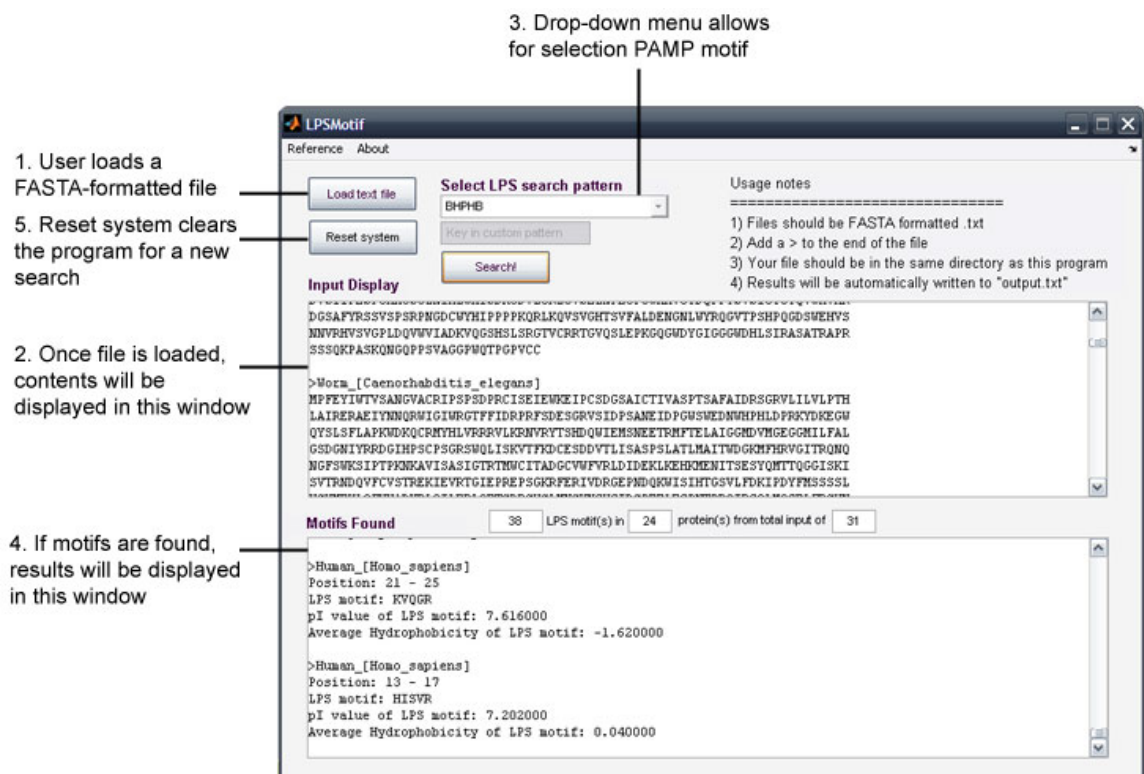


Figure 4.9: The LPSMotif program. A representative screen from LPSmotif shows the input and output options of the program and their general usage description.

4.2.3.2 hTectonin contains LPS-binding motifs

Proteins harboring LPS-binding motifs, with alternating basic-hydrophobic/polar residues, have been shown to bind LPS via the LA moiety, which is the most conserved bioactive pathophysiological centre of the LPS molecule.

Earlier, we have shown that GBP, our model Tectonin protein, is able to specifically bind LPS via sugar moieties and that HDMS results revealed that the site of LA interaction containing the LPS-binding motif of *BHPHB* corresponding to the amino acid sequence of *HINGK* (Figure 3.35A). This prompted us to hypothesize that hTectonin might harbor this motif as well. Using the LPSMotif application, we have managed to identify two such motifs in the 6th and 11th Tectonin domains of the hTectonin, namely “*KVQGR*” and “*HISVR*” (Figure 4.10). LPSMotif also reconfirmed the detection of the motif in GBP, “*HINGK*”, as discovered experimentally in Chapter 2. Moreover, such patterns are also found in the hTectonin homologs in other species (Figure 4.11, boxed), which is another clue that such LPS-binding motifs might consistently exist in Tectonin proteins.



Figure 4.10: LPS-binding motifs in Tectonin proteins. GBP and hTectonin harbor the *BHB(P)HB* pattern. GBP contains the “*HINGK*” sequence of amino acids in its 1st Tectonin domain whereas hTectonin contains the “*KVQGR*” and “*HISVR*” sequence of amino acids in its 6th and 11th (hTectonin) Tectonin domains respectively.

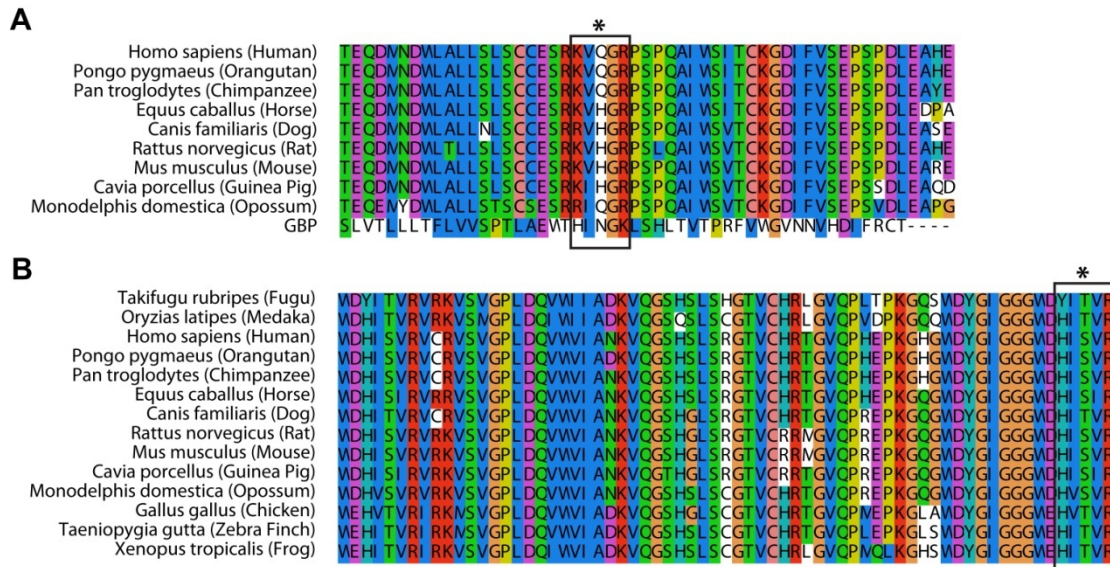


Figure 4.11: hTectonin LPS-binding motifs conserved in other species. The LPS-binding motif of the pattern *BHB(P)HB* (asterisk, boxed) in (A) hTectonin6 and (B) hTectonin11 - is well conserved in other species.

4.2.3.3 Designed predicted LPS-binding peptides are hydrophilic, synthetically feasible and suitable for binding analysis

In order to validate the computationally predicted LPS-binding motifs in the hTectonin domains, we synthesized representative Tectonin peptides derived from domains 6 and 11 of hTectonin (named hTect6 & hTect11), and the efficacy of binding of LA was compared with peptides derived from the GBP Tectonin domains 6 & 1 (named GBP6-1), with which the LPS-binding sequences peptides were previously identified via HDMS.

For GBP, an additional peptide was designed - named GBP6-1(tail) – that included the C-terminal “tail” region which does not form the major 6 Tectonin β -propeller structure of GBP (see Figure 3.22). This peptide was included to test if the LPS-binding function can be purely conferred to the Tectonin domain regions.

The peptides were designed to be approximately 20-25 residues in length, while retaining the LPS-binding motif within the sequence and minimizing the hydrophobic content. Peptides from the Tectonin domains of GBP and hTectonin that did not contain the LPS-binding motif were also designed to serve as controls (named GBP3, hTect1, hTect8; after the domains they were derived from). Based on the hydrophobicity plots (Figure 4.12), the final peptides designed were hydrophilic in nature, making them suitable for synthesis, purification and analysis.

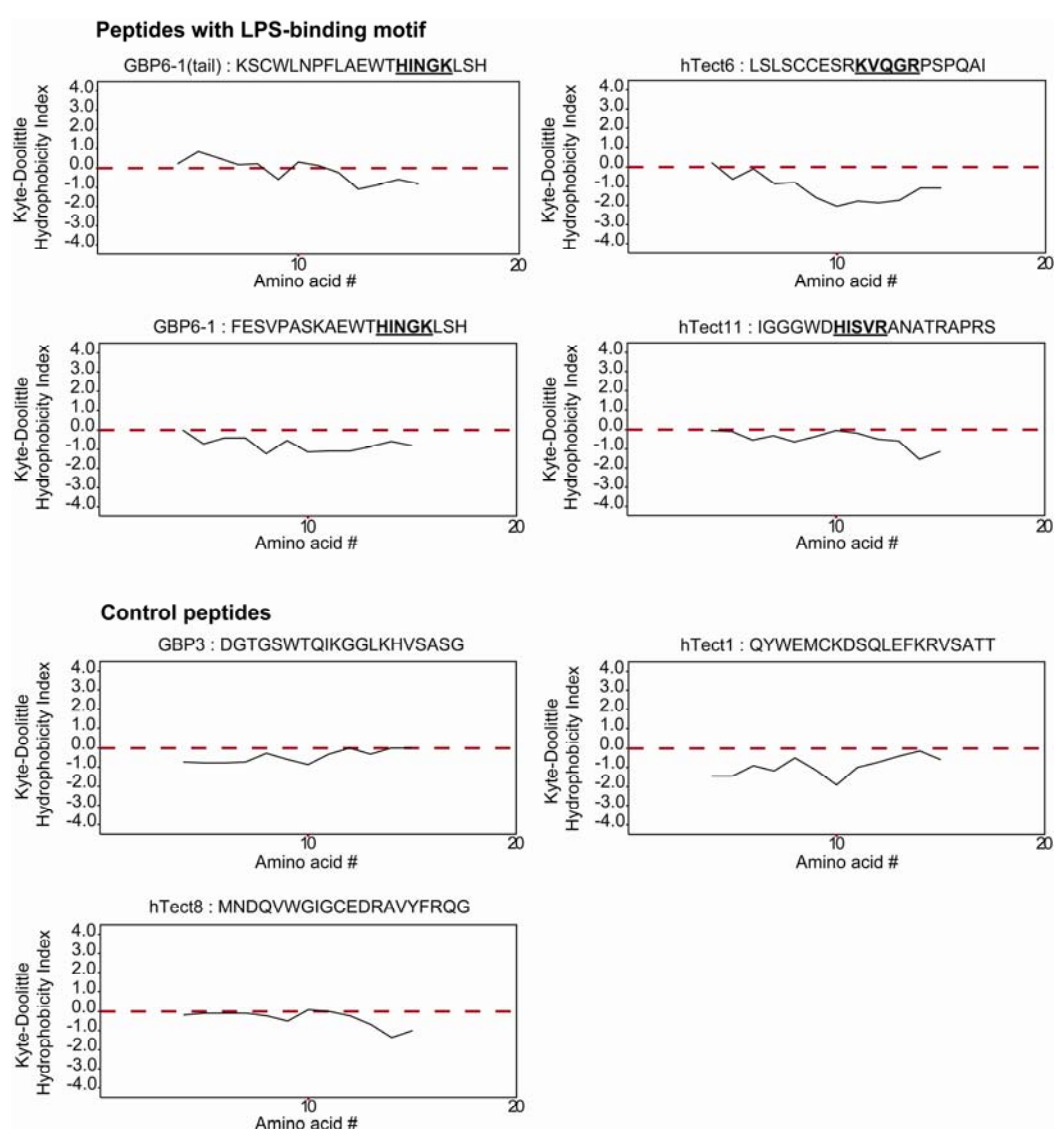


Figure 4.12: Hydrophobicity plots analysis of GBP and hTectonin peptides. The designed Tectonin peptides from GBP and hTectonin together with the respective control peptides show that they are similarly hydrophilic (area below red dashed line) in nature and suitable for synthesis and study. The LPS-binding motifs of *BHPHB* are bold and underlined.

4.2.3.4 hTectonin- and GBP-derived Tectonin peptides bind LPS with high affinity

The designed peptides were then tested for LPS binding via surface plasmon resonance analysis. The hTectonin peptides were flown over LA immobilized on the Biacore HPA chip and showed that indeed they bind to LA at affinities of K_D 10^{-7} to 10^{-8} M, which are similar to the affinities showed by the GBP peptides (Figure 4.13, Table 4.2).

GBP6-1(tail) and GBP6-1 bound LA with similar affinity (Figure 4.12C,D), indicating that the C-terminal 'tail' portion of GBP is not necessary for binding and neither does it adversely affect the LPS-binding motif. Therefore we can ascribe the LPS-binding to purely the Tectonin segments of the protein, consistent with our postulate that the Tectonin domain is important for pathogen interaction.

The hTectonin and the GBP peptides also exhibited similar level of binding affinity to ReLPS and LPS (Figure 4.14, Table 4.2). This corroborates our hypothesis and demonstrates the pathogen binding ability of Tectonin domains and its functional conservation across species, from the horseshoe crab GBP to the human hTectonin. This shows that hTectonin and GBP are potentially homologous proteins based on the shared Tectonin domain motif and they have a shared function in pathogen binding.

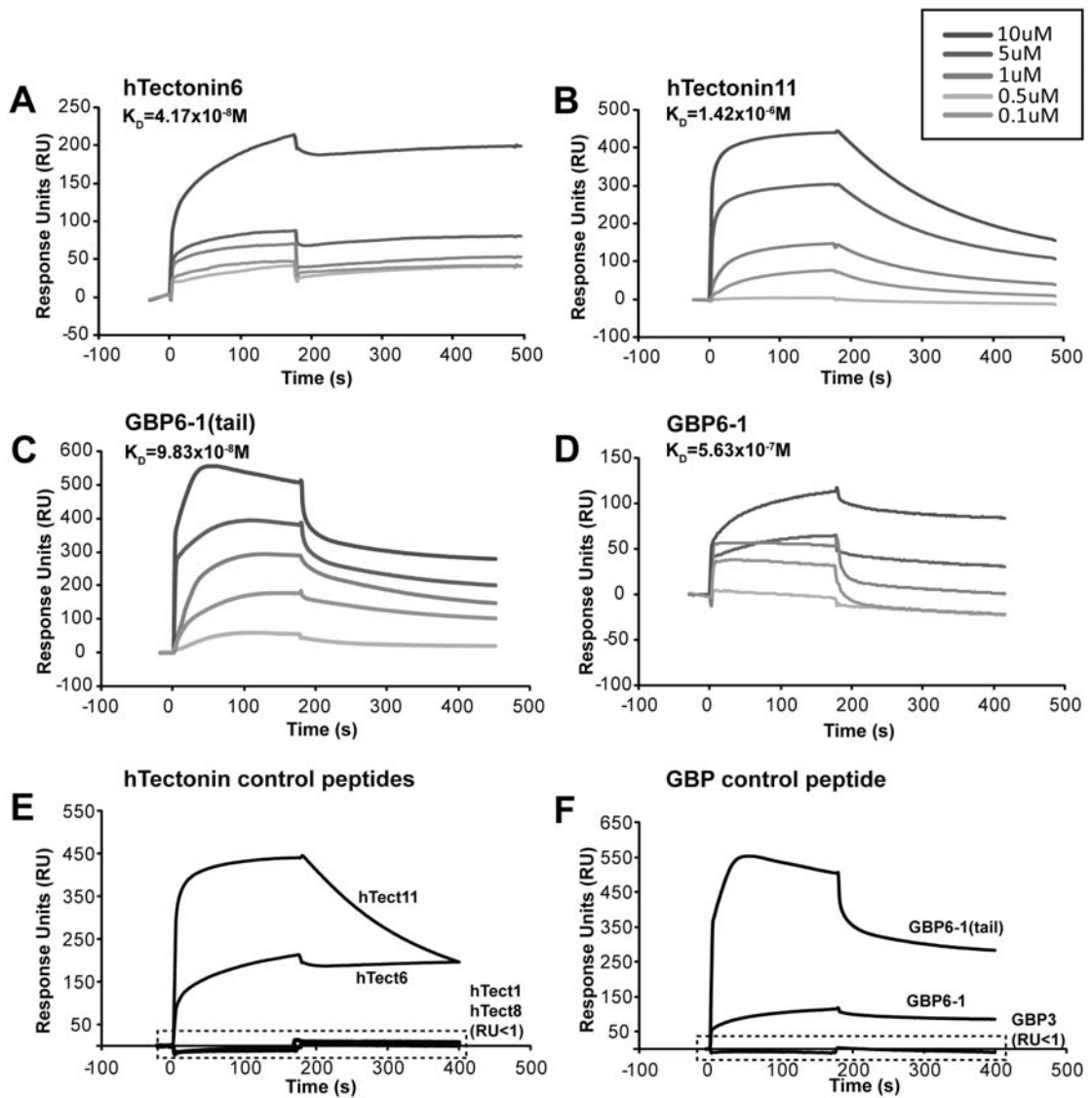


Figure 4.13: Peptides derived from the Tectonin domains of GBP and hTectonin bind LA with high affinity. (A-B) hTectonin peptides, namely hTectonin6 and hTectonin11 containing the predicted LPS-binding motif are able to bind the bacterial structure. **(C-D)** In GBP, the exclusion of an C-terminal ‘tail’ loop which is not part of the beta-propeller Tectonin structure - GBP6-1- gives similar binding affinity with the peptide designed to include this tail region, showing that it does not play an important role in the binding to LA, and that the function comes from within the Tectonin domains. **(E-F)** Several peptides from hTectonin and GBP that did not contain the LPS-binding motif were also derived as controls to measure the specificity of the peptides that were shown to bind to LA. hTect1 and hTectonin8 (derived from Tectonin domains 1 & 8 of hTectonin respectively) and GBP3 (derived from GBP Tectonin domain 3) did not produce significant binding to the LA surface.

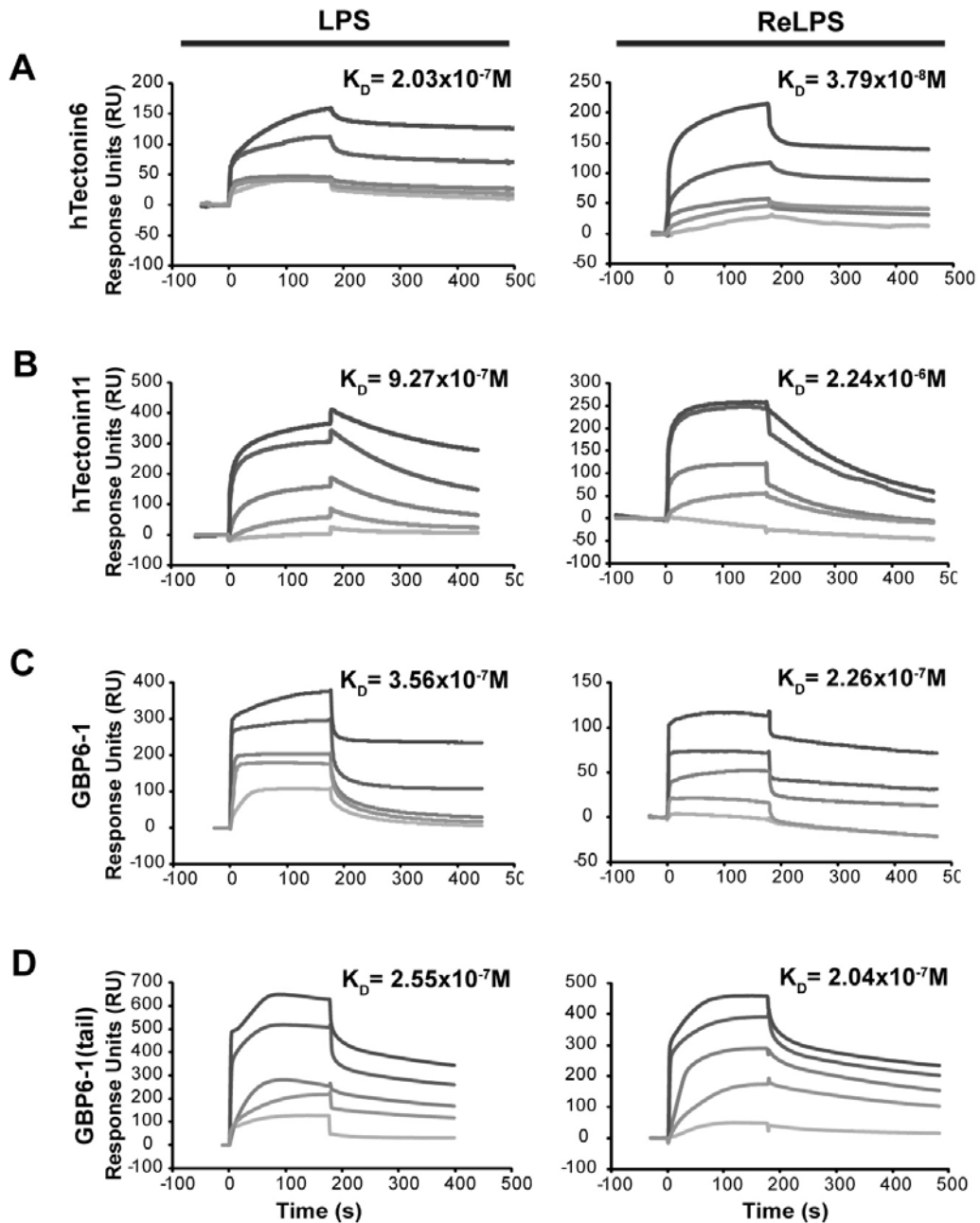


Figure 4.14: Peptides derived from the Tectonin domains of GBP and hTectonin also bind LPS and ReLPS with high affinity. Similar to lipid A, the peptides derived from hTectonin – namely (A) hTectonin6, (B) hTectonin11 – and peptides derived from GBP – (C) GBP6-1 and (D) GBP6-1(tail) also bound tightly to LPS and ReLPS.

Table 4.2: Dissociation constants of Tectonin peptides when bound to LPS, ReLPS, LA

Bacterial ligand	Peptide	Sequence*	K_D (mol⁻¹)
LPS	GBP6-1(tail)	KSCWLN <u>PFLAEWTHING</u> KLSH	2.55x10 ⁻⁷
	GBP6-1	FESVPASKAEW <u>THING</u> KLSH	3.56x10 ⁻⁷
	hTectonin6	LSLSCCESR <u>KVQGR</u> PSPQAI	2.03x10 ⁻⁷
	hTectonin 11	IGGGWD <u>HISVR</u> ANATRAPRS	9.27x10 ⁻⁷
ReLPS	GBP6-1(tail)		2.04x10 ⁻⁷
	GBP6-1		2.26x10 ⁻⁷
	hTectonin 6		3.79x10 ⁻⁸
	hTectonin 11		2.24x10 ⁻⁶
LA	GBP6-1(tail)		9.83x10 ⁻⁸
	GBP6-1		5.63x10 ⁻⁷
	hTectonin 6		4.17x10 ⁻⁸
	hTectonin 11		1.42x10 ⁻⁶

* LPS-binding motifs are underlined

4.3 Summary

In this chapter, we report our discovery of hTectonin, a human protein hitherto classified as being hypothetical, and by structure-function analyses, we inferred its function as an immune-related protein. We show that it retains its function by virtue of possessing Tectonin domains, similar to its invertebrate counterparts. We show that hTectonin contains multiple homologs widespread in the vertebrate kingdom, implying that it is not a one-off protein in the human, but rather an important one conserved throughout many species. We also discovered that the hTectonin gene is expressed in human leukocytes. This is interesting, as a recent addition to the human database of proteins showed another human leukocyte Tectonin protein called the leukolectin. This further indicated that hTectonin may be immune-related. Like its limulus counterpart GBP, which interacts with an important complement initiator CRP, we find that hTectonin also interacts with a cognate complement lectin, Ficolin, which is a homolog of CL-5. This interaction is specific, involving only the Tectonin

domains within the hTectonin protein. We also identified LPS-binding motifs within the hTectonin protein sequence which are also conserved in the thus discovered potential mammalian homologs of hTectonin and are located in highly conserved regions of the sequence, implying that these residues may play a functional role in lipid binding and form part of a highly conserved β -propeller fold. The affinity of these motifs in hTectonin for bacterial LPS and the truncated active forms of the endotoxin molecule (ReLPS and lipid A) was verified experimentally. We propose that hTectonin is a novel human protein that forms a beta-propeller structure which is involved in protein-protein interaction with immune related protein(s), and it simultaneously detects and binds pathogens. Thus, the hTectonin plays a vital role in immune defense, which is conserved over a vast number of organisms.

4.4 Common features of GBP and hTectonin

When the Tectonin proteins were first identified in the slime mold and the horseshoe crabs, they were found to have immune-like potential by being able to bind bacteria. Work in the lab and this thesis has further proven it is indeed an immune protein – by demonstrating that GBP is able to form vast PRR networks, binding bacteria and host proteins simultaneously with high affinity and specificity, and is able to affect endotoxic activity in LPS.

The two PRRs interacting with GBP have homologs in the human system, and thus it was highly possible that there exists a human Tectonin protein. This was indeed the case as we found the hTectonin, which also had homologs in several other mammalian species, proving that it is an important protein in the vertebrates as well.

Further tests revealed similarities between GBP and hTectonin –

- 1) PRR interaction – hTectonin was able to interact with ficolin, which parallels the interaction of GBP with CL5 (the ficolin homolog) in the horseshoe crab. hTectonin also interacted with proteins in the human leukocyte found via yeast 2-hybrid library screening.
- 2) Bacteria binding with high affinity - both GBP and hTectonin displayed the *BHB(P)HB* LPS-binding motif, and via surface plasmon resonance, was able to bind LPS, ReLPS and lipid A with high affinity and specificity.

This shows that hTectonin could likely be the human Tectonin homolog of GBP that was yet to be uncovered until the work in this thesis. Much of the work in the human system has been focused on the adaptive immune system, and here we have shown that innate immunity is also vital and has much potential to be further explored. This thesis has shown that the Tectonin proteins can function as immune proteins and are most likely immune proteins. However, we are just beginning to uncover the potential of the Tectonin proteins. Its conservation from the mold, horseshoe crab right up to the humans shows its vital place in the immune system, and is something that should not be ignored in our studies and interpretation of how our immune system reacts in overcoming pathogen invasion.

CHAPTER 5

CONCLUSION

The innate immune system has been described as the first line of defense of the host in response to pathogen invasion. Also, it is the only protection available to the invertebrates. Although the vertebrates possess the additional protection conferred by the adaptive immune system, innate immunity still proves to be important, especially during the first minutes to hours, and days of an infection.

The innate immune system is endowed with a large repertoire of plasma proteins collectively known as PRRs, and many of them are lectins. However, this particular group of lectin PRRs – the Tectonins - which finds its origin in the slime mold - has hitherto only been studied amongst the invertebrates: the horseshoe crab, the mushroom and the mold.

Having examined and identified that the Tectonin domains are responsible for the recognition and binding of bacterial ligands (LPS), we then sought to find mammalian counterparts of the lectin, particularly in the human. While the Tectonins have been characterized in the invertebrates, only minimal information was available – amino acid sequences that form β -propeller repeats, formation of a β -propeller structure and the ability to bind the bacterial LPS. Until recently, studies in our lab showed that GBP, the horseshoe crab Tectonin lectin from this family interacts with innate immune proteins in the host (Ng et al. 2007). This is the first instance where the Tectonin protein has been shown to be involved in protein-protein interaction (see Figure 1.13), and specifically, interaction amongst PRRs in the immune system (see

Figures 1.14-1.15). GBP was shown to be an infection-induced interaction partner of CRP, a key protein in the complement pathway. However, the mechanism of action of these interactions, and their conditions remained uncharacterized until the work done in this thesis.

In this research, by using both computational and experimental approaches, we identified the regions of GBP, CRP and LPS that mediate their interactions, and showed that infection conditions enhanced GBP:LPS interaction 10-fold and GBP:CRP interaction 1000-fold, which was likely brought on by the depletion of Ca^{2+} . The infection condition was proven irreversible since subsequent manipulations of Ca^{2+} levels were unable to return the binding affinities of these proteins back to basal (pre-infection) level.

The availability of the modeled structure enabled us to visualize experimental data – revealing distinct surfaces of interaction between protein-pathogen and protein-protein. Also, through empirical means such as binding kinetics studies, we found that the conformational changes on GBP upon infection are irreversible, suggesting that the proteins should become processed after fulfilling their duties as pathogen sensors, for example, the complement pathway, where proteins are commonly taken up by macrophages. These findings help to reveal the structural and functional basis of GBP and the Tectonin domain-containing proteins involved in defense against microbial infections.

In search for a human counterpart of the horseshoe crab GBP, the second part of this thesis documents our discovery of a human protein, the hTectonin, previously

classified in the human genome as being hypothetical. By structure-function analyses, we inferred the function of hTectonin as immune-related. We showed that this protein retained its function by virtue of possessing Tectonin domains, similar to its invertebrate counterparts. Sequence matching indicated that the hTectonin is a protein with low homology but phylogenetically related by domain architecture to known proteins with Tectonin domains, functioning as immune proteins. We show through SMART domain comparison that hTectonin contains multiple homologs widespread in the vertebrate kingdom, implying that it is not a one-off protein in the human, but rather an important one conserved throughout many species.

We also discovered that the hTectonin gene is expressed in the human leukocytes, and it interacts with immune-regulatory proteins in the leukocyte. This is interesting, as a recent addition to the human database of proteins showed another human leukocyte Tectonin protein called the leukolectin which further implicates hTectonin to be immune-related. We also identified LPS-binding motifs within the hTectonin protein sequence and validated their capacity for binding bacterial LPS and the truncated active forms of the endotoxin molecule (ReLPS and LA). We propose that hTectonin is a novel human protein that forms a β -propeller structure which is involved in protein-protein interactions with immune related protein(s), and it simultaneously detects and binds pathogens. Thus, the hTectonin plays a vital role in immune defense, which is conserved over a vast number of organisms.

Such a conservation of the Tectonin domains in protein-protein partnerships suggests that hTectonin and GBP share homologies in their secondary structure and function. In evolution, the horseshoe crab and the human are separated by ~500 million years,

yet the remarkable conservation in domain-architecture and potential function suggest a critical role of GBP and the Tectonin-domain-containing proteins in general, in the frontline defense against microbial infection.

CHAPTER 6

FUTURE PERSPECTIVES

The findings in this thesis have opened new directions of further interest. The following questions can be addressed and experiments can be designed for future studies:-

(1) What are the physiological and pathophysiological significance of the Tectonin proteins?

We have observed that the invertebrate Tectonin GBP is able to interact with CRP, a key protein in the complement cascade. Based on results that suggest the eventual turnover of these proteins, how do they mediate downstream action? Is hTectonin involved in immune regulation of bacterial, viral and cancerous non-self proteins? *In vitro* and *in vivo* studies (as hTectonin possesses a homolog in mice) in immune-related cell lines and real-time tracking of the interactions using immunostaining or FRET-analysis can help us visualize the events. Expression studies might help us understand the regulatory effects of the Tectonins upon pathogenic infections.

Our yeast 2-hybrid screening of hTectonin partner proteins also indicated that hTectonin interacts with immune-related proteins. If they are indeed important molecules involved in immune-related activities in the human leukocytes, these would even further prove the importance of hTectonin in the mammalian (in particular, human) immune system.

The SLA-2 and NCF-1 interaction partners are particularly interesting. SLA2 is similar to the Src family of tyrosine kinases. This molecule lacks a catalytic tyrosine kinase domain and is related to a previously identified protein, Src-like adapter protein (SLA), and is therefore designated SLA-2. Jurkat T-cells express SLA-2 protein and overexpression of SLA-2 in these cells negatively regulates T cell receptor signaling by IL-2 (Pandey et al. 2002). Defects in NCF1 have been suggested to lead to the inability of neutrophils and phagocytes to kill microbes, and patients suffer from life-threatening bacterial/fungal infections (Noack et al. 2001). This points to a possible role of NCF-1 in bacteria detection and elimination. Together with these newly elucidated potential interaction partners, and since Tectonin proteins have been shown to be bacteria-binding, it would be interesting to elucidate: (1) The effects of LPS stimulation on hTectonin, as a representative Tectonin protein in the human – by examining protein expression levels in appropriate immune cell lines and (2) The relationship of hTectonin to SLA-2 and NCF-1, in terms of protein expression, location and function of hTectonin in signaling pathways.

(2) Are the Tectonin peptides suitable for development as potential antimicrobial drugs?

The Tectonin peptides derived from both the GBP and hTectonin bind LPS with high affinity – the elucidation of their structure-activity relationships would enable a more detailed analysis of their mechanism of action and guide computational drug design and modifications, leading to its potential use as LPS-binding and LPS-neutralizing drugs. Site-directed mutagenesis studies could also be carried out on the regions where the LPS-binding motif appear on the proteins (and peptides) to determine, without question, that the LPS-binding ability comes from such motifs in Tectonins.

Also, in light of the revelation in Chapter 2 where GBP and the Tectonin-derived peptides were able to increase the endotoxic potential of LPS, these peptides could be used as triggers or enhancers of immune responsive action.

(3) Crystallographic structure analysis of GBP with CRP and LPS

In this thesis, we had attempted to crystallize GBP and CRP individually, to study the structural basis of their immune response action. However, since they work in concert and bind LPS strongly, perhaps co-crystallization of the high affinity complex should be considered as well. The X-ray crystallographic determination of this pathogen recognition complex will also serve to further validate interaction data from HDMS and yeast 2-hybrid results. NMR and molecular dynamics simulations of the interaction, together with isothermal titration calorimetry (ITC) with enthalpy and entropy measurements will help in the characterization and strengthening of the affinity data of the interactions between GBP, CRP and LPS.

CHAPTER 7

REFERENCES

- Aderem, A. and R. J. Ulevitch (2000). "Toll-like receptors in the induction of the innate immune response." Nature **406**(6797): 782-7.
- Anderson, K. V. (2000). "Toll signaling pathways in the innate immune response." Curr Opin Immunol **12**(1): 13-9.
- Andrade, M. A., C. Perez-Iratxeta and C. P. Ponting (2001). "Protein repeats: structures, functions, and evolution." J Struct Biol **134**(2-3): 117-31.
- Ariki, S., K. Koori, T. Osaki, K. Motoyama, K. Inamori and S. Kawabata (2004). "A serine protease zymogen functions as a pattern-recognition receptor for lipopolysaccharides." Proc Natl Acad Sci U S A **101**(4): 953-8.
- Aslam, S. N., M. A. Newman, G. Erbs, K. L. Morrissey, D. Chinchilla, T. Boller, T. T. Jensen, C. De Castro, T. Ierano, A. Molinaro, R. W. Jackson, M. R. Knight and R. M. Cooper (2008). "Bacterial Polysaccharides Suppress Induced Innate Immunity by Calcium Chelation." Curr Biol.
- Aurell, C. A. and A. O. Wistrom (1998). "Critical aggregation concentrations of gram-negative bacterial lipopolysaccharides (LPS)." Biochem Biophys Res Commun **253**(1): 119-23.
- Beisel, H. G., S. Kawabata, S. Iwanaga, R. Huber and W. Bode (1999). "Tachylectin-2: crystal structure of a specific GlcNAc/GalNAc-binding lectin involved in the innate immunity host defense of the Japanese horseshoe crab *Tachypleus tridentatus*." EMBO J **18**(9): 2313-22.
- Berger, D., S. Schleich, M. Seidelmann and H. G. Beger (1991). "Demonstration of an interaction between transferrin and lipopolysaccharide--an in vitro study." Eur Surg Res **23**(5-6): 309-16.
- Beutler, B. (2003). "Not "molecular patterns" but molecules." Immunity **19**(2): 155-6.
- Beutler, B. and A. Cerami (1988). "The history, properties, and biological effects of cachectin." Biochemistry **27**(20): 7575-82.
- Blackwell, J. M., S. Searle, T. Goswami and E. N. Miller (2000). "Understanding the multiple functions of Nramp1." Microbes Infect **2**(3): 317-21.
- Bondar, R. J., J. D. Teller, A. Bowanko and K. M. Kelly (1979). "Properties of *Limulus* amoebocyte lysate and the turbidimetric assay for the quantitative

- determination of gram negative bacterial endotoxin." Prog Clin Biol Res **29**: 435-51.
- Boraston, A. B., D. N. Bolam, H. J. Gilbert and G. J. Davies (2004). "Carbohydrate-binding modules: fine-tuning polysaccharide recognition." Biochem J **382**(Pt 3): 769-81.
- Bork, P. and R. F. Doolittle (1994). "Drosophila kelch motif is derived from a common enzyme fold." J Mol Biol **236**(5): 1277-82.
- Bossart-Whitaker, P., M. Carson, Y. S. Babu, C. D. Smith, W. G. Laver and G. M. Air (1993). "Three-dimensional structure of influenza A N9 neuraminidase and its complex with the inhibitor 2-deoxy 2,3-dehydro-N-acetyl neuraminic acid." J Mol Biol **232**(4): 1069-83.
- Brenowitz, M., C. Bonaventura and J. Bonaventura (1983). "Assembly and calcium-induced cooperativity of Limulus IV hemocyanin: a model system for analysis of structure-function relationships in the absence of subunit heterogeneity." Biochemistry **22**(20): 4707-13.
- Brogden, K. A. (2005). "Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria?" Nat Rev Microbiol **3**(3): 238-50.
- Brudler, R., C. R. Gessner, S. Li, S. Tyndall, E. D. Getzoff and V. L. Woods, Jr. (2006). "PAS domain allostery and light-induced conformational changes in photoactive yellow protein upon I2 intermediate formation, probed with enhanced hydrogen/deuterium exchange mass spectrometry." J Mol Biol **363**(1): 148-60.
- Brummelkamp, T. R., S. M. Nijman, A. M. Dirac and R. Bernards (2003). "Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-kappaB." Nature **424**(6950): 797-801.
- Bull, J. H., G. Ellison, A. Patel, G. Muir, M. Walker, M. Underwood, F. Khan and L. Paskins (2001). "Identification of potential diagnostic markers of prostate cancer and prostatic intraepithelial neoplasia using cDNA microarray." Br J Cancer **84**(11): 1512-9.
- Case, D. A., T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods (2005). "The Amber biomolecular simulation programs." J Comput Chem **26**(16): 1668-88.
- Casimir, C. M., H. N. Bu-Ghanim, A. R. Rodaway, D. L. Bentley, P. Rowe and A. W. Segal (1991). "Autosomal recessive chronic granulomatous disease caused by deletion at a dinucleotide repeat." Proc Natl Acad Sci U S A **88**(7): 2753-7.

- Cerenius, L. and K. Soderhall (2004). "The prophenoloxidase-activating system in invertebrates." Immunol Rev **198**: 116-26.
- Chen, S. C., C. H. Yen, M. S. Yeh, C. J. Huang and T. Y. Liu (2001). "Biochemical properties and cDNA cloning of two new lectins from the plasma of *Tachypleus tridentatus*: *Tachypleus* plasma lectin 1 and 2+." J Biol Chem **276**(13): 9631-9.
- Chiou, S. T., Y. W. Chen, S. C. Chen, C. F. Chao and T. Y. Liu (2000). "Isolation and characterization of proteins that bind to galactose, lipopolysaccharide of *Escherichia coli*, and protein A of *Staphylococcus aureus* from the hemolymph of *Tachypleus tridentatus*." J Biol Chem **275**(3): 1630-4.
- Chu, B. (1991). Laser Light Scattering: Basic Principles and Practice. Boston, Academic Press.
- Cioci, G., E. P. Mitchell, V. Chazalet, H. Debray, S. Oscarson, M. Lahmann, C. Gautier, C. Breton, S. Perez and A. Imberty (2006). "Beta-propeller crystal structure of *Psathyrella velutina* lectin: an integrin-like fungal protein interacting with monosaccharides and calcium." J Mol Biol **357**(5): 1575-91.
- Danner, R. L., R. J. Elin, J. M. Hosseini, R. A. Wesley, J. M. Reilly and J. E. Parillo (1991). "Endotoxemia in human septic shock." Chest **99**(1): 169-75.
- de Vries, S. J., A. D. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar and A. M. Bonvin (2007). "HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets." Proteins **69**(4): 726-33.
- Dhople, V., A. Krukemeyer and A. Ramamoorthy (2006). "The human beta-defensin-3, an antibacterial peptide with multiple biological functions." Biochim Biophys Acta **1758**(9): 1499-512.
- Dinarello, C. A. (1986). "Interleukin-1: amino acid sequences, multiple biological activities and comparison with tumor necrosis factor (cachectin)." Year Immunol **2**: 68-89.
- Ding, J. L. and B. Ho (2001). "A new era in pyrogen testing." Trends Biotechnol **19**(8): 277-81.
- Ding, J. L., M. A. Navas, 3rd and B. Ho (1993). "Two forms of factor C from the amoebocytes of *Carcinoscorpius rotundicauda*: purification and characterisation." Biochim Biophys Acta **1202**(1): 149-56.
- Ding, J. L., K. C. Tan, S. Thangamani, N. Kusuma, W. K. Seow, T. H. Bui, J. Wang and B. Ho (2005). "Spatial and temporal coordination of expression of

- immune response genes during *Pseudomonas* infection of horseshoe crab, *Carcinoscorpius rotundicauda*." Genes Immun **6**(7): 557-74.
- Ding, J. L., L. Wang and B. Ho (2004). "Current Genome-Wide Analysis on Serine Proteases in Innate Immunity." Current Genomics **5**(2): 9.
- Dominguez, C., R. Boelens and A. M. Bonvin (2003). "HADDOCK: a protein-protein docking approach based on biochemical or biophysical information." J Am Chem Soc **125**(7): 1731-7.
- Elgavish, S. and B. Shaanan (1997). "Lectin-carbohydrate interactions: different folds, common recognition principles." Trends Biochem Sci **22**(12): 462-7.
- Freceer, V., B. Ho and J. L. Ding (2000). "Interpretation of biological activity data of bacterial endotoxins by simple molecular models of mechanism of action." Eur J Biochem **267**(3): 837-52.
- Fulop, V. and D. T. Jones (1999). "Beta propellers: structural rigidity and functional diversity." Curr Opin Struct Biol **9**(6): 715-21.
- Galliano, M., L. Minchiotti, M. Campagnoli, A. Sala, L. Visai, A. Amoresano, P. Pucci, A. Casbarra, M. Cauci, M. Perduca and H. L. Monaco (2003). "Structural and biochemical characterization of a new type of lectin isolated from carp eggs." Biochem J **376**(Pt 2): 433-40.
- Gong, K., L. Li, J. F. Wang, F. Cheng, D. Q. Wei and K. C. Chou (2009). "Binding mechanism of H5N1 influenza virus neuraminidase with ligands and its implication for drug design." Med Chem **5**(3): 242-9.
- Gura, T. (2001). "Innate immunity. Ancient system gets new respect." Science **291**(5511): 2068-71.
- Hamuro, Y., S. J. Coales, J. A. Morrow, K. S. Molnar, S. J. Tuske, M. R. Southern and P. R. Griffin (2006). "Hydrogen/deuterium-exchange (H/D-Ex) of PPARgamma LBD in the presence of various modulators." Protein Sci **15**(8): 1883-92.
- Hancock, R. (1999). "Host defence (cationic) peptides: What is their future clinical potential?" Drugs **57**: 469-473.
- Hashimoto, H. (2006). "Recent structural studies of carbohydrate-binding modules." Cell Mol Life Sci **63**(24): 2954-67.
- Hata, K., K. Koseki, K. Yamaguchi, S. Moriya, Y. Suzuki, S. Yingsakmongkon, G. Hirai, M. Sodeoka, M. von Itzstein and T. Miyagi (2008). "Limited inhibitory

effects of oseltamivir and zanamivir on human sialidases." Antimicrob Agents Chemother **52**(10): 3484-91.

Hetenyi, C. and D. van der Spoel (2002). "Efficient docking of peptides to proteins without prior knowledge of the binding site." Protein Sci **11**(7): 1729-37.

Hetenyi, C. and D. van der Spoel (2006). "Blind docking of drug-sized compounds to proteins with up to a thousand residues." FEBS Lett **580**(5): 1447-50.

Hinshaw, L. B. (1984). Handbook of Endotoxin : Pathophysiology of Endotoxin. Amsterdam, Elsevier/North-Holland Biomedical.

Ho, B. (1983). "An improved Limulus gelation assay." Microbios Lett. **24**: 81-84.

Holland, P. C., D. Thompson, S. Hancock and D. Hodge (2002). "Calciphylaxis, proteases, and purpura: an alternative hypothesis for the severe shock, rash, and hypocalcemia associated with meningococcal septicemia." Crit Care Med **30**(12): 2757-61.

Holland, S. J., X. C. Liao, M. K. Mendenhall, X. Zhou, J. Pardo, P. Chu, C. Spencer, A. Fu, N. Sheng, P. Yu, E. Pali, A. Nagin, M. Shen, S. Yu, E. Chan, X. Wu, C. Li, M. Woisetschlager, G. Aversa, F. Kolbinger, M. K. Bennett, S. Molineaux, Y. Luo, D. G. Payan, H. S. Mancebo and J. Wu (2001). "Functional cloning of Src-like adapter protein-2 (SLAP-2), a novel inhibitor of antigen receptor signaling." J Exp Med **194**(9): 1263-76.

Horn, D. L., D. C. Morrison, S. M. Opal, R. Silverstein, K. Visvanathan and J. B. Zabriskie (2000). "What are the microbial components implicated in the pathogenesis of sepsis? Report on a symposium." Clin Infect Dis **31**(4): 851-8.

Horn, J. R., B. Kraybill, E. J. Petro, S. J. Coales, J. A. Morrow, Y. Hamuro and A. A. Kossiakoff (2006). "The role of protein dynamics in increasing binding affinity for an engineered protein-protein interaction established by H/D exchange mass spectrometry." Biochemistry **45**(28): 8488-98.

Howe, J., P. Garidel, M. Roessle, W. Richter, C. Alexander, K. Fournier, J. P. Mach, T. Waelli, R. M. Gorczynski, A. J. Ulmer, U. Zahringer, A. Hartmann, E. T. Rietschel and K. Brandenburg (2008). "Structural investigations into the interaction of hemoglobin and part structures with bacterial endotoxins." Innate Immun **14**(1): 39-49.

Huang, H. W. (2000). "Action of antimicrobial peptides: two-state model." Biochemistry **39**(29): 8347-52.

Hubbard, S. J. and J. M. Thornton (1993). NACCESS.

- Huh, C. G., J. Aldrich, J. Mottahedeh, H. Kwon, C. Johnson and R. Marsh (1998). "Cloning and characterization of Physarum polycephalum tectonins. Homologues of Limulus lectin L-6." J Biol Chem **273**(11): 6565-74.
- Iwaki, D., T. Osaki, Y. Mizunoe, S. N. Wai, S. Iwanaga and S. Kawabata (1999). "Functional and structural diversities of C-reactive proteins present in horseshoe crab hemolymph plasma." Eur J Biochem **264**(2): 314-26.
- Iwanaga, S. (2002). "The molecular basis of innate immunity in the horseshoe crab." Curr Opin Immunol **14**(1): 87-95.
- Iwanaga, S. and B. L. Lee (2005). "Recent advances in the innate immunity of invertebrate animals." J Biochem Mol Biol **38**(2): 128-50.
- Janeway, C. A. (2005). Immunobiology 6th Ed. New York, Garland Sciences.
- Janeway, C. A., Jr. and R. Medzhitov (2002). "Innate immune recognition." Annu Rev Immunol **20**: 197-216.
- Jawad, Z. and M. Paoli (2002). "Novel sequences propel familiar folds." Structure **10**(4): 447-54.
- Jeon, H., W. Meng, J. Takagi, M. J. Eck, T. A. Springer and S. C. Blacklow (2001). "Implications for familial hypercholesterolemia from the structure of the LDL receptor YWTD-EGF domain pair." Nat Struct Biol **8**(6): 499-504.
- Jiang, N., N. S. Tan, B. Ho and J. L. Ding (2007). "Respiratory protein-generated reactive oxygen species as an antimicrobial strategy." Nat Immunol **8**(10): 1114-22.
- Ju, B. G., S. Jeong, E. Bae, S. Hyun, S. B. Carroll, J. Yim and J. Kim (2000). "Fringe forms a complex with Notch." Nature **405**(6783): 191-5.
- Kaca, W. and R. Roth (1995). "Activation of complement by human hemoglobin and by mixtures of hemoglobin and bacterial endotoxin." Biochim Biophys Acta **1245**(1): 49-56.
- Kato, N., M. Ohta, N. Kido, H. Ito, S. Naito, T. Hasegawa, T. Watabe and K. Sasaki (1990). "Crystallization of R-form lipopolysaccharides from Salmonella minnesota and Escherichia coli." J Bacteriol **172**(3): 1516-28.
- Katsikis, P. D., S. P. Schoenberger and B. Pulendran (2007). "Probing the 'labyrinth' linking the innate and adaptive immune systems." Nat Immunol **8**(9): 899-901.
- Kawabata, S. and S. Iwanaga (1999). "Role of lectins in the innate immunity of horseshoe crab." Dev Comp Immunol **23**(4-5): 391-400.

- Kawabata, S. and R. Tsuda (2002). "Molecular basis of non-self recognition by the horseshoe crab tachylectins." Biochim Biophys Acta **1572**(2-3): 414-21.
- Kawasaki, H., T. Nose, T. Muta, S. Iwanaga, Y. Shimohigashi and S. Kawabata (2000). "Head-to-tail polymerization of coagulin, a clottable protein of the horseshoe crab." J Biol Chem **275**(45): 35297-301.
- Kiener, P. A., F. Marek, G. Rodgers, P. F. Lin, G. Warr and J. Desiderio (1988). "Induction of tumor necrosis factor, IFN-gamma, and acute lethality in mice by toxic and non-toxic forms of lipid A." J Immunol **141**(3): 870-4.
- Kimball, J. (1994). *Kimball's Biology*.
- Kovalenko, A., C. Chable-Bessia, G. Cantarella, A. Israel, D. Wallach and G. Courtois (2003). "The tumour suppressor CYLD negatively regulates NF-kappaB signalling by deubiquitination." Nature **424**(6950): 801-5.
- Kumar, A., M. T. Eby, S. Sinha, A. Jasmin and P. M. Chaudhary (2001). "The ectodermal dysplasia receptor activates the nuclear factor-kappaB, JNK, and cell death pathways and binds to ectodysplasin A." J Biol Chem **276**(4): 2668-77.
- Le Saux, A., P. M. Ng, J. J. Koh, D. H. Low, G. E. Leong, B. Ho and J. L. Ding (2008). "The macromolecular assembly of pathogen-recognition receptors is impelled by serine proteases, via their complement control protein modules." J Mol Biol **377**(3): 902-13.
- Lemaitre, B., J. M. Reichhart and J. A. Hoffmann (1997). "Drosophila host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms." Proc Natl Acad Sci U S A **94**(26): 14614-9.
- Letunic, I., T. Doerks and P. Bork (2009). "SMART 6: recent updates and new developments." Nucleic Acids Res **37**(Database issue): D229-32.
- Li, P., M. Sun, T. Wohland, D. Yang, B. Ho and J. L. Ding (2006). "Molecular mechanisms that govern the specificity of Sushi peptides for Gram-negative bacterial membrane lipids." Biochemistry **45**(35): 10554-62.
- Li, P., T. Wohland, B. Ho and J. L. Ding (2004). "Perturbation of Lipopolysaccharide (LPS) Micelles by Sushi 3 (S3) antimicrobial peptide. The importance of an intermolecular disulfide bond in S3 dimer for binding, disruption, and neutralization of LPS." J Biol Chem **279**(48): 50150-6.
- Li, Y., P. M. Ng, B. Ho and J. L. Ding (2007). "A female-specific pentraxin, CrOctin, bridges pattern recognition receptors to bacterial phosphoethanolamine." Eur J Immunol **37**(12): 3477-88.

- Litman, G. W., J. P. Cannon and L. J. Dishaw (2005). "Reconstructing immune phylogeny: new perspectives." Nat Rev Immunol **5**(11): 866-79.
- Lodish, H., A. Berk, L. Zipursky, P. Matsudaira, D. Baltimore and J. Darnell (1999). Molecular Cell Biology. New York, Scientific American Books.
- Loppnow, H., H. Brade, I. Durrbaum, C. A. Dinarello, S. Kusumoto, E. T. Rietschel and H. D. Flad (1989). "IL-1 induction-capacity of defined lipopolysaccharide partial structures." J Immunol **142**(9): 3229-38.
- Lovell, S. C., I. W. Davis, W. B. Arendall, 3rd, P. I. de Bakker, J. M. Word, M. G. Prisant, J. S. Richardson and D. C. Richardson (2003). "Structure validation by Calpha geometry: phi,psi and Cbeta deviation." Proteins **50**(3): 437-50.
- Maguire, M. E. (2006). "Magnesium transporters: properties, regulation and structure." Front Biosci **11**: 3149-63.
- Mandell, J. G., A. M. Falick and E. A. Komives (1998). "Measurement of amide hydrogen exchange by MALDI-TOF mass spectrometry." Anal Chem **70**(19): 3987-95.
- Matzinger, P. (2002). "The danger model: a renewed sense of self." Science **296**(5566): 301-5.
- Medzhitov, R. (2007). "Recognition of microorganisms and activation of the immune response." Nature **449**(7164): 819-26.
- Medzhitov, R. and C. Janeway, Jr. (2000). "Innate immunity." N Engl J Med **343**(5): 338-44.
- Medzhitov, R. and C. A. Janeway, Jr. (1997). "Innate immunity: impact on the adaptive immune response." Curr Opin Immunol **9**(1): 4-9.
- Miftari, M. H. and B. T. Walther (2009). Molecular cloning and characterization of the leukolectin gene isolated from the human leukocytes.
- Miyata, T., F. Tokunaga, T. Yoneya, K. Yoshikawa, S. Iwanaga, M. Niwa, T. Takao and Y. Shimonishi (1989). "Antimicrobial peptides, isolated from horseshoe crab hemocytes, tachyplesin II, and polyphemusins I and II: chemical structures and biological activity." J Biochem **106**(4): 663-8.
- Morrison, D. C. and J. L. Ryan (1987). "Endotoxins and disease mechanisms." Annu Rev Med **38**: 417-32.
- Moscona, A. (2005). "Oseltamivir resistance--disabling our influenza defenses." N Engl J Med **353**(25): 2633-6.

- Moscona, A. (2009). "Global transmission of oseltamivir-resistant influenza." N Engl J Med **360**(10): 953-6.
- Murakami, T., M. Niwa, F. Tokunaga, T. Miyata and S. Iwanaga (1991). "Direct virus inactivation of tachyplesin I and its isopeptides from horseshoe crab hemocytes." Chemotherapy **37**(5): 327-34.
- Natanson, C., P. W. Eichenholz, R. L. Danner, P. Q. Eichacker, W. D. Hoffman, G. C. Kuo, S. M. Banks, T. J. MacVittie and J. E. Parrillo (1989). "Endotoxin and tumor necrosis factor challenges in dogs simulate the cardiovascular profile of human septic shock." J Exp Med **169**(3): 823-32.
- Neer, E. J., C. J. Schmidt, R. Nambudripad and T. F. Smith (1994). "The ancient regulatory-protein family of WD-repeat proteins." Nature **371**(6495): 297-300.
- Ng, P. M., Z. Jin, S. S. Tan, B. Ho and J. L. Ding (2004). "C-reactive protein: a predominant LPS-binding acute phase protein responsive to *Pseudomonas* infection." J Endotoxin Res **10**(3): 163-74.
- Ng, P. M., A. Le Saux, C. M. Lee, N. S. Tan, J. Lu, S. Thiel, B. Ho and J. L. Ding (2007). "C-reactive protein collaborates with plasma lectins to boost immune response against bacteria." EMBO J **26**(14): 3431-40.
- Nickols, N. G. and P. B. Dervan (2007). "Suppression of androgen receptor-mediated gene expression by a sequence-specific DNA-binding polyamide." Proc Natl Acad Sci U S A **104**(25): 10418-23.
- Nikaido, H. and M. Vaara (1987). In *Escherichia coli* and *Salmonella typhimurium*. Washington, DC, ASM Publications.
- Nilges, M. (1995). "Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities." J Mol Biol **245**(5): 645-60.
- Nilges, M. (1996). "Structure calculation from NMR data." Curr Opin Struct Biol **6**(5): 617-23.
- Noack, D., J. Rae, A. R. Cross, B. A. Ellis, P. E. Newburger, J. T. Curnutte and P. G. Heyworth (2001). "Autosomal recessive chronic granulomatous disease caused by defects in NCF-1, the gene encoding the phagocyte p47-phox: mutations not arising in the NCF-1 pseudogenes." Blood **97**(1): 305-11.
- O'Croinin, D. F., A. D. Nichol, N. Hopkins, J. Boylan, S. O'Brien, C. O'Connor, J. G. Laffey and P. McLoughlin (2008). "Sustained hypercapnic acidosis during pulmonary infection increases bacterial load and worsens lung injury." Crit Care Med **36**(7): 2128-35.

- Old, L. J. (1988). "Tumor necrosis factor." Sci Am **258**(5): 59-60, 69-75.
- Ong, S. T., J. Z. Ho, B. Ho and J. L. Ding (2006). "Iron-withholding strategy in innate immunity." Immunobiology **211**(4): 295-314.
- Onishi, H. R., B. A. Pelak, L. S. Gerckens, L. L. Silver, F. M. Kahan, M. H. Chen, A. A. Patchett, S. M. Galloway, S. A. Hyland, M. S. Anderson and C. R. Raetz (1996). "Antibacterial agents that inhibit lipid A biosynthesis." Science **274**(5289): 980-2.
- Opal, S. M. (1995). "Lessons learned in recent clinical trials for sepsis." J Endotoxin Res **2**: 221-226.
- Oren, Z. and Y. Shai (1998). "Mode of action of linear amphipathic alpha-helical antimicrobial peptides." Biopolymers **47**(6): 451-63.
- Oubrie, A., H. J. Rozeboom, K. H. Kalk, A. J. Olsthoorn, J. A. Duine and B. W. Dijkstra (1999). "Structure and mechanism of soluble quinoprotein glucose dehydrogenase." EMBO J **18**(19): 5187-94.
- Pandey, A., N. Ibarrola, I. Kratchmarova, M. M. Fernandez, S. N. Constantinescu, O. Ohara, S. Sawasdikosol, H. F. Lodish and M. Mann (2002). "A novel Src homology 2 domain-containing molecule, Src-like adapter protein-2 (SLAP-2), which negatively regulates T cell receptor signaling." J Biol Chem **277**(21): 19131-8.
- Paoli, M. (2001). "Protein folds propelled by diversity." Prog Biophys Mol Biol **76**(1-2): 103-30.
- Papp-Wallace, K. M. and M. E. Maguire (2006). "Manganese transport and the role of manganese in virulence." Annu Rev Microbiol **60**: 187-209.
- Patil, P. B. and R. V. Sonti (2004). "Variation suggestive of horizontal gene transfer at a lipopolysaccharide (lps) biosynthetic locus in *Xanthomonas oryzae* pv. *oryzae*, the bacterial leaf blight pathogen of rice." BMC Microbiol **4**: 40.
- Pebay-Peyroula, E., G. Rummel, J. P. Rosenbusch and E. M. Landau (1997). "X-ray structure of bacteriorhodopsin at 2.5 angstroms from microcrystals grown in lipidic cubic phases." Science **277**(5332): 1676-81.
- Petsko, G. A. and R. Dagmar (2004). Protein structure and function, New Science Press.
- Pons, T., R. Gomez, G. Chinaea and A. Valencia (2003). "Beta-propellers: associated functions and their role in human diseases." Curr Med Chem **10**(6): 505-24.

- Ponting, C. P., R. Mott, P. Bork and R. R. Copley (2001). "Novel protein domains and repeats in *Drosophila melanogaster*: insights into structure, function, and evolution." Genome Res **11**(12): 1996-2008.
- Pristovsek, P. and J. Kidric (2004). "The search for molecular determinants of LPS inhibition by proteins and peptides." Curr Top Med Chem **4**(11): 1185-201.
- Raetz, C. R. (1986). "Molecular genetics of membrane phospholipid synthesis." Annu Rev Genet **20**: 253-95.
- Raetz, C. R. (1990). "Biochemistry of endotoxins." Annu Rev Biochem **59**: 129-70.
- Reeves, P. P. and L. Wang (2002). "Genomic organization of LPS-specific loci." Curr Top Microbiol Immunol **264**(1): 109-35.
- Regamey, A., D. Hohl, J. W. Liu, T. Roger, P. Kogerman, R. Toftgard and M. Huber (2003). "The tumor suppressor CYLD interacts with TRIP and regulates negatively nuclear factor kappaB activation by tumor necrosis factor." J Exp Med **198**(12): 1959-64.
- Retief, J. D. (2000). "Phylogenetic analysis using PHYLIP." Methods Mol Biol **132**: 243-58.
- Rietschel, E. T. (1984). Handbook of Endotoxin : Chemistry of Endotoxin. Amsterdam, Elsevier/North-Holland Biomedical.
- Rini, J. M. (1995). "Lectin structure." Annu Rev Biophys Biomol Struct **24**: 551-77.
- Roeder, A., C. J. Kirschning, R. A. Rupec, M. Schaller, G. Weindl and H. C. Korting (2004). "Toll-like receptors as key mediators in innate antifungal immunity." Med Mycol **42**(6): 485-98.
- Rungrotmongkol, T., V. Frecer, W. De-Eknamkul, S. Hannongbua and S. Miertus (2009). "Design of oseltamivir analogs inhibiting neuraminidase of avian influenza virus H5N1." Antiviral Res **82**(1): 51-8.
- Saito, T., S. Kawabata, M. Hirata and S. Iwanaga (1995). "A novel type of limulus lectin-L6. Purification, primary structure, and antibacterial activity." J Biol Chem **270**(24): 14493-9.
- Schlessinger, A. and B. Rost (2005). "Protein flexibility and rigidity predicted from sequence." Proteins **61**(1): 115-26.
- Schlessinger, A., G. Yachdav and B. Rost (2006). "PROFbval: predict flexible and rigid residues in proteins." Bioinformatics **22**(7): 891-3.

- Schroder, H. C., H. Ushijima, A. Krasko, V. Gamulin, N. L. Thakur, B. Diehl-Seifert, I. M. Muller and W. E. Muller (2003). "Emergence and disappearance of an immune molecule, an antimicrobial lectin, in basal metazoa. A tachylectin-related protein in the sponge *Suberites domuncula*." J Biol Chem **278**(35): 32810-7.
- Schultz, J., F. Milpetz, P. Bork and C. P. Ponting (1998). "SMART, a simple modular architecture research tool: identification of signaling domains." Proc Natl Acad Sci U S A **95**(11): 5857-64.
- Sharon, N. (1993). "Lectin-carbohydrate complexes of plants and animals: an atomic view." Trends Biochem Sci **18**(6): 221-6.
- Shrive, A. K., A. M. Metcalfe, J. R. Cartwright and T. J. Greenhough (1999). "C-reactive protein and SAP-like pentraxin are both present in *Limulus polyphemus* haemolymph: crystal structure of *Limulus* SAP." J Mol Biol **290**(5): 997-1008.
- Sinz, A. (2003). "Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes." J Mass Spectrom **38**(12): 1225-37.
- Springer, T. A. (1998). "An extracellular beta-propeller module predicted in lipoprotein and scavenger receptors, tyrosine kinases, epidermal growth factor precursor, and extracellular matrix components." J Mol Biol **283**(4): 837-62.
- Srinivasan, N., H. E. White, J. Emsley, S. P. Wood, M. B. Pepys and T. L. Blundell (1994). "Comparative analyses of pentraxins: implications for protomer assembly and ligand binding." Structure **2**(11): 1017-27.
- Sturtinova, e. a. (1995). Inflammation & Fever from Pathophysiology : Principles of Disease, Academic Electronic Press.
- Takeda, K. and S. Akira (2005). "Toll-like receptors in innate immunity." Int Immunol **17**(1): 1-14.
- Tan, N. S., B. Ho and J. L. Ding (2000). "High-affinity LPS binding domain(s) in recombinant factor C of a horseshoe crab neutralizes LPS-induced lethality." FASEB J **14**(7): 859-70.
- Tan, N. S., M. L. Ng, Y. H. Yau, P. K. Chong, B. Ho and J. L. Ding (2000). "Definition of endotoxin binding sites in horseshoe crab factor C recombinant sushi proteins and neutralization of endotoxin by sushi peptides." FASEB J **14**(12): 1801-13.

- Thompson, D., M. B. Pepys and S. P. Wood (1999). "The physiological structure of human C-reactive protein and its complex with phosphocholine." Structure **7**(2): 169-77.
- Thompson, J. D., T. J. Gibson and D. G. Higgins (2002). "Multiple sequence alignment using ClustalW and ClustalX." Curr Protoc Bioinformatics **Chapter 2: Unit 23**.
- Tracey, K. J., B. Beutler, S. F. Lowry, J. Merryweather, S. Wolpe, I. W. Milsark, R. J. Hariri, T. J. Fahey, 3rd, A. Zentella, J. D. Albert and et al. (1986). "Shock and tissue injury induced by recombinant human cachectin." Science **234**(4775): 470-4.
- Trompouki, E., E. Hatzivassiliou, T. Tschritzis, H. Farmer, A. Ashworth and G. Mosialos (2003). "CYLD is a deubiquitinating enzyme that negatively regulates NF-kappaB activation by TNFR family members." Nature **424**(6950): 793-6.
- Tsutsui, Y. and P. L. Wintrode (2007). "Hydrogen/deuterium exchange-mass spectrometry: a powerful tool for probing protein structure, dynamics and interactions." Curr Med Chem **14**(22): 2344-58.
- Tzeng, Y. L., A. Datta, V. K. Kolli, R. W. Carlson and D. S. Stephens (2002). "Endotoxin of *Neisseria meningitidis* composed only of intact lipid A: inactivation of the meningococcal 3-deoxy-D-manno-octulosonic acid transferase." J Bacteriol **184**(9): 2379-88.
- van der Ley, P. and L. Steeghs (2003). "Lessons from an LPS-deficient *Neisseria meningitidis* mutant." J Endotoxin Res **9**(2): 124-8.
- van Dijk, A. D., R. Boelens and A. M. Bonvin (2005). "Data-driven docking for the study of biomolecular complexes." FEBS J **272**(2): 293-312.
- Varghese, J. N., W. G. Laver and P. M. Colman (1983). "Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 Å resolution." Nature **303**(5912): 35-40.
- Volpp, B. D., W. M. Nauseef, J. E. Donelson, D. R. Moser and R. A. Clark (1989). "Cloning of the cDNA and functional expression of the 47-kilodalton cytosolic component of human neutrophil respiratory burst oxidase." Proc Natl Acad Sci U S A **86**(18): 7195-9.
- Wang, J., N. S. Tan, B. Ho and J. L. Ding (2002). "Modular arrangement and secretion of a multidomain serine protease. Evidence for involvement of proline-rich region and N-glycans in the secretion pathway." J Biol Chem **277**(39): 36363-72.

- Wang, J. M., P. Cieplak and P. A. Kollman (2000). "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?" Journal of Computational Chemistry **21**(12): 25.
- Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton (2009). "Jalview Version 2--a multiple sequence alignment editor and analysis workbench." Bioinformatics **25**(9): 1189-91.
- Weis, W. I. and K. Drickamer (1996). "Structural basis of lectin-carbohydrate recognition." Annu Rev Biochem **65**: 441-73.
- Wileman, T. E., M. R. Lennartz and P. D. Stahl (1986). "Identification of the macrophage mannose receptor as a 175-kDa membrane protein." Proc Natl Acad Sci U S A **83**(8): 2501-5.
- Wimmerova, M., E. Mitchell, J. F. Sanchez, C. Gautier and A. Imberty (2003). "Crystal structure of fungal lectin: six-bladed beta-propeller fold and novel fucose recognition mode for *Aleuria aurantia* lectin." J Biol Chem **278**(29): 27059-67.
- Yadid, I. and D. S. Tawfik (2007). "Reconstruction of functional beta-propeller lectins via homo-oligomeric assembly of shorter fragments." J Mol Biol **365**(1): 10-7.
- Yeaman, M. R. and N. Y. Yount (2003). "Mechanisms of antimicrobial peptide action and resistance." Pharmacol Rev **55**(1): 27-55.
- Yu, L., M. Tan, B. Ho, J. L. Ding and T. Wohland (2006). "Determination of critical micelle concentrations and aggregation numbers by fluorescence correlation spectroscopy: aggregation of a lipopolysaccharide." Anal Chim Acta **556**(1): 216-25.
- Zhu, Y., P. M. Ng, L. Wang, B. Ho and J. L. Ding (2006). "Diversity in lectins enables immune recognition and differentiation of wide spectrum of pathogens." Int Immunol **18**(12): 1671-80.
- Zhu, Y., S. Thangamani, B. Ho and J. L. Ding (2005). "The ancient origin of the complement system." EMBO J **24**(2): 382-94.

PUBLICATIONS

A Novel Human Tectonin Protein with Multivalent β -Propeller Folds Interacts with Ficolin and Binds Bacterial LPS

Diana Hooi Ping Low^{1,2}, Zhiwei Ang², Quan Yuan², Vladimir Frecer^{3,4}, Bow Ho⁵, Jianzhu Chen^{1,6}, Jeak Ling Ding^{1,2*}

1 Computational and Systems Biology, Singapore-MIT Alliance, Singapore, Singapore, **2** Department of Biological Sciences, National University of Singapore, Singapore, Singapore, **3** Laboratory of Molecular Biostructural and Nanomaterial Modeling, AREA Science Park, Trieste, Italy, **4** Cancer Research Institute, Slovak Academy of Sciences, Bratislava, Slovakia, **5** Department of Microbiology, National University of Singapore, Singapore, Singapore, **6** Koch Institute for Integrative Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

Abstract

Background: Although the human genome database has been completed a decade ago, ~50% of the proteome remains hypothetical as their functions are unknown. The elucidation of the functions of these hypothetical proteins can lead to additional protein pathways and revelation of new cascades. However, many of these inferences are limited to proteins with substantial sequence similarity. Of particular interest here is the Tectonin domain-containing family of proteins.

Methodology/Principal Findings: We have identified hTectonin, a hypothetical protein in the human genome database, as a distant ortholog of the limulus galactose binding protein (GBP). Phylogenetic analysis revealed strong evolutionary conservation of hTectonin homologues from parasite to human. By computational analysis, we showed that both the hTectonin and GBP form β -propeller structures with multiple Tectonin domains, each containing β -sheets of 4 strands per β -sheet. hTectonin is present in the human leukocyte cDNA library and immune-related cell lines. It interacts with M-ficolin, a known human complement protein whose ancient homolog, carcinolectin (CL5), is the functional protein partner of GBP during infection. Yeast 2-hybrid assay showed that only the Tectonin domains of hTectonin recognize the fibrinogen-like domain of the M-ficolin. Surface plasmon resonance analysis showed real-time interaction between the Tectonin domains 6 & 11 and bacterial LPS, indicating that despite forming 2 β -propellers with its different Tectonin domains, the hTectonin molecule could precisely employ domains 6 & 11 to recognise bacteria.

Conclusions/Significance: By virtue of a recent finding of another Tectonin protein, leukolectin, in the human leukocyte, and our structure-function analysis of the hypothetical hTectonin, we propose that Tectonin domains of proteins could play a vital role in innate immune defense, and that this function has been conserved over several hundred million years, from invertebrates to vertebrates. Furthermore, the approach we have used could be employed in unraveling the characteristics and functions of other hypothetical proteins in the human proteome.

Citation: Low DHP, Ang Z, Yuan Q, Frecer V, Ho B, et al. (2009) A Novel Human Tectonin Protein with Multivalent β -Propeller Folds Interacts with Ficolin and Binds Bacterial LPS. PLoS ONE 4(7): e6260. doi:10.1371/journal.pone.0006260

Editor: Stefan Bereswill, Charité-Universitätsmedizin Berlin, Germany

Received: June 12, 2009; **Accepted:** June 19, 2009; **Published:** July 16, 2009

Copyright: © 2009 Low et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a grant from the Ministry of Education (AcRF Tier 2, awarded to JL Ding and B Ho). We also thank the Singapore-MIT Alliance for research scholarship and funding support to DHP Low. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dbsdjl@nus.edu.sg

Introduction

Advances in sequence genomics have resulted in an accumulation of a large number of protein sequences derived from genome sequences. Although the human genome database has been completed a decade ago, about 50% of the human proteome still remains hypothetical as their functions are unknown [1]. The elucidation of the functions of these hypothetical proteins can lead to additional protein pathways and revelation of new cascades, thus completing our fragmentary knowledge on the proteome complex. Furthermore, information on the network of protein-protein interactions will increase logarithmically. New hypothetical proteins may serve as disease markers and pharmacological targets.

The prime targets for the discovery of functional proteins are those which show homology to counterparts in lower species by way of sequence similarities and domain conservation. An alternate approach is to examine the proteins of invertebrates that do not have homologs in the vertebrate system. One example of such a group of proteins is the Tectonin domain-containing proteins in humans. Tectonin domain containing proteins, which belong to a subclass of proteins of the larger β -propeller family, have thus far only been studied in the fish, horseshoe crab, slime mold and sponge [2–5]. Tectonin domains were first reported in the Tectonins I and II proteins of the slime mold, *Physarum polycephalum*. The Tectonins I and II were characterized to have repeats of Tectonin domains [2]. Because the proteins are located

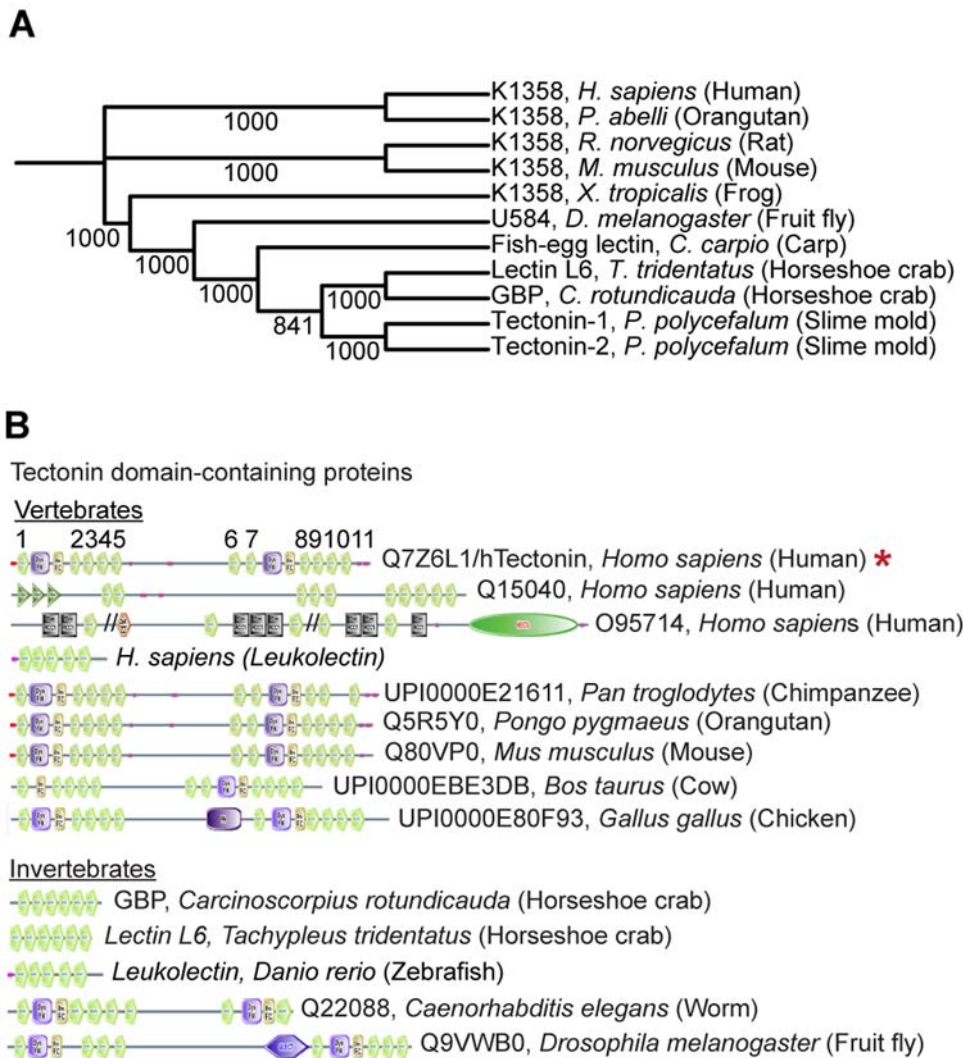


Figure 1. hTectonin is distantly related to the invertebrate Tectonins. (A) The phylogenetic tree constructed after a PSI-Search query using the invertebrate Tectonins revealed K1358 family of proteins as closely related Tectonin domain containing proteins in the mammals and also in lower species like the frog. The numbers at the nodes are an indication of the level of confidence for the branches as determined by bootstrap analysis (1000 bootstrap replicates). (B) Bioinformatics domain analysis utilizing SMART [22,23] shows existence of Tectonin domain-containing proteins both in invertebrates and vertebrates from the horseshoe crab lectins, worm, up to humans. Of interest in this study is the protein hTectonin (red asterisk) which appear to have homologues in other species as well, for example in *P. troglodytes* (chimpanzee), *P. pygmaeus* (orangutan), *M. musculus* (mouse), *G. gallus* (chicken), *C. elegans* (worm) and *D. melanogaster* (fruitfly). doi:10.1371/journal.pone.0006260.g001

at the surface of this organism, where they were postulated to scavenge food including bacteria, the Tectonin proteins have been speculated to function as bacterial sensors. Tachylectin-1 in the horseshoe crab, *Tachypleus tridentatus*, also has Tectonin domain classification [4,6], and was shown to be able to bind bacterial lipopolysaccharide (LPS). Study on the Tectonin protein, LEC_SUBDO, of the sponge, *Suberites domuncula*, also revealed a possible LPS-binding function [3]. A recent investigation on the galactose-binding protein (GBP) in the horseshoe crab (*Carcinoscorpius rotundicauda*), a protein consisting of only 6 Tectonin domains revealed that the Tectonin domains function to differentiate host from pathogen and simultaneously bridge a host-pathogen interactome (Low et al., unpublished).

An exhaustive search in the databases for vertebrate proteins failed to reveal any potential homologs with significant sequence similarity, indicating that perhaps these Tectonin domain-

containing proteins (henceforth referred to as Tectonin proteins) have evolved through the species, although more recently, other proteins with Tectonin domains are being uncovered, for example, the human leukolectin (GenBank Accession No. ACM77812). There are many examples of other families of meiosis-related proteins, kinetochores, cell gap contacts and nuclear pore complexes which show no homology at the primary amino-acid sequence level. However, they hint at the conservation of their domain architecture organization. Furthermore, the three-dimensional structure of functionally important domains in proteins in the budding yeast, nematode, *Drosophila*, *Arabidopsis*, and human have been conserved [7–11]. Here, we have used several databases like SCOP, CATH, SMART, which also employ domain and secondary structure classification for structure sorting and function prediction, to search for β-propeller structures and possibly distance relationships by domain conservation. This is

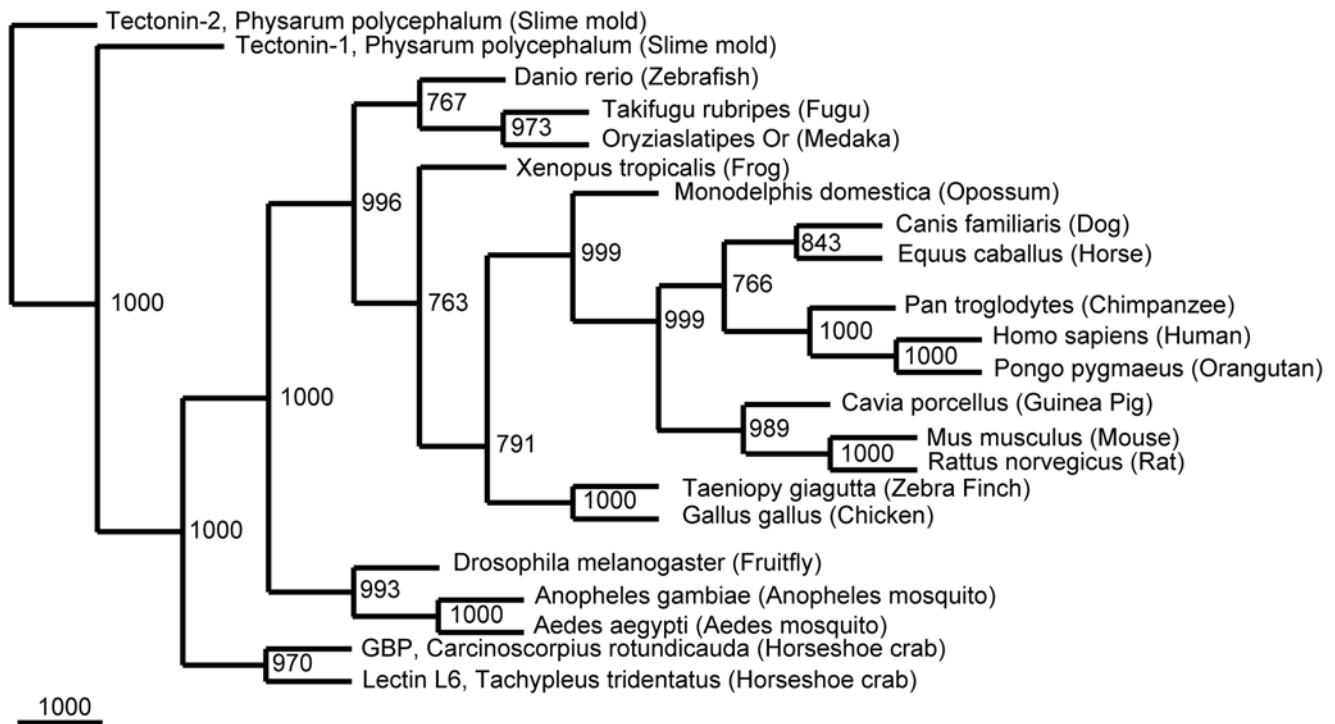


Figure 2. hTectonin gene is widespread across many species. The phylogenetic tree of hTectonin homologues constructed by primary sequence similarity shows its prevalence and conservation among a vast number of different species, right down to the worm, *C. elegans*. The human hTectonin protein was used as a query sequence in BLAST. Top hits were then compiled and multiple sequence alignment based on a guide tree was done using CLUSTALW [32] and the alignment was edited with Jalview [33]. The tree was constructed using the neighbour joining algorithm of the PHYLIP package. The numbers at the nodes are an indication of the level of confidence for the branches as determined by bootstrap analysis (1000 bootstrap replicates).

doi:10.1371/journal.pone.0006260.g002

especially useful when searching for related proteins with low sequence homology or when sequences have diversified through evolution from the invertebrates to the mammals. We thus seek to identify Tectonins in the vertebrates, and compare their domain architecture and function with ancient homologs from the invertebrates in order to gain insights into their functional conservation in the vertebrates, particularly in view of host-pathogen interactions.

By using known invertebrate Tectonin proteins, we performed domain- and conserved position-specific iterated sequence searches, and identified a potential human homolog, which we dubbed the hTectonin. We also discovered that the domain architecture of hTectonin is well conserved throughout the different species, suggesting that it is an important functional protein. Sequence motif analysis, and prediction of the secondary and tertiary structures suggests that hTectonin is a β-propeller protein, in accordance to the definition of the Tectonin domain. Specifically, only the Tectonin domains of hTectonin were found to interact with the fibrinogen-like domain of M-ficolin, an important complement initiator [12]. In addition, the hTectonin domains 6 and 11 also exhibited LPS-binding properties. The specificity of recognition of LPS by certain Tectonin domains is consistent with the invertebrate Tectonins such as the limulus GBP. We suggest that hTectonin forms a β-propeller structure involved in protein-protein interaction amongst host proteins and also in pathogen-detection, thus playing a vital role in bridging the host immune defense proteins to the invading pathogen, and this phenomenon is probably conserved over a vast number of organisms.

Results

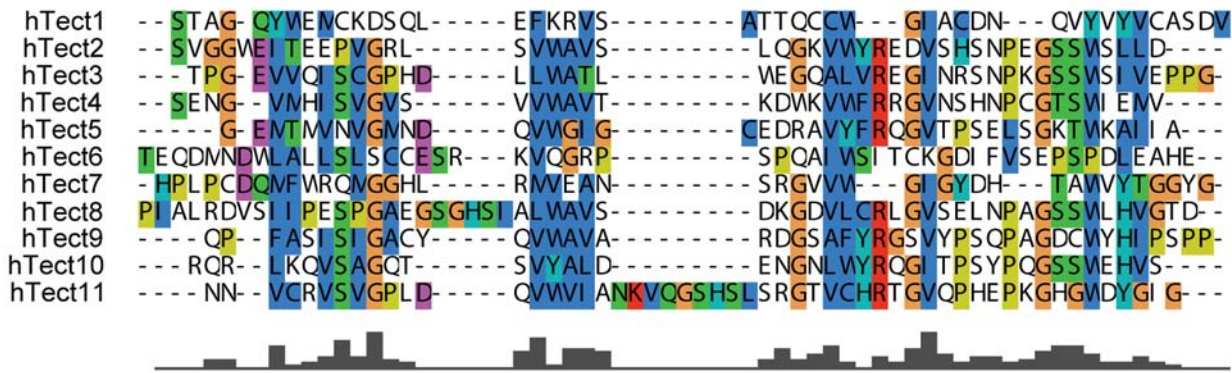
hTectonin identified from the human genome database – a hypothetical protein?

In mammals, the identity and role of proteins with Tectonin domains are unknown. Those identified or studied in the invertebrates [2–4,6,13–21] as well as the first vertebrate, fish, exhibit immune defense properties. Here, we sought to examine whether the Tectonin domains are structurally and functionally conserved in the mammals. A position-iterated search using known Tectonin domain-containing proteins in the invertebrates revealed a family of vertebrate Tectonin proteins to be distantly related (Figure 1A). This includes the human protein, Q7Z6L1, which is one of 3 human hypothetical proteins (GenBank Accession No.s: Q7Z6L1, Q15040 and O95714) that contain the Tectonin domain architecture [22,23], when a domain architectural search was done on Tectonin domain-containing proteins. Q7Z6L1 codes for a predominantly Tectonin domain-containing protein (Figure 1B), suggesting that the domains probably form an essential part of the molecular structure and play a vital role. Furthermore, the high architectural homology of Q7Z6L1, from the slime mold to the human, suggests its evolutionary conservation and functional significance (Figure 2). We thus selected Q7Z6L1 which codes for ‘hTectonin’ for molecular expression, and further structural and functional analyses.

hTectonin consists of β-propeller secondary structure

From the multiple sequence alignment (MSA) of the Tectonin domains, we confirmed a pattern of sequence repeats of 40 to 50

Sequence conservation in the Tectonin domains β-sheets



Secondary structure prediction (H=helix, C=coil, E=beta sheet)

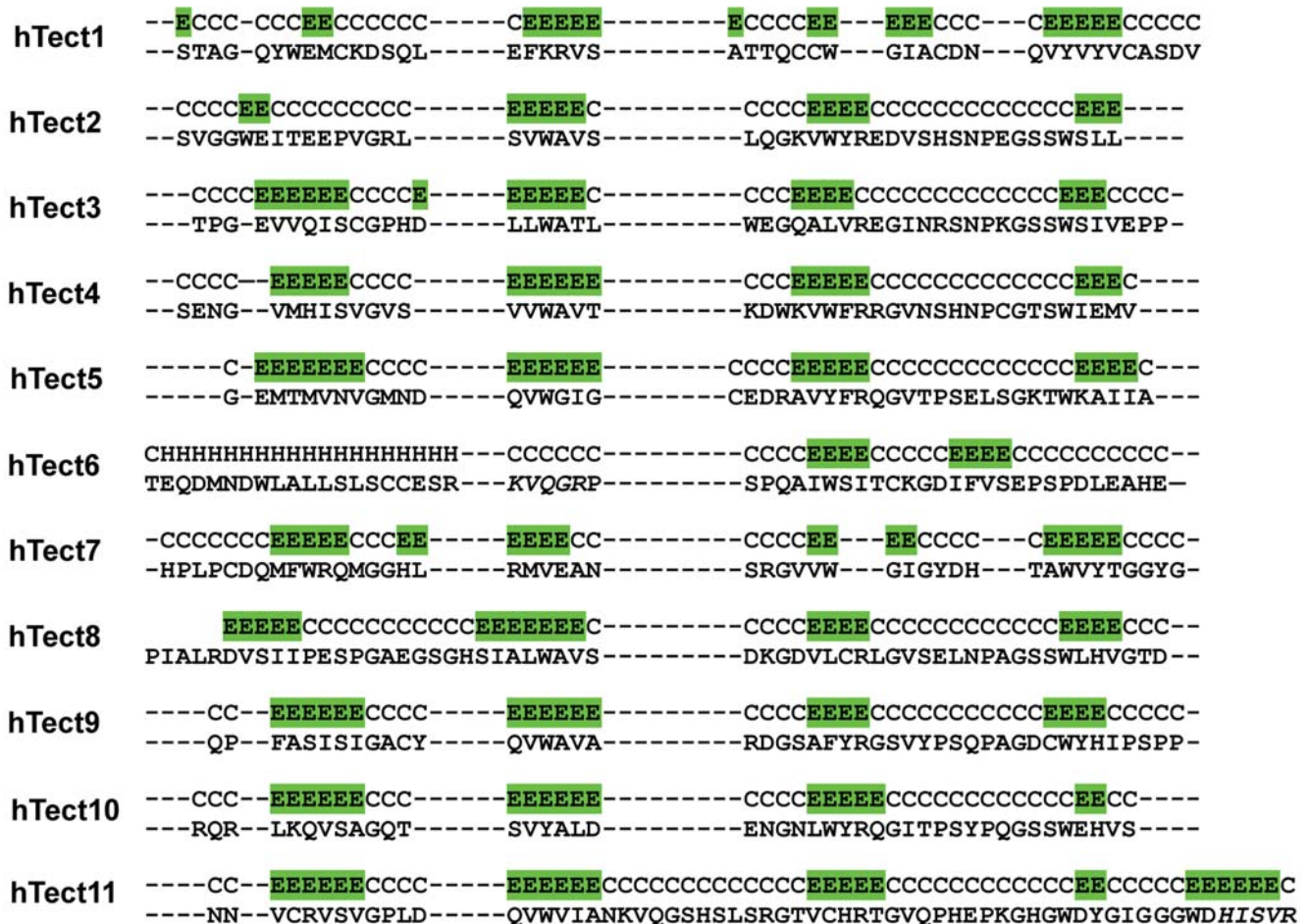


Figure 3. hTectonin forms β-sheets in its Tectonin repeats. CLUSTALW alignment of the individual Tectonin domains and PSIPRED secondary structure prediction indicates that the 11 Tectonin domains of hTectonin contain 4 highly conserved repeats that form β-strands (highlighted in green), a motif that is characteristic of the β-propeller fold. The LPS-binding motifs are in red font. E, β-sheet; C, Coil; H, Helix. doi:10.1371/journal.pone.0006260.g003

residues in length, which is a unique characteristic of β-propellers [9,10]. In addition, secondary structure prediction of hTectonin by PSIPRED [24,25] predicted these conserved repeats to form the β-strands of a β-sheet topology, consistent with β-propeller architecture (Figure 3).

hTectonin interacts with ficolin through its Tectonin domains

Based on our observations that a Tectonin protein, GBP (GenBank Accession No. AAV65031.1), interacts with two complement proteins, C-reactive protein (CRP) and carcinolec-

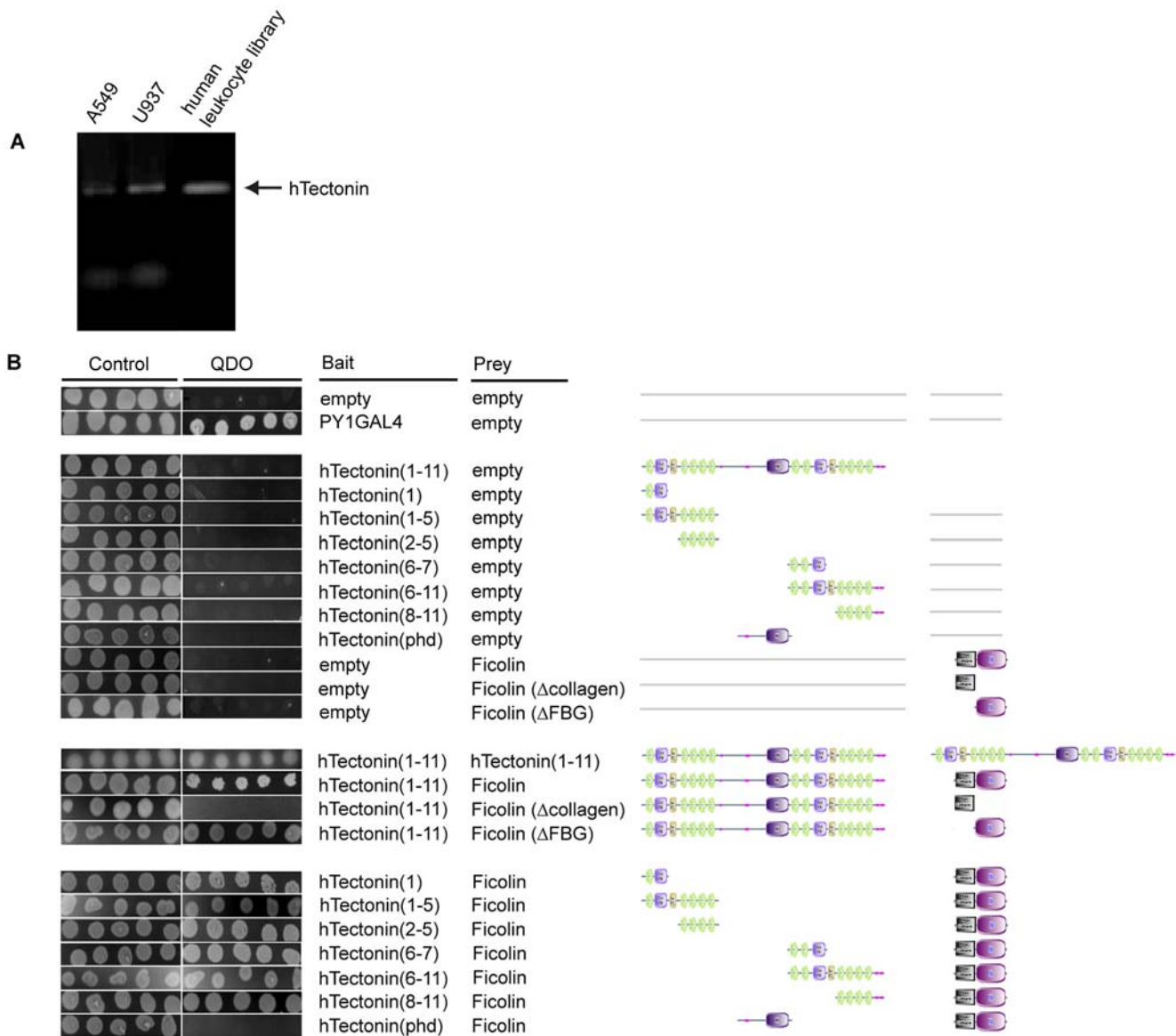


Figure 4. hTectonin exists and interacts with immune-related genes. (A) hTectonin cDNA is found in the human T cells (A549), monocytes (U937) and leukocytes. (B) hTectonin interacts with ficolin. Yeast 2-hybrid shows that hTectonin interacts (i) with itself, suggesting the possibility of oligomerization, as observed in other beta-propeller proteins; and (ii) with ficolin, a human complement protein. Furthermore, interaction with ficolin specifically occurs through the Tectonin domains of the hTectonins. This demonstrates a possible functional conservation of Tectonin domains since the Tectonin domains of GBP (horseshoe crab Tectonin lectin) was shown to interact with carcinolectin-5, a homologue of ficolin [26]. doi:10.1371/journal.pone.0006260.g004

tin (CL5), and is therefore immune-related, we reasoned that the hTectonin might play a similar role in immune defense. We tested and showed that the hTectonin gene is expressed in the human T cell line (A549), monocytes (U937) and the human leukocytes (Figure 4A), corroborating its immune relevance. Based on the rationale that (i) hTectonin is an architectural homolog of GBP and (ii) as a pathogen pattern-recognition receptor, GBP interacts with CL5 [26], which is a homolog of the human ficolin, we performed yeast 2-hybrid analysis using hTectonin as bait and the three isoforms of ficolins (L-, H- and M-ficolin) as prey. Results showed that the hTectonin (clone QZ7L1) interacts specifically with M-ficolin (GenBank Accession No. O00602) (Figure 4B). M-ficolin has in turn been shown to interact with the CRP [26]. Since both the CRP and M-ficolin

are key proteins of the complement classical and lectin pathways, respectively, this is the first evidence for the potential function of a human Tectonin domain-containing protein in frontline immune defense. Further delineation of hTectonin to isolate its functional domains showed that only the sub-clones expressing the predicted Tectonin domains interacted with M-ficolin. Furthermore, only the fibrinogen-like (FBG) domain of M-ficolin was shown to interact with the hTectonin, concurring with recent findings that the FBG domain is responsible for ligand-binding [12]. These results suggest that the protein-protein interaction between the *hypothetical* hTectonin and M-ficolin is not random, but structurally and positionally specific, and that the hTectonin is potentially involved in immune regulation, acting through its Tectonin domains.

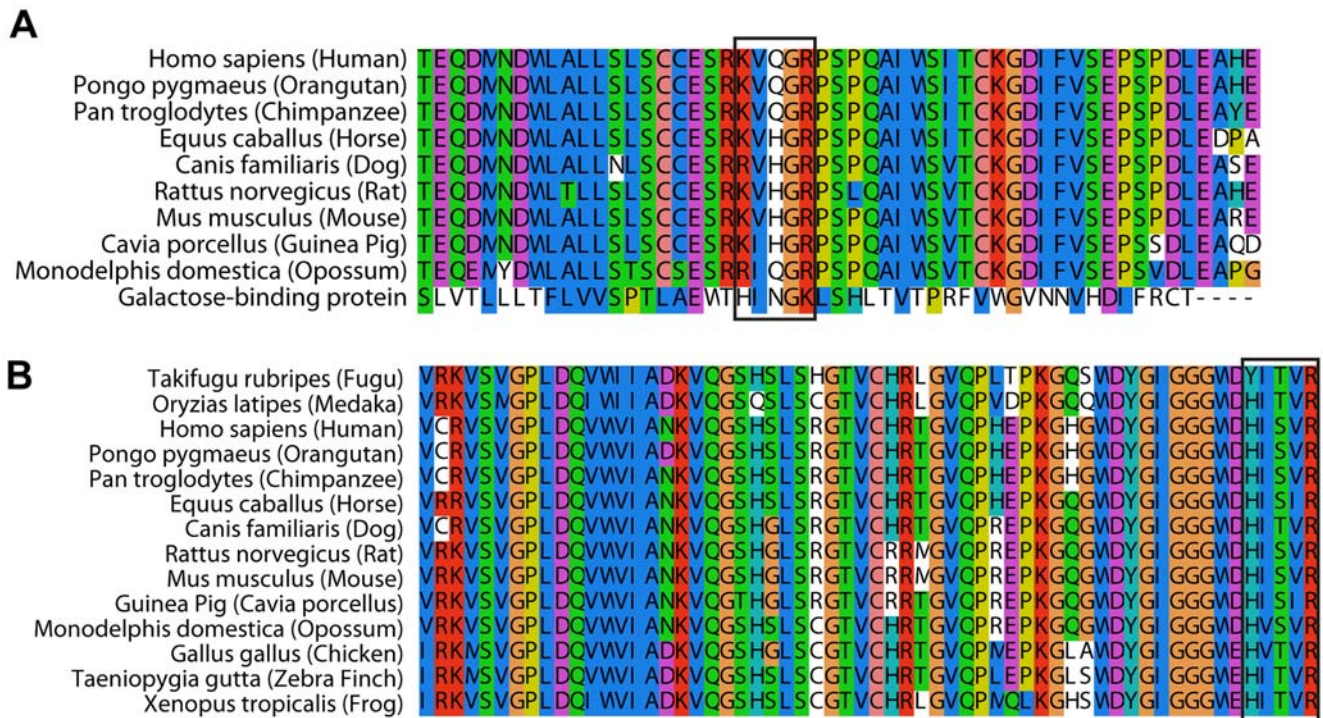


Figure 5. The LPS-binding motifs of hTectonin are conserved in other species. The LPS-binding motif of the pattern BHB(P)HB (blue box) in - (A) hTect6 and (B) hTect11 - are well-conserved in other species. doi:10.1371/journal.pone.0006260.g005

Tectonin domains harbor high avidity LPS-binding motifs

Gram negative bacterial endotoxin or lipopolysaccharide (LPS) is a prominent and well-studied representative pathogen-associated molecular pattern. Proteins harboring LPS-binding motifs, with alternating basic-hydrophobic/polar residues (BHB(P)HB), have been shown to bind LPS via the lipid A moiety [27,28], which is the most conserved bioactive pathophysiological centre of the LPS molecule (Supporting Figure S1A). Based on the BHB(P)HB pattern, we identified two such motifs in the 6th and 11th Tectonin domains of the hTectonin and found that these motifs were well-conserved among the mammalian homologs of hTectonin in addition to being in a region of high sequence conservation (Figure 5). Representative Tectonin peptides were synthesized around the BHB(P)HB motifs in Tectonin domains 6 & 11, and their efficacy of binding of lipid A was compared with peptides derived from the GBP Tectonin domains 1 & 6 (Supporting Figure S1B), where similar BHB(P)HB motifs exist. Real-time biointeraction of these Tectonin peptides to lipid A immobilized on biacore HPA chip showed that indeed the hTectonin peptides bound the lipid A at affinities of K_D 10^{-7} to 10^{-8} M, which are similar to the GBP peptides (Figure 6 and Table 1). We also showed that both the hTectonin and the GBP peptides exhibited similar level of binding affinity to ReLPS and LPS (Figure 6 and Supporting Figure S1 A,C). Table 1 summarises and compares the binding affinities of various peptides derived from the hTectonin and GBP. This corroborates our hypothesis and demonstrates the pathogen-binding ability of Tectonin domains and its functional conservation across species, from horseshoe crab GBP to human hTectonin.

Discussion

In order to classify and complete the functional characterization of the human proteome, many of the unknown proteins are usually

inferred from their counterparts in other species. This seems to be an easy option if the proteins share high sequence similarity, as they can be matched to each other by performing a simple sequence matching. However, the task is more complicated if the proteins do not show homology in their primary sequences. Nevertheless, many related proteins show conserved functionality more in terms of domain and structural conservation.

In this paper, we report our discovery of a human Tectonin protein, hitherto classified as being hypothetical. By structure-function analyses, we inferred its function as an immune-related protein. We showed that similar to its invertebrate counterparts, the hTectonin protein functions via its Tectonin domains. Furthermore, a distance PSI-BLAST sequence matching indicates that although the hTectonin shows low sequence homology, it is phylogenetically related to known proteins with Tectonin domains, functioning as immune proteins. By SMART domain comparison, we show that hTectonin contains multiple homologs widespread in the vertebrate kingdom, implying that it is not a one-off protein in the human proteome, but rather, an important one conserved throughout many species. We also discovered that the hTectonin gene is expressed in the human leukocytes. This is interesting, as a recent addition to the human database of proteins showed another human leukocyte Tectonin protein called the leukolectin (GenBank Accession No. ACM77812.1) [29], which also exhibits five Tectonin domain repeats (Figure 1B). This further implicates hTectonin to be immune-related. Like its limulus counterpart, GBP, which interacts with an important complement initiator (CRP), we find that the hTectonin also interacts with a cognate complement lectin, Ficolin. Furthermore, this interaction is specific, involving only the Tectonin domains within the hTectonin protein. We also identified LPS-binding motifs within two of the Tectonin domains which are located in the highly conserved sequence of the β-propeller fold. The affinity

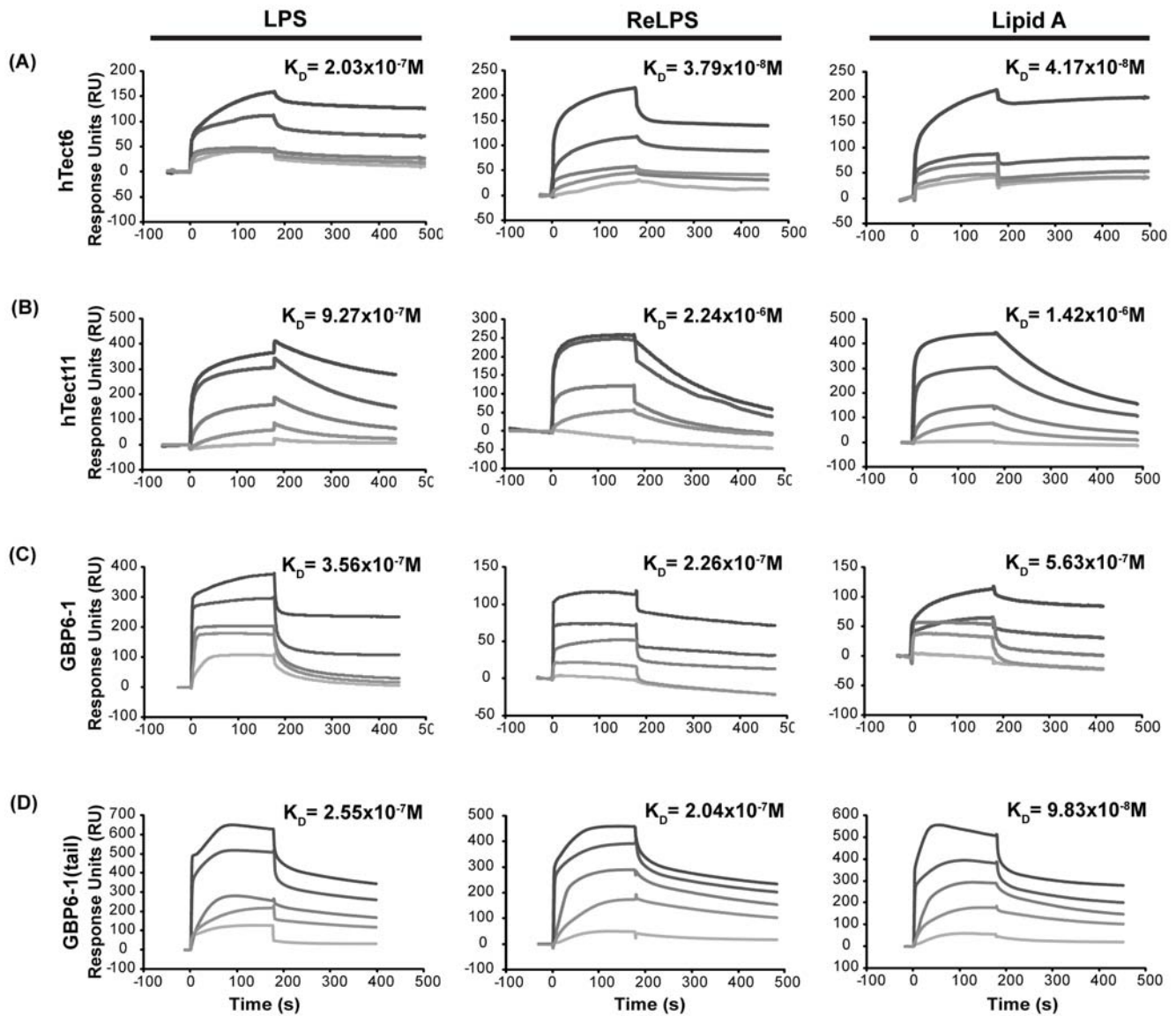


Figure 6. hTectonin peptides and GBP peptides bind LPS, ReLPS and lipid A with high affinity. hTectonin peptides, namely hTect6 and hTect11, containing the predicted LPS-binding motif [27,28] are able to bind the bacterial endotoxin. In GBP, the exclusion of the C-terminal tail loop which is not part of the β-propeller Tectonin structure (Supporting Figure S1B) - GBP6-1- gives similar binding affinity with the peptide designed to include this tail region, showing that it does not play an important role in the binding to lipid A, and that the function comes from within the Tectonin domains. Control peptides derived from non-Tectonin regions showed no binding to LPS (see Supporting Figure S1C), thus confirming the specificity of interaction with LPS via the Tectonin domains.
doi:10.1371/journal.pone.0006260.g006

of these motifs for bacterial LPS and the truncated active forms of the endotoxin molecule (ReLPS and lipid A) was verified experimentally. Thus, we propose that hTectonin is a novel human protein that forms a β-propeller structure which is involved in protein-protein interaction with immune-related proteins such as ficolin, and it simultaneously interacts with pathogens via PAMPs like LPS. Thus, the hTectonin plays a vital role in immune defense, which is conserved over a vast number of organisms.

Materials and Methods

Identification of Tectonin proteins

Tectonin domain containing proteins were identified using domain search on the SMART database [22,23]. A position-

specific iterated search using the primary sequence on PSI-Search on the EMBL server was performed using GBP as the query sequence. Related sequences were chosen after 2 iterations of PSI-Search. Hits were put through the SMART prediction server to confirm their propensity to form Tectonin domains. Multiple sequence alignment was carried out on the curated list of proteins using Promals3D [30]. A phylogenetic tree was then constructed from sequences showing strong domain alignments using PHYLIP [31] with a bootstrap value of 1000.

Human hTectonin cDNA clones

The hTectonin (Q7Z6L1) cDNA was obtained from iDNA Open-Biosystems (MHS1010-9205594). The cDNA was subcloned into pGBKT7 and pGADT7 vectors for the yeast 2-hybrid experiments.

Table 1. Dissociation constants of Tectonin peptides when bound to LPS, ReLPS, lipid A.

Bacterial ligand	Peptide	Sequence (LPS-binding motif underlined)	K _D (mol ⁻¹)
LPS	GBP6-1(tail)	KSCWLNPF <u>LAEW</u> THINGKLSH	2.55×10 ⁻⁷
	GBP6-1	FESVPASKA <u>EW</u> THINGKLSH	3.56×10 ⁻⁷
	hTectonin6	L ^S L ^S CCESR ^{KVQGR} PSPQAI	2.03×10 ⁻⁷
	hTectonin11	IGGGWDHISVRANATRAPRS	9.27×10 ⁻⁷
ReLPS	GBP6-1(tail)		2.04×10 ⁻⁷
	GBP6-1		2.26×10 ⁻⁷
	hTectonin6		3.79×10 ⁻⁸
	hTectonin11		2.24×10 ⁻⁶
Lipid A	GBP6-1(tail)		9.83×10 ⁻⁸
	GBP6-1		5.63×10 ⁻⁷
	hTectonin6		4.17×10 ⁻⁸
	hTectonin11		1.42×10 ⁻⁶

doi:10.1371/journal.pone.0006260.t001

Yeast-2-hybrid assay for protein-protein interaction

Co-transformations of the different bait and prey plasmids into *S. cerevisiae* AH109 strain were performed in accordance to [16]. The full-length and subclones of hTectonin and the ficolin cDNAs (without their signal sequences) were each fused to the DNA-binding domain of Gal4 in the bait plasmid pGBKT7 (BD Biosciences), or to the activation domain of Gal4 in the prey plasmid pGADT7-Rec (BD Biosciences). For selection, synthetic complete (SC) media lacking Leu and Trp (SC-Trp-Leu) or lacking Leu, Trp, His and adenine (quadruple dropout, QDO medium) were used. Transformants containing bait and prey plasmids were selected on SC-Trp-Leu by incubation for 3.5 days at 30°C. Resulting colonies were suspended in water and replated on SC-Trp-Leu and QDO agar at 30°C. The negative control was co-transformed with a recombinant plasmid and an empty prey or bait plasmid. The positive control was co-transformed with a plasmid expressing the full-length Gal4 transcriptional activator together with the empty pGADT7-Rec vector. Positive transformants were selected in SC media lacking Trp. hTectonin plasmid in pGBKT7 was then transformed into the library-positive yeast. DNA from resulting colonies from the co-transformation on QDO agar were extracted and identified through sequencing.

References

- Human Genome Project Information. Oak Ridge, Tennessee: U.S. Department of Energy Office of Science, Office of Biological and Environmental Research.
- Huh CG, Aldrich J, Mottahedeh J, Kwon H, Johnson C, et al. (1998) Cloning and characterization of Physarum polycephalum tectonins. Homologues of Limulus lectin L-6. *J Biol Chem* 273: 6565–6574.
- Schroder HC, Ushijima H, Krasko A, Gamulin V, Thakur NL, et al. (2003) Emergence and disappearance of an immune molecule, an antimicrobial lectin, in basal metazoa. A tachylectin-related protein in the sponge *Suberites domuncula*. *J Biol Chem* 278: 32810–32817.
- Kawabata S, Tsuda R (2002) Molecular basis of non-self recognition by the horseshoe crab tachylectins. *Biochim Biophys Acta* 1572: 414–421.
- Galliano M, Minchiotti L, Campagnoli M, Sala A, Visai L, et al. (2003) Structural and biochemical characterization of a new type of lectin isolated from carp eggs. *Biochem J* 376: 433–440.
- Kawabata S, Iwanaga S (1999) Role of lectins in the innate immunity of horseshoe crab. *Dev Comp Immunol* 23: 391–400.
- Basak S, Banerjee R, Mukherjee I, Das S (2009) Influence of domain architecture and codon usage pattern on the evolution of virulence factors of *Vibrio cholerae*. *Biochem Biophys Res Commun* 379: 803–805.

Peptide design and synthesis

The hTectonin protein sequence was scanned for LPS-binding motif. Two potential sites with the BHPHB pattern were found in hTectonin domains 6 (KVQGR) and 11 (HISVR). Henceforth, these peptides are referred to as hTec peptides (hTec6 and hTec11). The hTec peptide length and region surrounding the LPS-binding motif was chosen and optimized based on hydrophilicity and solubility values. The h-Tec6 was: LSLSCCESR^{KVQGR}PSPQAI and hTec11 was IGGGWDHISVRANATRAPRS. For comparison, one BHPHB site was found in the limulus GBP, Tectonin domain 1 (HINGK). The GBP peptides are: GBP6-1(tail) KSCWLNPF^{LAEW}THINGKLSH and GBP6-1(no tail) FESVPASKA^{EW}THINGKLSH, which are annotated based on the amino acid residues which encompass the domains 6-to-1. Peptides were also designed from the combination of GBP Tectonin domains 1 and 6. The peptides were synthesized by Genemed Synthesis, Inc., USA, and purified to >95% under pyrogen-free conditions.

Surface plasmon resonance analysis of the peptides

Surface plasmon resonance analysis for real-time biointeraction between the Tectonin peptides and bacterial LPS was performed using a Biacore 2000 instrument (Biacore AB). LPS, ReLPS and lipid A from *Salmonella minnesota* (List Biologicals, UK) were diluted to 0.25 mg/ml in 20 mM sodium phosphate, 150 mM NaCl, pH 7.4 and immobilized on the surface of an HPA sensor chip (Biacore AB) according to the manufacturer's specifications. Binding of the Tectonin peptides to the immobilized ligands was measured at a flow rate of 20 µl/min in 10 mM Tris, 150 mM NaCl, pH 7.4. Regeneration of the chip surface was achieved by injection of 20 µl 0.1 M NaOH until steady baseline was achieved. The dissociation constant, KD was calculated using BiaEvaluation software, version 3.2.

Supporting Information

Figure S1 (A) Structure of the bacterial LPS. LPS structure and the truncated forms, ReLPS and lipid A. (B) The structure of GBP, with the tail (circled) at the C-terminal end, which does not form the β-propeller structure of GBP. (C) Control Tectonin peptides which do not harbor the LPS-binding motif of BHPHB do not bind lipid A.

Found at: doi:10.1371/journal.pone.0006260.s001 (0.24 MB TIF)

Author Contributions

Conceived and designed the experiments: DHPL QY VF JLD. Performed the experiments: DHPL ZA QY. Analyzed the data: DHPL ZA VF BH JC JLD. Wrote the paper: DHPL ZA VF BH JC JLD.

15. Kuo TH, Chuang SC, Chang SY, Liang PH (2006) Ligand specificities and structural requirements of two Tachypleus plasma lectins for bacterial trapping. *Biochem J* 393: 757–766.
16. Le Saux A, Ng PM, Koh JJ, Low DH, Leong GE, et al. (2008) The macromolecular assembly of pathogen-recognition receptors is impelled by serine proteases, via their complement control protein modules. *J Mol Biol* 377: 902–913.
17. Iwanaga S (2002) The molecular basis of innate immunity in the horseshoe crab. *Curr Opin Immunol* 14: 87–95.
18. Ponting CP, Mott R, Bork P, Copley RR (2001) Novel protein domains and repeats in *Drosophila melanogaster*: insights into structure, function, and evolution. *Genome Res* 11: 1996–2008.
19. Saito T, Kawabata S, Hirata M, Iwanaga S (1995) A novel type of limulus lectin-L6. Purification, primary structure, and antibacterial activity. *J Biol Chem* 270: 14493–14499.
20. Iwanaga S, Lee BL (2005) Recent advances in the innate immunity of invertebrate animals. *J Biochem Mol Biol* 38: 128–150.
21. Mali B, Soza-Ried J, Frohme M, Frank U (2006) Structural but not functional conservation of an immune molecule: a tachylectin-like gene in *Hydractinia*. *Dev Comp Immunol* 30: 275–281.
22. Letunic I, Copley RR, Pils B, Pinkert S, Schultz J, et al. (2006) SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* 34: D257–260.
23. Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95: 5857–5864.
24. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292: 195–202.
25. Bryson K, McGuffin LJ, Marsden RL, Ward JJ, Sodhi JS, et al. (2005) Protein structure prediction servers at University College London. *Nucleic Acids Res* 33: W36–38.
26. Ng PM, Le Saux A, Lee CM, Tan NS, Lu J, et al. (2007) C-reactive protein collaborates with plasma lectins to boost immune response against bacteria. *EMBO J* 26: 3431–3440.
27. Frecer V, Ho B, Ding JL (2000) Interpretation of biological activity data of bacterial endotoxins by simple molecular models of mechanism of action. *Eur J Biochem* 267: 837–852.
28. Frecer V, Ho B, Ding JL (2004) De novo design of potent antimicrobial peptides. *Antimicrob Agents Chemother* 48: 3349–3357.
29. Miftari MH, Walther BT (2009) Molecular cloning and characterization of the leukolectin gene isolated from the human leukocytes.
30. Pei J, Kim BH, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 36: 2295–2300.
31. Retief JD (2000) Phylogenetic analysis using PHYLIP. *Methods Mol Biol* 132: 243–258.
32. Thompson JD, Gibson TJ, Higgins DG (2002) Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*; Chapter 2: Unit 2.3.
33. Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189–1191.

The Macromolecular Assembly of Pathogen-Recognition Receptors is Impelled by Serine Proteases, via Their Complement Control Protein Modules

Agnès Le Saux¹, Patricia Miang Lon Ng², Joanne Jing Yun Koh¹, Diana Hooi Ping Low^{1,3}, Geraldine E-Ling Leong¹, Bow Ho^{4†} and Jeak Ling Ding^{1,3*†}

¹Department of Biological Sciences, National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore

²NanoBioanalytics Laboratory, Faculty of Engineering, National University of Singapore, Singapore 117574, Singapore

³Singapore-MIT Alliance, Computational Systems Biology, Singapore 117576, Singapore

⁴Department of Microbiology, Yong Loo Lin School of Medicine, National University of Singapore, 5 Science Drive 2, Singapore 117597, Singapore

Received 18 September 2007;
received in revised form
11 January 2008;
accepted 16 January 2008
Available online
30 January 2008

Edited by J. Karn

Although the innate immune response is triggered by the formation of a stable assembly of pathogen-recognition receptors (PRRs) onto the pathogens, the driving force that enables this PRR–PRR interaction is unknown. Here, we show that serine proteases, which are activated during infection, participate in associating with the PRRs. Inhibition of serine proteases gravely impairs the PRR assembly. Using yeast two-hybrid and pull-down methods, we found that two serine proteases in the horseshoe crab *Carcinoscorpius rotundicauda* are able to bind to the following three core members of PRRs: galactose-binding protein, Carcinolectin-5 and C-reactive protein. These two serine proteases are (1) Factor C, which activates the coagulation pathway, and (2) C2/Bf, a protein from the complement pathway. By systematic molecular dissection, we show that these serine proteases interact with the core “pathogen-recognition complex” via their complement control protein modules.

© 2008 Elsevier Ltd. All rights reserved.

Keywords: pathogen-recognition receptor assembly; serine proteases; protein–protein interactions; innate immunity; yeast two-hybrid method

*Corresponding author. E-mail address: dbsdjl@nus.edu.sg.

† B.H. and J.L.D. are senior coauthors.

Abbreviations used: PRR, pathogen-recognition receptor; LPS, lipopolysaccharide; CRP, C-reactive protein; CL5, Carcinolectin-5; GBP, galactose-binding protein; Tryp-SP, trypsin-like serine protease; FC, Factor C; CCP, complement control protein; GlcNAc, N-acetylglucosamine; MS, mass spectrometry; TBS, Tris-buffered saline; QDO, quadruple dropout; GST, glutathione S-transferase; SC, synthetic complete.

Introduction

The innate immune system uses germline-encoded pathogen-recognition receptors (PRRs), which are evolutionarily conserved.¹ The PRRs recognize conserved microbial cell wall components referred to as pathogen-associated molecular patterns, such as lipopolysaccharide (LPS) from Gram-negative bacteria.^{2,3}

The horseshoe crab relies solely on its innate immune system to combat microbial infection.^{4–9} We have recently shown that the *Carcinoscorpius rotundicauda* C-reactive protein (CRP), Carcinolectin-5

(CL5) and the galactose-binding protein (GBP) constitute the core complex of pathogen-recognition assembly that binds to invading pathogens.¹⁰ CRP is a major hemolymph protein that is conserved through evolution.^{11–13} CL5 is homologous to human ficolins. GBP was isolated from the Japanese horseshoe crab *Tachypleus tridentatus* as a Sepharose-binding protein.^{14,15} It is associated with a trypsin-like protease activity when chromatographed through an LPS-Sepharose affinity column.¹⁵ However, the serine protease(s) that coeluted with GBP remains unidentified.

In the horseshoe crab hemolymph, the two innate immune pathways that function via serine proteases are complement activation and blood coagulation. C2/Bf, a functional homolog of the vertebrate trypsin-like serine proteases (Tryp-SPs) C2 and Bf, is involved in C3 activation in *C. rotundicauda* hemolymph.⁷ Circulating C2/Bf zymogen could be the serine protease associated with GBP. On the other hand, the *C. rotundicauda* Factor C (FC), a serine protease sensitive to LPS,^{16–24} could be another candidate. During Gram-negative infection, LPS activates FC, which initiates the coagulation pathway to incapacitate the invading pathogen. FC is expressed and mainly stored in granules within hemocytes. It is also synthesized and secreted by the hepatopancreas.²⁵ This strongly suggests a circulating form of the FC zymogen, which acts as a sentinel to rapidly respond to the LPS, to initiate the degranulation of hemocytes and to trigger the coagulation cascade. FC is also found on the hemocyte membrane, enhancing its exposure to the invading pathogens.²⁶ These observations reinforce the assumption that FC plays a key role in frontline immune response and could be the endopeptidase associated with GBP.¹⁵ Both C2/Bf and FC are multidomain proteins. They each comprise, among other domains, five complement control protein (CCP) modules, also known as short consensus repeats or Sushi domains, due to their secondary folding structure²⁷ and homology to C1r and C1s complement factors.²⁸

Here, we show that serine proteases previously thought to be responsible for activating downstream functions of immune effectors also participate in regulating upstream pathogen recognition. In the presence of protease inhibitors, the binding of GBP and CL5 to Sepharose beads and to bacteria is significantly reduced. In contrast, the pathogen-induced recruitment of CRP to the bacterial surface was previously shown to be unaffected by serine proteases.¹⁰ Therefore, in view of gaining insights into the mechanisms underlying the interactions among these PRRs, it is pertinent to identify and characterize the serine protease(s) involved in driving these interactions.

Using a yeast two-hybrid approach, we observed that the CCP repeats of the two serine proteases, FC and C2/Bf, harbor strong interactions with GBP, CL5-C isoform and CRP-1 isoform, but not with CRP-2 isoform. Pull-down assays confirmed the interactions between GBP and the two serine pro-

teases. We propose that FC and/or C2/Bf is the serine protease associated with GBP. This association triggers a rapid degranulation of hemocytes, leading to blood coagulation and/or complement activation, respectively, thence the destruction of the pathogens. Based on a protein–protein interaction map of different PRRs, we propose a model to explain how serine proteases drive a dynamic macromolecular assembly of PRRs, contributing to a fast and efficient immune response.

Results

The profile of PRRs copurified with GBP is altered by protease inhibitors

GBP is a galactose-recognizing lectin that binds to Sepharose beads, a polymer made of alternating units of galactose and 3,6-anhydrogalactose. Since CNBr-activated Tris-reacted Sepharose beads are reported to bind human plasma lectins such as ficolins^{29,30} (CL5 homologue), we chose this resin to isolate other hemolymph lectins together with GBP and to test whether the inhibition of serine proteases would affect the profile of the proteins copurified. The bound proteins were eluted with 0.4 M *N*-acetylglucosamine (GlcNAc), a ligand that dissociates GBP from the beads. The protease inhibitor cocktail greatly reduced the binding of GBP (p26 and p52), and proteins of p35–40, to the Sepharose beads (Fig. 1a). p26 and p52 were confirmed by Western blot to be a GBP monomer and a GBP dimer, respectively (Fig. 1a). Mass spectrometry (MS) identified p35 as CL5-C (Fig. 1b). Presumably, p40 was CL5-B.¹⁰ Another protein, p34 (Fig. 1a), was detected by anti-Sushi-1 antibody (Fig. 1a). This antibody is directed against the first CCP domain of FC, a well-exposed region in the FC protein^{18,23} (see Materials and Methods and Supplementary Fig. 2a) and, therefore, may react with a proteolytic fragment containing CCP from FC or other CCP-domain proteins. The same is observed with naïve and infected hemolymph (Fig. 1a). We also note that the binding effect induced by protease inhibitors is specific, as some hemolymph proteins are affected, while others either are unaffected or, in the case of hemocyanin (p72–75), even display enhanced binding to the Sepharose beads (Fig. 1a). Although, at this juncture, it is unclear why the binding of hemocyanin is improved when serine proteases are inactivated, this observation suggests that proteases can up-regulate or down-regulate the assembly of PRRs to pathogens. Taken together, these results indicate that the efficiency of recruitment of a panel of PRRs is strongly dependent on serine proteases.

The binding of GBP, CL5 and hemocyanin to bacteria is regulated by a serine protease

To test whether the binding of PRRs to bacteria could also be affected by protease inhibitors as much

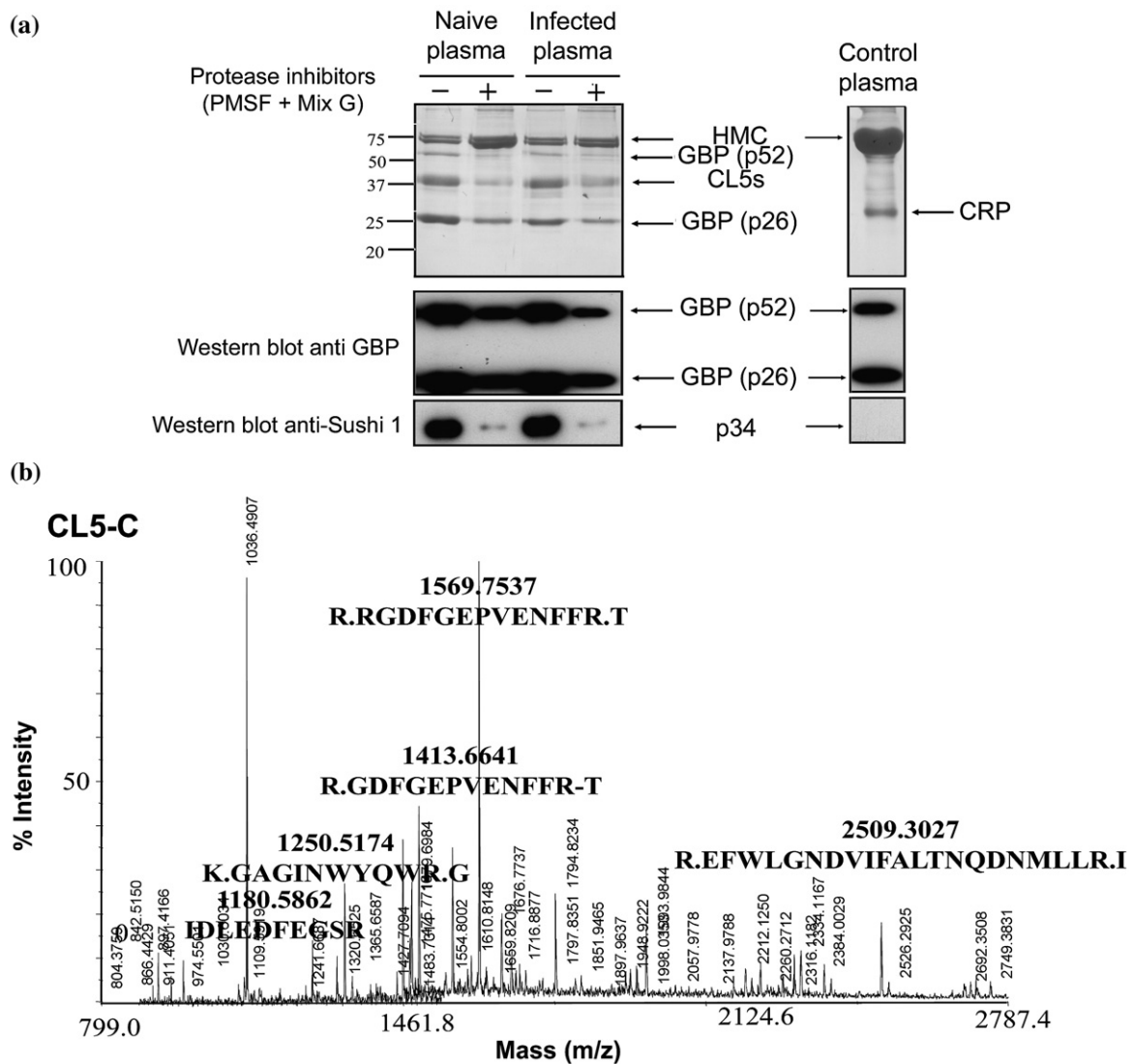


Fig. 1. Serine protease mediates the recruitment of GBP and CL5 to Sepharose beads. (a) GBP, CL5 and a CCP-containing protein show enhanced interaction with CNBr-activated Tris-reacted Sepharose in the absence of protease inhibitors. Coomassie-blue-stained SDS-PAGE of the hemolymph proteins eluted from Sepharose beads with 0.4 M GlcNAc in the absence (-) or in the presence (+) of protease inhibitors (PMSF + Mix G). GBP (p26 and p52) is confirmed by Western blot (lower panel). The anti-Sushi-1 antibody reacts with a p34 protein, presumably a proteolytic fragment of either FC or C2/Bf (lower panel). A 0.2% aliquot of the naïve hemolymph was loaded in the last lane (control hemolymph). (b) Peptide mass fingerprint of trypsin-digested proteins from the p35 shows peaks belonging to CL5-C. The m/z values of peaks that are unidentified are in smaller font.

as that to Sepharose beads, we incubated *Pseudomonas aeruginosa* with either Tris-buffered saline (TBS; 100 mM Tris-Cl pH 7.4 and 150 mM NaCl) or 10% hemolymph in TBS for 30 min with and without phenylmethylsulfonyl fluoride (PMSF), a serine protease inhibitor. The presence of PMSF decreased the binding of p26, p40 and p52 to the bacteria (Fig. 2a). MS identification showed these proteins to be a GBP monomer and a GBP dimer (p26 and p52, respectively) and CL5-B (p40) (Supplementary Fig. 1). Scanning and integrating the intensity of the protein bands corresponding to GBP and CL5, using ImageJ 1.38 \times ,³¹ showed that about 2.5-fold more of these proteins are bound to bacteria when serine proteases are active. This is observed with both the

naïve and the infected hemolymph (Fig. 2a). Conversely, the binding of hemocyanin to bacteria appeared to be slightly improved when serine proteases were inhibited. A parallel study in our laboratory has shown that hemocyanin is susceptible to proteolysis by bacterial proteases.³² Thus, besides suppressing the serine proteases in the host's hemolymph, PMSF could have inhibited the *P. aeruginosa* proteases from proteolysing the hemocyanin, thus allowing it to accumulate on the bacteria.

It is possible that the binding of GBP and CL5 is unaffected by the proteases *per se* and that proteases merely aided in the elution step of the bound proteins. To verify this possibility, the same

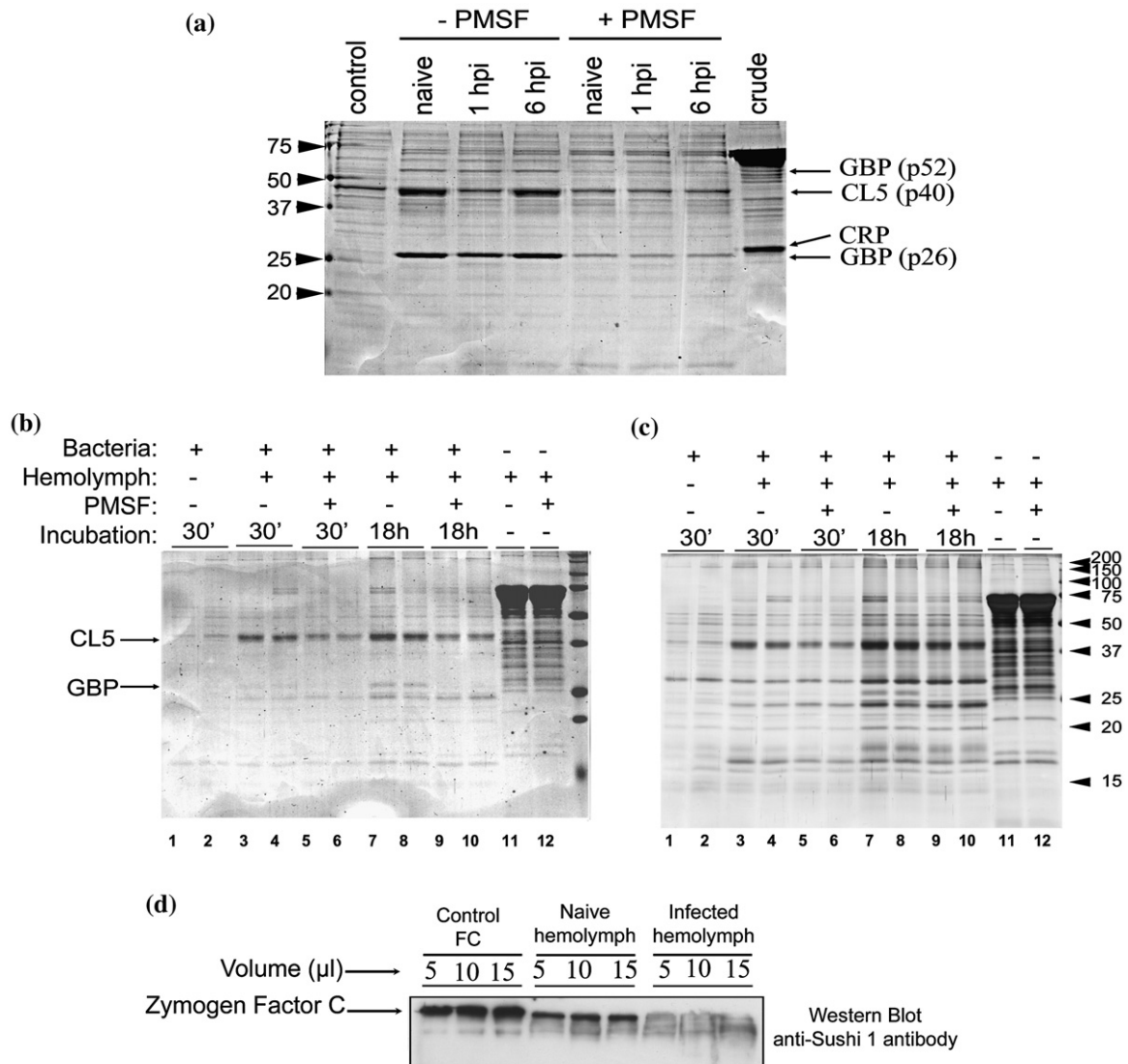


Fig. 2. GBP and CL5 showed enhanced interaction with bacteria in the absence of PMSF. *P. aeruginosa* was incubated with TBS or 10% hemolymph in TBS for 30 min in the absence (-) or in the presence (+) of PMSF. (a) Silver-stained SDS-PAGE gel of proteins eluted from the bacteria with 0.4 M GlcNAc. GBP and CL5 were identified by MS (Supplementary Fig. 1). A 0.2% aliquot of the hemolymph that was used in each treatment was loaded in the last lane. In (b) and (c), samples are treated similarly to (a), but bacteria were incubated with hemolymph for either 30 min or 18 h in the absence (-) or in the presence (+) of PMSF, and proteins were eluted with 4 M urea. This was performed to ensure a more complete elution of the proteins bound to bacteria and to demonstrate that active proteases have an intrinsic effect on the binding of GBP and CL5 to bacteria. SDS-PAGE gel was (b) Coomassie-blue-stained and (c) silver-stained. (d) FC circulates in the naïve hemolymph. Different amounts of naïve and infected hemolymph have been analyzed by Western blot using anti-Sushi-1 antibody. A full-length recombinant FC (~130 kDa; from the PyroGene® Assay kit; Lonza, Inc.) was used as a control.

experiment was repeated, but elution was performed with a strong denaturant (i.e., 4 M urea) instead of GlcNAc. Urea unfolds proteins and causes complete elution of the hemolymph proteins bound to bacteria. Results show that consistently more proteins are eluted (including some bacterial proteins), but most importantly, the amount of GBP and CL5 eluted was still more prominent when proteases were active during incubation with bacteria (Fig. 2b and c). This demonstrates that proteases impel the binding of GBP and CL5 to bacteria. Scanning the Coomassie-blue-stained gel

(Fig. 2b) and integrating the band intensities (ImageJ 1.38³¹) showed that about 2.3- and 1.8-fold more GBP and CL5, respectively, bound to bacteria when the proteases were active, a result close to the one observed when elution was performed with GlcNAc. Altogether, these results confirm our observation that GBP is associated with a serine protease¹⁵ that upregulates the binding of GBP and CL5 to the bacteria, while it downregulates the binding of hemocyanin. Nevertheless, the identity of this serine protease remains to be elucidated.

The FC zymogen is present only in the naïve hemolymph

Immunodetection using anti-Sushi-1 antibody suggests that a CCP-domain protein, potentially FC, could be the serine protease that copurifies with GBP in the absence of PMSF (Fig. 1a). The multi-domain organization of FC and its process of activation are shown in Supplementary Fig. 2a. The FC is located in large intracellular granules,⁴ as well as on the surface of the hemocytes.²⁶ Since expression and secretion from the hepatopancreas have been reported for FC,³³ it is conceivable that FC is present in the naïve hemolymph at a low level. When we analyzed for its presence in hemolymph, a full-length CCP-containing protein at a size expected for FC (>100 kDa) was observed in the naïve, but not in the infected, hemolymph (Fig. 2d). The recombinant FC provided in the PyroGene® Recombinant Factor C Assay kit (Lonza, Inc.) was used as a control. This prompted us to assume that a low level of the FC zymogen circulates in the naïve hemolymph. Nevertheless, we cannot fully exclude the possibilities that, (i) during exsanguination, some hemocyte exocytosis may have occurred to release intracellular FC to the hemolymph, and (ii) the anti-Sushi-1 antibody may cross-react with other CCP-containing proteins present in the hemolymph such as C2/Bf, the other known serine protease zymogen that also contains CCP domains. During infection, FC is activated, resulting in the release/loss of the N-terminal domain containing the CCP modules (Supplementary Fig. 2a). Absence of a full-length CCP-containing protein in infected hemolymph further suggests that what is detected in the naïve hemolymph is a zymogen that becomes undetectable in the infected hemolymph.

FC interacts with GBP, CL5-C and CRP-1

We used the yeast two-hybrid approach to verify whether FC interacts with GBP and CL5-C, the two PRRs that bind better to bacteria in the absence of protease inhibitors. We also tested for an interaction between FC and the CRP-1 and CRP-2 isoforms. Even though the binding of CRP to bacteria is not directly affected by protease inhibitors, the CRP still constitutes the core of PRRs, which forms the stable pathogen-recognition complex.¹⁰

We first ruled out the possibility of an auto-activation by cotransformation of the yeast with GBP, CL5-C, CRP-1, CRP-2 or FC cloned in pGBKT7 and pGADT7-Rec (empty plasmid). Transformants were verified for the absence of growth in a quadruple dropout (QDO) medium (Supplementary Data) (Fig. 3a and b). Next, we analyzed whether the full-length FC could interact with GBP, CL5-C, CRP-1 and CRP-2. The growth on QDO plates shows that FC interacts with GBP, CRP-1 and CL5-C (Fig. 3b), but does not interact with the CRP-2 isoform.

The interaction between GBP and FC was further confirmed by glutathione S-transferase (GST) pull-

down assay. The recombinant GST-GBP fusion protein was used to pull down the FC from the PyroGene® Assay kit (Lonza, Inc.). After pulldown with and without protease inhibitors, we looked for the presence of FC with the anti-Sushi-1 antibody. Figure 3c shows that FC interacts with GST-GBP (lanes 4 and 5), but not with GST (lanes 2 and 3). Additionally, we show that under non-pyrogen-free condition, LPS induced the activation of the zymogen FC, resulting in a proteolytic fragment p70 (lanes 4 and 5), thus further authenticating the presence of a functional FC. When FC was auto-activated by LPS and proteolysed, the presence of protease inhibitors had no effect on the FC interaction with PRRs or on its control over the PRR-PRR interaction.

To map the region of FC that interacts with the PRRs, we subcloned various fragments of FC (Supplementary Figs. 3 and 4) as fusions to Gal4 DNA-binding domain in the pGBKT7 vector and studied their interaction with GBP, CRP-1, CRP-2 or CL5-C into the yeast. Supplementary Figs. 3 and 4 show that the region encompassing the five CCP modules (FC CCP1-CCP5) specifically and strongly interacts with GBP, CRP-1 and CL5-C. Single CCP modules exhibited interesting results: CCP1, which interacts with LPS,²¹⁻²³ does not interact at all with those PRRs. CCP5, which is located proximal to the Tryp-SP domain, strongly interacts with GBP and CL5-C, but more weakly with CRP-1. The combination of CCP5 and the Tryp-SP domains showed a strong interaction with GBP and CL5-C, but a weaker interaction with CRP-1 (Supplementary Figs. 3 and 4). The first four CCP modules of FC showed either some autoactivation or no interaction. The proline-rich domain shows a very weak interaction with CL5-C only, and the Tryp-SP domain weakly interacts with CL5-C and GBP (Supplementary Figs. 3 and 4). At this juncture, no conclusion can be drawn on why some regions of FC showed autoactivation in the yeast two-hybrid assay. Nevertheless, the CCP5 module appears to be crucial for the specific binding to GBP and CL5-C, and, to a lesser extent, to CRP-1. Being the last module to be released during the process of FC activation (Supplementary Fig. 2a), it would be interesting in the future to study the structure-function relationship of CCP5 in more detail.

The 5xCCP from C2/Bf interacts with GBP, CL5-C and CRP-1

Another CCP-containing serine protease, the C2/Bf, had been previously identified in our laboratory.⁷ C2/Bf is found in the hemolymph and is involved in complement activation. Therefore, we tested whether this immune-response-related CCP-containing protease is also able to interact with PRRs. The multidomain organization and process of activation of C2/Bf are depicted in Supplementary Fig. 2b. We used the yeast two-hybrid approach to test for interaction between C2/Bf and the CL5-C, CRP-1 and CRP-2 isoforms. The 5xCCP region of

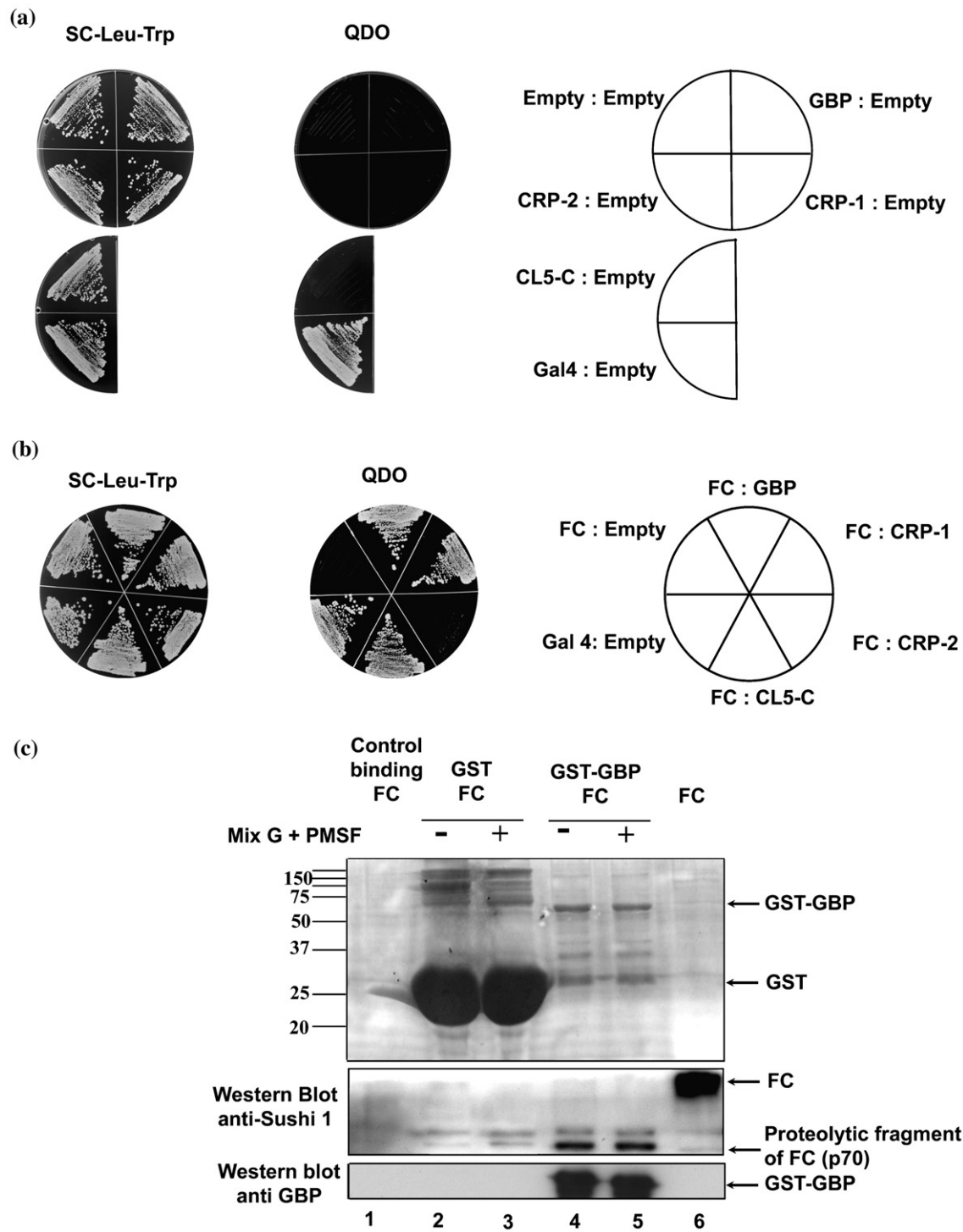


Fig. 3. FC interacts with GBP, CRP-1 and CL5-C, but not with CRP-2. (a) Controls showing the absence of autoactivation of the Gal4 promoter in yeast cells by the prey and bait proteins studied. Growth on SC-Leu-Trp (Trp and Leu dropouts) agar indicates the presence of both plasmids. Growth on QDO (QDO lacking Trp, Leu, His and Ade) agar indicates whether there is autoactivation when yeast cells are cotransformed with bait plasmids shown in the figure and an empty prey plasmid (pGADT7-Rec). (b) Yeast two-hybrid result shows a specific interaction between FC and GBP, CRP-1 and CL5-C, but not CRP-2. Growth on SC-Leu-Trp agar indicates the presence of both plasmids. Growth on QDO agar indicates whether there is autoactivation when yeast cells are cotransformed with pGBKT7-FC and an empty plasmid (pGADT7-Rec), or protein-protein interactions in the other cases. (c) GST pull-down assay confirming the interaction between FC and GBP. The upper panel shows the SDS-PAGE (Coomassie-blue-stained) of the proteins eluted after pulldown, and the two lower panels show Western blots using anti-Sushi-1 antibody and anti-GBP antibody, respectively.

C2/Bf was studied in parallel with the full-length C2/Bf.

Results show that the 5xCCP of C2/Bf interacts with GBP, CL5-C and CRP-1, but consistent with FC,

the 5xCCP also does not interact with the CRP-2 isoform (Fig. 4a). Interestingly, the full-length C2/Bf did not interact with CL5-C, CRP-1 and CRP-2, and only weakly with GBP after a longer incubation of

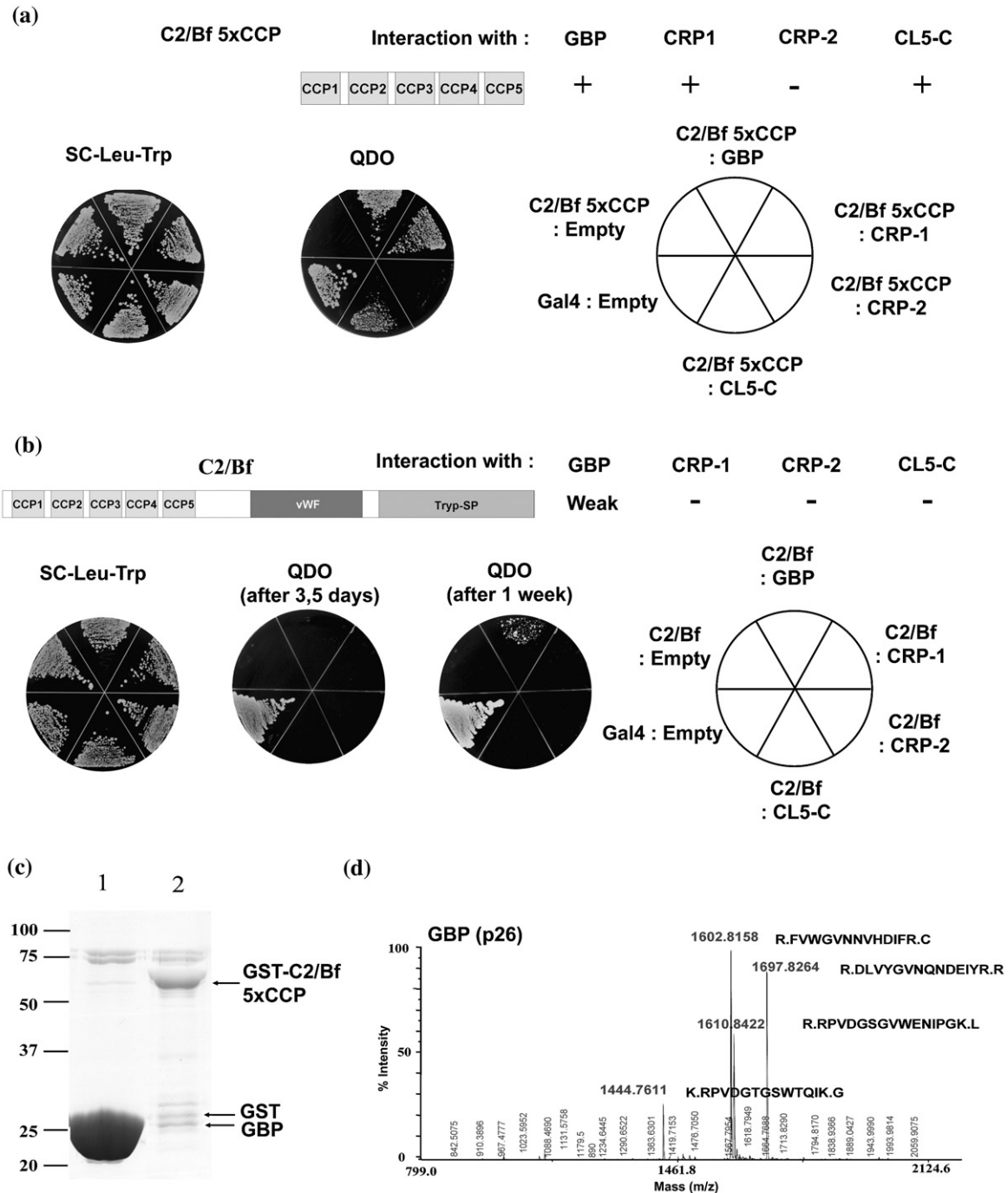


Fig. 4. Yeast two-hybrid method shows that C2/Bf 5xCCP interacts with GBP, CL5-C and CRP-1, but not with CRP-2. (a) The yeast cells were cotransformed, restreaked on SC-Leu-Trp and QDO plates, and incubated at 30 °C for 3.5 days, unless otherwise mentioned. Growth on SC-Leu-Trp agar indicates the presence of both plasmids. Growth on QDO agar indicates whether there is autoactivation when yeast cells are cotransformed with pGBKT7-C2/Bf 5xCCP and an empty plasmid (pGADT7-Rec), or protein–protein interactions in the other cases. (b) C2/Bf interacts weakly with GBP. The procedure was performed as described above. (c) GST pull-down assay using GST as control (lane 1) or GST–C2/Bf 5xCCP (lane 2) confirms the interaction between C2/Bf 5xCCP and GBP, and (d) peptide mass fingerprint of trypsin-digested proteins from the p26 shows peaks belonging to GBP. The *m/z* values of peaks that are unidentified are in smaller font.

7 days instead of 3.5 days (Fig. 4b). However, we cannot conclude whether the slow growth or lack of growth is due to a weak interaction or a problem linked to the lower level of C2/Bf in the nucleus available for interaction. Nevertheless, our results suggest that the 5xCCP of C2/Bf is specific and sufficient to interact with the three PRRs: GBP, CL5-C and CRP-1.

The binding of GBP to the 5xCCP of C2/Bf was further confirmed by GST pull-down assay. The recombinant GST–C2/Bf 5xCCP fusion protein was used to pull down interacting proteins from infected horseshoe crab hemolymph (Fig. 4c). A p26 protein associated with the fusion protein GST–C2/Bf 5xCCP, and not with the control GST, was identified by MS to be the GBP monomer (Fig. 4d), thus confirming the results observed by the yeast two-hybrid method.

Further delineation of the C2/Bf domains that interact with GBP, CL5-C and CRP-1

Different subfragments of C2/Bf were expressed in fusion to the Gal4 DNA-binding domain in pGBKT7. The individual CCP1 and CCP2 modules induced autoactivation, but single CCP3, CCP4 and CCP5 showed specific interactions with GBP, CL5-C and CRP-1. A combination of CCP1 and CCP2 specifically interacted with GBP, CL5-C and CRP-1, whereas a combination of CCP3 and CCP4 failed to interact with any of these PRRs (Supplementary Figs. 5 and 6). None of the fragments tested interacted with CRP-2. Thus, the CCP domains of C2/Bf specifically interact with the CRP-1 isoform. It is not clear why, individually, CCP3 and CCP4 interact with the PRRs, whereas in tandem, they fail to interact. Conversely, CCP1 and CCP2 individually showed unspecific responses, but in tandem, they interacted specifically with the PRRs. Further studies are required to clarify this dichotomy and to analyze whether the von Willebrand factor type A and Tryp-SP domains could also interact with some of the PRRs. Nevertheless, the single modules of CCP3, CCP4 and CCP5 are sufficient for the interaction with GBP, CL5-C and CRP-1.

In conclusion, we have consistently demonstrated that the CCP modules in FC and C2/Bf are necessary and sufficient for the interaction with GBP, CRP-1 and CL5-C. The FC and C2/Bf, and perhaps other serine proteases, are the driving forces that regulate the molecular assembly of the pathogen-recognition complex.

Discussion

Despite having only the innate immune system, horseshoe crabs have thrived in a microbiologically harsh habitat. Hemolymph PRRs are essential in the first line of defense to bind to the invading bacteria.¹⁰ In this article, we show that the binding of representative PRRs, such as GBP and CL5, to Sepharose beads, as well as to bacteria, is regulated

by serine proteases. Inhibition of serine proteases drastically reduced the binding of these PRRs (Figs. 1a and 2a–c). Thus, in addition to their commonly known functions in the activation of coagulation and complement pathways, respectively, FC and C2/Bf are also involved in regulating the binding of frontline PRRs to the pathogen surface.

Using a yeast two-hybrid approach and confirmation by pull-down methods, we showed that FC and/or C2/Bf impels the macromolecular assembly of PRRs on the pathogen. Firstly, FC, the initiator protein in the coagulation cascade that responds to Gram-negative infection, interacts with GBP, CL5 and CRP-1, but not with CRP-2 (Fig. 3b). The interaction with GBP has been confirmed by pull-down assay (Fig. 3c). The region encompassing the five CCP domains of FC is clearly required for this interaction, particularly the fifth CCP module, which is the most critical domain that interacts with the PRRs (Supplementary Figs. 3 and 4). Upon infection, FC is activated and triggers a G-protein-mediated exocytosis,²⁶ leading to the release of more FC, as well as the other innate immune molecules (Fig. 5). Secondly, consistent with the FC, the 5xCCP domain of C2/Bf, rather than the whole zymogen, binds to GBP, CL5 and CRP-1 (Fig. 4a and b). The interaction between the 5xCCP and GBP has also been confirmed by pull-down assay (Fig. 4c and d). Consistently, no interaction was found with the CRP-2 isoform, suggesting the specificity of the PRR–PRR collaboration. The ability of the other domains of C2/Bf to interact with the PRRs remains to be studied. C2/Bf is a complement protein circulating in the hemolymph. Once activated, C2/Bf participates in the activation of C3 to form the C3 convertase, which further activates the lectin complement pathway. We postulate that, under naïve conditions, C2/Bf may be in a conformation where its five CCPs are not exposed and, therefore, does not allow its binding to PRRs, since complement is not activated. During infection, the complement protein C3 binds to C2/Bf, which may change its conformation and exposes its 5xCCP region, thereby allowing its binding to the PRRs. C2/Bf can then be activated by Factor D (identified in the granules of horseshoe crab hemocytes³⁴) to form the C3 convertase (Fig. 5). Meanwhile, the cleaved C2/Bf 5xCCP fragment may remain bound to PRRs while attached to the pathogen surface. Therefore, such synchronized actions suggest that the interaction between C2/Bf and PRRs may help stabilize the macromolecular assembly to the pathogen. Additionally, *in vivo*, the presence of C3 bound to the pathogens may be a prerequisite for the binding of C2/Bf to the PRRs.

Taken together, we have demonstrated that the CCP modules of FC and C2/Bf are the crucial determinants of the interaction with three core PRRs. The CCPs are structures known to be important in protein–protein interactions.^{35,36} The CCP modules of FC and C2/Bf share 15.5–43.6% homology (Supplementary Fig. 7a), which may appear rather low. CCP modules exist in a wide

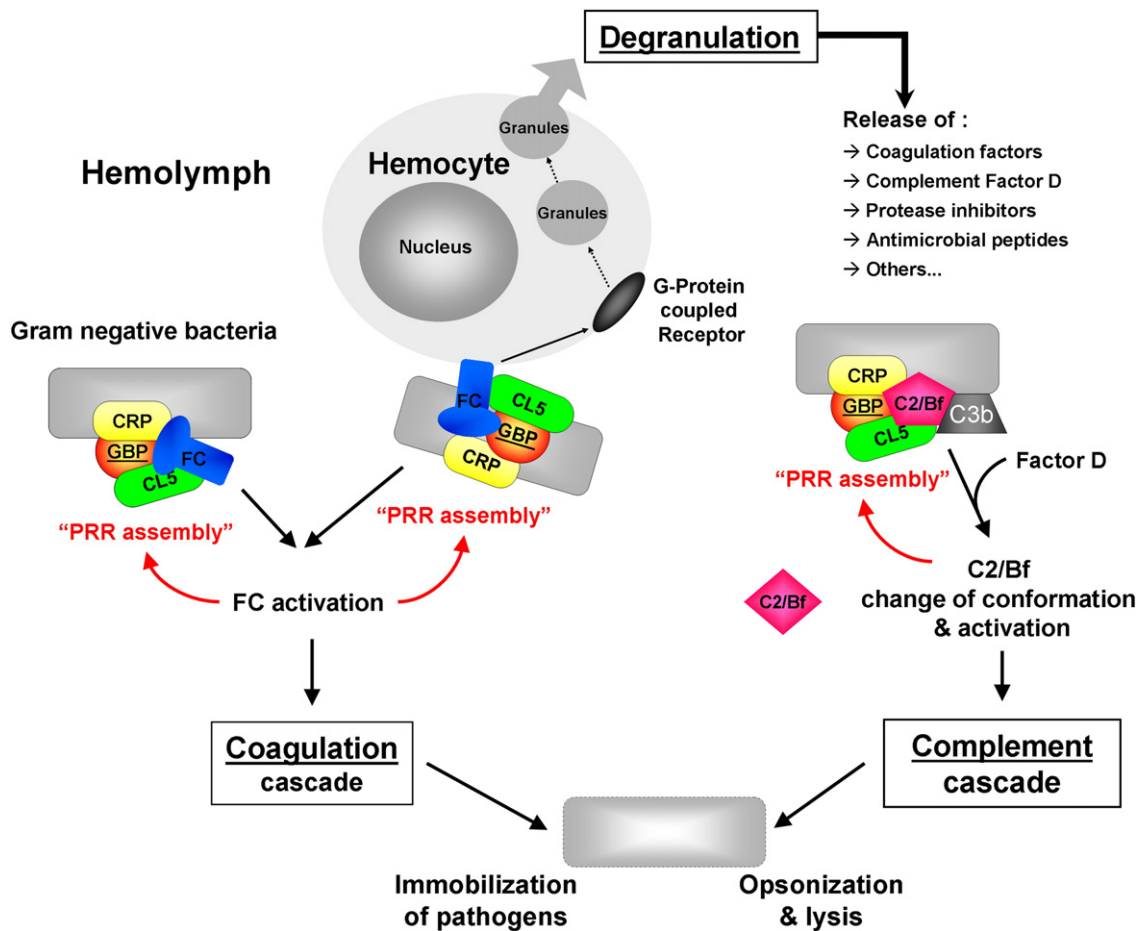


Fig. 5. Serine proteases regulate the assembly of PRRs to prompt the innate immune response. The model shows that FC and C2/Bf promote PRR assembly on pathogens. Up to now, serine proteases were known for only activating downstream functions of immune system, triggering rapid degranulation of hemocytes to release coagulation components and concomitantly activating both the coagulation cascade and the complement pathways to destroy the pathogens effectively. Our study shows that the positive feedback of the serine proteases in regulating the upstream pathogen-recognition assembly conceivably strengthens the immune response.

variety of complement and adhesion proteins, and their structure is known to be based on a β -sandwich arrangement: one face that is made up of three β -strands hydrogen-bonded to form a triple-stranded region at its center and the other face that is formed from two separate β -strands.^{35,37} Most of the CCP modules have four cysteine residues, with a glycine and a tryptophan highly conserved in similar positions (Supplementary Fig. 7b). The CCP5 of FC, which strongly interacts with GBP, CL5 and CRP-1, contains only three of the four highly conserved cysteines. The first one is replaced by a serine residue (Supplementary Fig. 7b). During the process of LPS-mediated activation of the FC, the N-terminal region is first cleaved off to result in an intermediate FC consisting of CCP5-Tryp-SP. Only in a second proteolysis is the CCP5 module processed to release a fully active FC enzyme (Supplementary Fig. 2a). Thus, it is interesting to note that CCP5, which is still able to interact with several PRRs, also remains as the final CCP linked to the Tryp-SP domain during the process of FC

activation. In FC, it has been shown that CCP1 and CCP3 bind LPS.^{18,21-24} Here, we observed that CCP1 does not interact with any of the PRRs studied. Thus, although the secondary structures of the CCP modules are expected to be very similar to one another, their abilities to interact with PRRs might remain very different. Thus, we propose that the CCP modules retain a unique identity and strong specificity towards their interacting partners. Therefore, *in vivo*, a tight regulation of the serine protease activities by protease inhibitors is envisaged. Nevertheless, further studies are needed to show how the molecular interactions between the serine proteases and PRRs are maintained and regulated *in vivo*.

In conclusion, our findings demonstrate the importance of serine proteases at the very frontline of the immune response, *viz.* the assembly of the PRRs onto the pathogens, in addition to their known roles in coagulation and complement activation. This is a new role described for serine proteases. In naïve hemolymph, serine proteases are present but inactive. During infection, they are activated, and

they impel the binding of PRRs to the pathogens through their CCP modules. A stable assembly of the PRRs incapacitates the microbial invader and boosts the immune response by triggering the degranulation of hemocytes, coagulation cascade and complement activation.

Materials and Methods

Organisms

Horseshoe crabs (*C. rotundicauda*) were collected from the Kranji estuary of Singapore. The animals were handled in accordance with national and institutional guidelines stipulated by the National Advisory Committee for Laboratory Animal Research, Singapore. The infection of horseshoe crabs and the bacterial strains used in this work are described in Supplementary Data.

The yeast strain used is *Saccharomyces cerevisiae* AH109 (BD Biosciences) (MATa, trp1-901, leu2-3,112, ura3-52, his3-200, gal4 Δ , gal80 Δ , LYS2::GAL1_{UAS}-GAL1_{TATA}-HIS3, GAL2_{UAS} GAL2_{TATA}-ADE2, URA3::MEL1_{UAS}-MEL1_{TATA}-LacZ and MEL1). The yeast culture conditions are described in Supplementary Data.

Biochemical reagents

Anti-mouse and anti-rabbit antibodies were obtained from DAKO. Anti-goat antibody was obtained from GE Healthcare. Anti-Sushi-1 antibody was raised in New Zealand white rabbits against a synthetic peptide derived from the CCP1 domain (Sushi 1) of FC. Its amino acid sequence is GFKLKGMARISCLPNGQWSNFPPKCIRES-CAMVSS. Anti-CRP and anti-GBP antibodies were raised in New Zealand white rabbits. Anti-CRP antibodies were raised against the *Limulus polyphemus* CRP (Sigma). This protein shows 80% sequence homology to the *C. rotundicauda* CRP. Specificity of the anti-CRP was confirmed by immunoprecipitation of its target antigen from hemolymph, followed by analysis of the antigen by MS. GBP was purified from the hemolymph by using CNBr-activated Tris-reacted Sepharose (GE Healthcare) and eluted with 0.4 M GlcNAc. The purity of GBP obtained by SDS-PAGE extraction was verified by MS and used to raise anti-GBP antibodies. The protease inhibitor cocktail Mix G was obtained from Serva. It contains 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride and aprotinin from bovine-lung-targeting serine proteases, E-64 (a cysteine protease inhibitor), leupeptin (a cysteine and Tryp-SP inhibitor), and ethylenediaminetetraacetic acid (which targets the metalloproteases). PMSF, which inhibits serine proteases, was obtained from Sigma. Recombinant FC is from the PyroGene[®] Recombinant Factor C Assay kit (Lonza, Inc.). All the cDNA, plasmid constructs and primers used for this work are described in Supplementary Data. GST pull-down assay conditions are also described in Supplementary Data.

Effect of protease inhibitors on the binding of PRRs to galactose

One milliliter of naïve or infected hemolymph was directly incubated with a 75% slurry of CNBr-activated Tris-reacted Sepharose. Briefly, CNBr-activated Sepharose 4 Fast Flow (Amersham and GE Healthcare) was

Tris-reacted and blocked with ethanolamine before overnight incubation at 4 °C (by end-over-end rotation) with hemolymph in the presence or in the absence of protease inhibitors (PMSF+Mix G). The beads were washed five times with ice-cold TBS (100 mM Tris-Cl pH 7.4 and 150 mM NaCl) before elution with 50 μ l of 0.4 M GlcNAc in TBS for 2 h at room temperature. Proteins eluted were analyzed with SDS-PAGE and identified with MS or Western blot (see Mass Spectrometry and Western Blots sections in Supplementary Data).

Effect of PMSF on the binding of PRRs to bacteria

For all treatments, bacteria were freshly grown for 2–3 h in tryptic soy broth, at 37 °C. Bacteria were washed thrice in saline and resuspended in a volume of TBS to yield an OD₆₀₀ of 10.0/ml. This suspension was then used as bacterial “beads” for incubation with horseshoe crab hemolymph. Before incubation, hemolymph was preclarified by centrifugation at 10,000g for 10 min and then diluted 10-fold in TBS in the presence or in the absence of 1 mM PMSF. A 1-ml final volume of diluted hemolymph was incubated with the bacteria. The hemolymph proteins bound to the bacteria were eluted with 0.4 M GlcNAc, analyzed by SDS-PAGE and silver-stained. To ensure a complete elution of all the hemolymph proteins bound to bacteria and thus to demonstrate that elution of protease-treated samples is not improved due to on-column proteolytic digestion, we also performed the elution with 4 M urea. The proteins eluted were analyzed by SDS-PAGE, Coomassie-blue-stained and silver-stained. The intensity of the Coomassie-blue-stained GBP and CL5 bands was densitometrically scanned and integrated using ImageJ 1.38 \times .³¹

Yeast transformation and yeast two-hybrid analysis

Cotransformations of the different bait and prey plasmids into *S. cerevisiae* AH109 strain were performed in accordance with standard protocols.³⁸ All the fragments and full-length FC, C2/Bf, CRP-1, CRP-2, GBP and CL5-C genes (without their signal sequence) were each fused to the DNA-binding domain of Gal4 in the bait plasmid pGBKT7 (BD Biosciences), or to the activation domain of Gal4 in the prey plasmid pGADT7-Rec (BD Biosciences). For selection, synthetic complete (SC) media lacking Leu and Trp (SC-Trp-Leu) or lacking Leu, Trp, His and adenine (QDO medium) were used (see Supplementary Data). Transformants containing bait and prey plasmids were selected on SC-Trp-Leu by incubation for 3.5 days at 30 °C. Resulting colonies were suspended in water and replated on SC-Trp-Leu and QDO agar at 30 °C for up to a maximum of 7 days. The proteins were tested for autoactivation by cotransformation of their recombinant plasmid with an empty prey or bait plasmid. The positive control was cotransformed with a plasmid expressing the full-length Gal4 transcriptional activator together with the empty pGADT7-Rec vector.

Accession numbers

The GenBank accession numbers of the gene and proteins studied in this article are as follows: GBP, AY647278; CRP-1 isoform, AY647271; CRP-2 isoform, AY647272; CL5-C, DQ250746; FC, S77063; C2/Bf variant 1, AY647279.

Acknowledgements

This work was supported by grants from the BioMedical Research Council (A*STAR) and Academic Research Fund (Tier 2 and MoE) awarded to J. L. Ding and B. Ho.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2008.01.045](https://doi.org/10.1016/j.jmb.2008.01.045)

References

- Kairies, N., Beisel, H. G., Fuentes-Prior, P., Tsuda, R., Muta, T., Iwanaga, S. *et al.* (2001). The 2.0-Å crystal structure of tachylectin 5A provides evidence for the common origin of the innate immunity and the blood coagulation systems. *Proc. Natl Acad. Sci. USA*, **98**, 13519–13524.
- Iwanaga, S. & Lee, B. L. (2005). Recent advances in the innate immunity of invertebrate animals. *J. Biochem. Mol. Biol.* **38**, 128–150.
- Medzhitov, R. & Janeway, C., Jr (2000). Innate immune recognition: mechanisms and pathways. *Immunol. Rev.* **173**, 89–97.
- Iwanaga, S. (2002). The molecular basis of innate immunity in the horseshoe crab. *Curr. Opin. Immunol.* **14**, 87–95.
- Iwanaga, S. & Kawabata, S. (1998). Evolution and phylogeny of defense molecules associated with innate immunity in horseshoe crab. *Front. Biosci.* **3**, D973–D984.
- Kurata, S., Ariki, S. & Kawabata, S. (2006). Recognition of pathogens and activation of immune responses in *Drosophila* and horseshoe crab innate immunity. *Immunobiology*, **211**, 237–249.
- Zhu, Y., Thangamani, S., Ho, B. & Ding, J. L. (2005). The ancient origin of the complement system. *EMBO J.* **24**, 382–394.
- Ding, J. L., Tan, K. C., Thangamani, S., Kusuma, N., Seow, W. K., Bui, T. H. *et al.* (2005). Spatial and temporal coordination of expression of immune response genes during *Pseudomonas* infection of horseshoe crab, *Carcinoscorpius rotundicauda*. *Genes Immun.* **6**, 557–574.
- Wang, X. W., Tan, N. S., Ho, B. & Ding, J. L. (2006). Evidence for the ancient origin of the NF- κ B/I κ B cascade: its archaic role in pathogen infection and immunity. *Proc. Natl Acad. Sci. USA*, **103**, 4204–4209.
- Ng, P. M. L., Le Saux, A., Lee, C. M., Tan, N. S., Lu, J., Thiel, S. *et al.* (2007). C-reactive protein collaborates with plasma lectins to boost immune response against bacteria. *EMBO J.* **26**, 3431–3440.
- Iwaki, D., Osaki, T., Mizunoe, Y., Wai, S. N., Iwanaga, S. & Kawabata, S. (1999). Functional and structural diversities of C-reactive proteins present in horseshoe crab hemolymph plasma. *Eur. J. Biochem.* **264**, 314–326.
- Ng, P. M., Jin, Z., Tan, S. S., Ho, B. & Ding, J. L. (2004). C-reactive protein: a predominant LPS-binding acute phase protein responsive to *Pseudomonas* infection. *J. Endotoxin Res.* **10**, 163–174.
- Robey, F. A. & Liu, T. Y. (1981). Limulin: a C-reactive protein from *Limulus polyphemus*. *J. Biol. Chem.* **256**, 969–975.
- Kuo, T. H., Chuang, S. C., Chang, S. Y. & Liang, P. H. (2006). Ligand specificities and structural requirements of two *Tachypleus* plasma lectins for bacterial trapping. *Biochem. J.* **393**, 757–766.
- Chiou, S. T., Chen, Y. W., Chen, S. C., Chao, C. F. & Liu, T. Y. (2000). Isolation and characterization of proteins that bind to galactose, lipopolysaccharide of *Escherichia coli*, and protein A of *Staphylococcus aureus* from the hemolymph of *Tachypleus tridentatus*. *J. Biol. Chem.* **275**, 1630–1634.
- Ding, J. L. & Ho, B. (2001). A new era in pyrogen testing. *Trends Biotechnol.* **19**, 277–281.
- Tan, N. S., Ho, B. & Ding, J. L. (2000). High-affinity LPS binding domain(s) in recombinant factor C of a horseshoe crab neutralizes LPS-induced lethality. *FASEB J.* **14**, 859–870.
- Tan, N. S., Ng, M. L., Yau, Y. H., Chong, P. K., Ho, B. & Ding, J. L. (2000). Definition of endotoxin binding sites in horseshoe crab factor C recombinant sushi proteins and neutralization of endotoxin by sushi peptides. *FASEB J.* **14**, 1801–1813.
- Li, A., Lee, P. Y., Ho, B., Ding, J. L. & Lim, C. T. (2007). Atomic force microscopy study of the antimicrobial action of Sushi peptides on Gram negative bacteria. *Biochim. Biophys. Acta*, **1768**, 411–418.
- Li, C., Ng, M. L., Zhu, Y., Ho, B. & Ding, J. L. (2003). Tandem repeats of Sushi3 peptide with enhanced LPS-binding and -neutralizing activities. *Protein Eng.* **16**, 629–635.
- Li, P., Sun, M., Ho, B. & Ding, J. L. (2006). The specificity of Sushi peptides for endotoxin and anionic phospholipids: potential application of POPG as an adjuvant for anti-LPS strategies. *Biochem. Soc. Trans.* **34**, 270–272.
- Li, P., Sun, M., Wohland, T., Ho, B. & Ding, J. L. (2006). The molecular mechanism of interaction between sushi peptide and *Pseudomonas* endotoxin. *Cell. Mol. Immunol.* **3**, 21–28.
- Li, P., Sun, M., Wohland, T., Yang, D., Ho, B. & Ding, J. L. (2006). Molecular mechanisms that govern the specificity of Sushi peptides for Gram-negative bacterial membrane lipids. *Biochemistry*, **45**, 10554–10562.
- Li, P., Wohland, T., Ho, B. & Ding, J. L. (2004). Perturbation of lipopolysaccharide (LPS) micelles by Sushi 3 (S3) antimicrobial peptide. The importance of an intermolecular disulfide bond in S3 dimer for binding, disruption, and neutralization of LPS. *J. Biol. Chem.* **279**, 50150–50156.
- Wang, L., Ho, B. & Ding, J. L. (2003). Transcriptional regulation of limulus factor C: repression of an NF κ B motif modulates its responsiveness to bacterial lipopolysaccharide. *J. Biol. Chem.* **278**, 49428–49437.
- Ariki, S., Koori, K., Osaki, T., Motoyama, K., Inamori, K. & Kawabata, S. (2004). A serine protease zymogen functions as a pattern-recognition receptor for lipopolysaccharides. *Proc. Natl Acad. Sci. USA*, **101**, 953–958.
- Ichinose, A., Bottenus, R. E. & Davie, E. W. (1990). Structure of transglutaminases. *J. Biol. Chem.* **265**, 13411–13414.
- Gaboriaud, C., Rossi, V., Fontecilla-Camps, J. C. & Arlaud, G. J. (1998). Evolutionary conserved rigid module-domain interactions can be detected at the sequence level: the examples of complement and blood coagulation proteases. *J. Mol. Biol.* **282**, 459–470.

29. Le, Y., Lee, S. H., Kon, O. L. & Lu, J. (1998). Human L-ficolin: plasma levels, sugar specificity, and assignment of its lectin activity to the fibrinogen-like (FBG) domain. *FEBS Lett.* **425**, 367–370.
30. Le, Y., Tan, S. M., Lee, S. H., Kon, O. L. & Lu, J. (1997). Purification and binding properties of a human ficolin-like protein. *J. Immunol. Methods*, **204**, 43–49.
31. Rasband, W. S. (1997–2007). *ImageJ*, US National Institutes of Health, Bethesda, MD, USA.
32. Jiang, N., Tan, N. S., Ho, B. & Ding, J. L. (2007). Respiratory protein-generated reactive oxygen species as an antimicrobial strategy. *Nat. Immunol.* **8**, 1114–1122.
33. Wang, J., Tan, N. S., Ho, B. & Ding, J. L. (2002). Modular arrangement and secretion of a multidomain serine protease. Evidence for involvement of proline-rich region and N-glycans in the secretion pathway. *J. Biol. Chem.* **277**, 36363–36372.
34. Kawabata, S., Tokunaga, F., Kugi, Y., Motoyama, S., Miura, Y., Hirata, M. & Iwanaga, S. (1996). Limulus factor D, a 43-kDa protein isolated from horseshoe crab hemocytes, is a serine protease homologue with antimicrobial activity. *FEBS Lett.* **398**, 146–150.
35. Lehtinen, M. J., Meri, S. & Jokiranta, T. S. (2004). Interdomain contact regions and angles between adjacent short consensus repeat domains. *J. Mol. Biol.* **344**, 1385–1396.
36. Muta, T., Miyata, T., Misumi, Y., Tokunaga, F., Nakamura, T., Toh, Y. *et al.* (1991). Limulus factor C. An endotoxin-sensitive serine protease zymogen with a mosaic structure of complement-like, epidermal growth factor-like, and lectin-like domains. *J. Biol. Chem.* **266**, 6554–6561.
37. Norman, D. G., Barlow, P. N., Baron, M., Day, A. J., Sim, R. B. & Campbell, I. D. (1991). Three-dimensional structure of a complement control protein module in solution. *J. Mol. Biol.* **219**, 717–725.
38. Schiestl, R. H. & Gietz, R. D. (1989). High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr. Genet.* **16**, 339–346.

APPENDIX



Genemed Synthesis, Inc.
6203 Wood Lake Center Dr., Bldg. 2, San Antonio, TX 78244, USA
Toll free (800) 344-5337; Phone: (210) 745-5988; Fax (210) 745-5992
Email: info@genemedsyn.com, Website: www.genemedsyn.com

PEPTIDE DATA SHEET

LOT NO: 72886

NAME: hTect6

SEQUENCE: H-Leu-Ser-Leu-Ser-Cys-Cys-Glu-Ser-Arg-Lys-Val-Gln-Gly-Arg-Pro-Ser-Pro-Gln-Ala-Ile-OH

EXPECTED MASS: 2159.52

AVERAGE MASS: 720.70[M+3H+]3+


SOULUBILITY: Not tested

AMOUNT DELIVERED: 10.8mg 5.2mg 5.6mg

PURITY: 95.29%

COMMENTS: White to off white powder

DATE: 1/8/09

VERIFIED BY: 



Genemed Synthesis, Inc.
6203 Wood Lake Center Dr., Bldg. 2, San Antonio, TX 78244, USA
Toll free (800) 344-5337; Phone: (210) 745-5988; Fax (210) 745-5992
Email: info@genemedsyn.com, Website: www.genemedsyn.com

Received
9/3/09

PEPTIDE DATA SHEET

LOT NO: 73306

NAME: hTect1

SEQUENCE: H-Gln-Tyr-Trp-Glu-Met-Cys-Lys-Asp-Ser-Gln
Leu-Glu-Phe-Lys-Arg-Val-Ser-Ala-Thr-Thr-OH

EXPECTED MASS: 2450.79

AVERAGE MASS: 818.13[M+3H+]3+

SOULUBILITY: Not tested 5.8mg 2mM
5.2mg 1.18mL

AMOUNT DELIVERED: 11mg 1.06mL

PURITY: 96.64%

COMMENTS: White to off white powder

DATE: 3/6/09

VERIFIED BY: _____



Received
9/3/09

PEPTIDE DATA SHEET

LOT NO: 73304

NAME: GBP34

SEQUENCE: H-Asn-Gly-Glu-Trp-Glu-Leu-Val-Asp-Gly-Ser-
Leu-Lys-Gln-Val-Asp-Gly-Gly-Arg-Asp-Leu-OH

EXPECTED MASS: 2187.37

AVERAGE MASS: 730.30[M+3H+]³⁺

SOULUBILITY: Not tested

AMOUNT DELIVERED: 10mg

PURITY: 95.16%

COMMENTS: White to off white powder

DATE: 3/6/09

VERIFIED BY: _____

5mg 2mM
1.14uL
5mg 1.14uL



Genemed Synthesis, Inc.
6203 Wood Lake Center Dr., Bldg. 2, San Antonio, TX 78244, USA
Toll free (800) 344-5337; Phone: (210) 745-5988; Fax (210) 745-5992
Email: info@genemedsyn.com, Website: www.genemedsyn.com

PEPTIDE DATA SHEET

LOT NO: 72887

NAME: hTect11

SEQUENCE: H-Ile-Gly-Gly-Gly-Trp-Asp-His-Ile-Ser-Val-Arg-Ala-Asn-Ala-Thr-Arg-Ala-Pro-Arg-Ser-OH

EXPECTED MASS: 2121.36

AVERAGE MASS: 708.15[M+3H+]3+

SOULUBILITY: Not tested

AMOUNT DELIVERED: 10.4mg

PURITY: 97.95%
5.4mg
5.0mg

COMMENTS: White to off white powder

DATE: 1/8/09

VERIFIED BY: 