STRUCTURAL AND FUNCTIONAL STUDIES OF INTERACTIONS BETWEEN β-1,3-GLUCAN AND THE N-TERMINAL DOMAINS OF β-1,3-GLUCAN RECOGNITION PROTEINS INVOLVED IN INSECT INNATE IMMUNITY

by

HUAIEN DAI

M.S., Kansas State University, 2006

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Biochemistry Graduate Group

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Abstract

Insect β-1,3-glucan recognition protein (βGRP), a soluble receptor in the hemolymph, binds to the surfaces of bacteria and fungi and activates serine protease cascades that promote destruction of pathogens by means of melanization or expression of antimicrobial peptides. Delineation of mechanistic details of these processes may help develop strategies to control insect-borne diseases and economic losses. Multi-dimensional nuclear magnetic resonance (NMR) techniques were employed to solve the solution structure of the Indian meal moth (*Plodia interpunctella*) βGRP N-terminal domain (N-βGRP), which is sufficient to activate the prophenoloxidase (proPO) pathway resulting in melanin formation. This is the first determined three-dimensional structure of N-βGRP, which adopts an immunoglobulin fold. Addition of laminarin, a β-1,3 and β-1,6 link-containing glucose polysaccharide (~6 kDa) that activates the proPO pathway, to N-βGRP results in the loss of NMR cross-peaks from the backbone ^{15}N-^{1}H groups of the protein, suggesting the formation of a large complex. Analytical ultracentrifugation (AUC) studies of formation of the N-βGRP:laminarin complex show that ligand binding induces self-association of the protein-carbohydrate complex into a macro structure, likely containing six protein and three laminarin molecules (~102 kDa). The macro complex is quite stable, as it does not undergo dissociation upon dilution to submicromolar concentrations. The structural model thus derived from this study for the N-βGRP:laminarin complex in solution differs from the one in which a single N-βGRP molecule has been proposed to bind to a triple-helical form of laminarin on the basis of a X-ray crystal structure of the N-βGRP:laminarihexaose complex. AUC studies and phenoloxidase activation measurements made with designed mutants of N-βGRP indicate that electrostatic interactions between the ligand-bound protein molecules contribute to the stability of the N-βGRP:laminarin complex and that a decreased stability results in a reduction of proPO activation. These novel findings suggest that ligand-induced self-association of the βGRP:β-1,3-glucan complex may form a platform on a microbial surface for recruitment of downstream proteases, as a means of amplification of the pathogen recognition signal. In the case of the homolog of GNBPA2 from *Anopheles gambiae*, the malaria-causing *Plasmodium* carrier, multiligand specificity was characterized, suggesting a functional diversity of the immunoglobulin domain structure.

STRUCTURAL AND FUNCTIONAL STUDIES OF INTERACTIONS BETWEEN β-1,3-GLUCAN AND THE N-TERMINAL DOMAINS OF β-1,3-GLUCAN RECOGNITION PROTEINS INVOLVED IN INSECT INNATE IMMUNITY

by

HUAIEN DAI

M.S., Kansas State University, 2006

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Biochemistry Graduate Group

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2013

Approved by:

Major Professor
Ramaswamy Krishnamoorthi

Abstract

Insect β-1,3-glucan recognition protein (βGRP), a soluble receptor in the hemolymph, binds to the surfaces of bacteria and fungi and activates serine protease cascades that promote destruction of pathogens by means of melanization or expression of antimicrobial peptides. Delineation of mechanistic details of these processes may help develop strategies to control insect-borne diseases and economic losses. Multi-dimensional nuclear magnetic resonance (NMR) techniques were employed to solve the solution structure of the Indian meal moth (*Plodia interpunctella*) βGRP N-terminal domain (N-βGRP), which is sufficient to activate the prophenoloxidase (proPO) pathway resulting in melanin formation. This is the first determined three-dimensional structure of N-βGRP, which adopts an immunoglobulin fold. Addition of laminarin, a β-1,3 and β-1,6 link-containing glucose polysaccharide (~6 kDa) that activates the proPO pathway, to N-βGRP results in the loss of NMR cross-peaks from the backbone ^{15}N-^{1}H groups of the protein, suggesting the formation of a large complex. Analytical ultracentrifugation (AUC) studies of formation of the N-βGRP:laminarin complex show that ligand binding induces self-association of the protein-carbohydrate complex into a macro structure, likely containing six protein and three laminarin molecules (~102 kDa). The macro complex is quite stable, as it does not undergo dissociation upon dilution to submicromolar concentrations. The structural model thus derived from this study for the N-βGRP:laminarin complex in solution differs from the one in which a single N-βGRP molecule has been proposed to bind to a triple-helical form of laminarin on the basis of a X-ray crystal structure of the N-βGRP:laminarihexaose complex. AUC studies and phenoloxidase activation measurements made with designed mutants of N-βGRP indicate that electrostatic interactions between the ligand-bound protein molecules contribute to the stability of the N-βGRP:laminarin complex and that a decreased stability results in a reduction of proPO activation. These novel findings suggest that ligand-induced self-association of the βGRP:β-1,3-glucan complex may form a platform on a microbial surface for recruitment of downstream proteases, as a means of amplification of the pathogen recognition signal. In the case of the homolog of GNBPA2 from *Anopheles gambiae*, the malaria-causing *Plasmodium* carrier, multiligand specificity was characterized, suggesting a functional diversity of the immunoglobulin domain structure.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMP	antimicrobial peptide
AUC	analytical ultracentrifugation
βGRP	β-1,3-glucan recognition protein
BMRB	Biological Magnetic Resonance Bank
DAP	diaminopimelic acid
GNBP	Gram-negative bacteria-binding protein
HP	hemolymph proteinase
HSQC	heteronuclear single quantum correlation
ITC	isothermal titration calorimetry
kDa	kilodaltons
LPS	lipopolysaccharide
LTA	lipotechoic acid
MALDI	matrix assisted laser desorption ionization
MBP	microbe binding protein
MS	mass spectrometry
MSP	modular serine protease
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect spectroscopy
PAGE	polyacrylamide gel electrophoresis
PAMP	pathogen-associated molecular pattern
PDB	Protein Data Bank
PGRP	peptidoglycan recognition protein
proPO	prophenoloxidase
PRR	pattern recognition receptor
RT-PCR	reverse transcriptase polymerase chain reaction
SDS	sodium dodecyl sulfate
s^*	apparent sedimentation coefficient
$S_{20,w}$	sedimentation coefficient corrected to standard state of water at 20°C

Acknowledgements

I thank Dr. Ramaswamy Krishnamoorthi, my advisor, for his encouragement, accommodating support, willingness to help, and his instructions. This work would not have been possible without his guidance to combine a spectrum of biochemical and biophysical techniques.

I thank Dr. Michael Kanost for his insights and support. His earnest and open-minded approach to science has led me to broaden my knowledge in different areas of biochemistry, and to an area of challenge and discovery. In particular, I would like to express my gratitude to him for providing me with an opportunity to work with a team of wonderful researchers.

I thank my committee members, Dr. Ruth Welti and Dr. Karl Kramer, for their valuable comments and suggestions, and for their great patience.

I thank Dr. David VanderVelde and Dr. Asokan Anbanandam at the University of Kansas for their help with NMR experiments. I thank Dr. Michal Zolkiewski for teaching me biophysical techniques. I thank Dr. Yasuaki Hiromasa for his generous contribution to this work with his AUC expertise, and for his enthusiasm for research.

I thank the former and current members of the Kanost lab: Dr. Maureen Gorman, who taught me much of molecular biology, Dr. Neal Dittmer, Dr. Emily Ragan, Dr. Chunju An, Dr. Changzhong Sheng, Dr. Jayne Christen, Dr. Jeff Fabrick, Dr. Daisuke Takahashi, Lucinda Sullivan, Sandi Yungeberg, Stewart Gardner, Lisa Brummett, Zeyu (Sara) Peng and Zhen Li.

I thank all the faculty, staff and graduate students of the Department of Biochemistry and Molecular Biophysics for their assistance. My thanks are to: Dr. Gerald Reeck, Dr. Om Prakash, Dr. Lawrence Davis, Dr. Subbaratnam Muthukrishnan, Yuxi Gong, Crystal Sapp, Dave Manning, Sue-Yi Huang, Melinda Bainter and Rebecca Darkow de Rodriguez.

Finally, I would like to thank my friends and family. I thank Alvaro Herrera, Feng Cui, Rohit Kamat, Matthew Heerman, Ting Zhang, Xiangming Li, Xin Zhu, Yi Yang and Ying Ding, for their help and friendship. I am deeply indebted to my family, which has given me the greatest support during this long journey to accomplish my academic goal.

Chapter 1 - Introduction

Studies on insect innate immunity have important relevance to human health. An understanding of how insects respond to pathogen invasion may help develop strategies for insect control with few or no ill effects on humans or the environment. That would be useful in dealing with agricultural pests and human disease-causing insects. Furthermore, advances in understanding insect immunity have paved the way to obtain insights into human immune responses (Janeway & Medzhitov, 2002). The past two decades have witnessed rapid developments and breakthrough discoveries regarding insect innate immune responses, such as the work on the *Drosophila melanogaster* serine protease cascade leading to the Toll pathway (Lemaitre et al., 1996; Lemaitre & Hoffmann, 2007).

Unlike vertebrates that have both innate and adaptive immune systems, insects as well as other arthropods have only innate immunity. The innate immune system is not only the precursor to the adaptive immune system in vertebrates, but also acts as a regulator of adaptive immunity (Iwasaki & Medzhitov, 2010). It forms the first line of defense against pathogens invading the host, and innate immune responses are immediate, potent and non-specific compared to the adaptive ones. Although they have no adaptive immune system, insects produce antimicrobial peptides within a few hours after infctions (AMPs; Haine et al., 2008). The immediately acting responses, including melanization and engulfment, are highly-efficient to clear most of invading pathogens, before the AMP response is mounted (Schneider & Chambers, 2008). Delineation of activation pathways of such immune responses may help us uncover mechanistic details of antimicrobial responses.

Hosts detect invading pathogens through their pathogen recognition receptors (PRRs), which sense the pathogen-associated molecular patterns (PAMPs) and initiate appropriate immune responses (Janeway & Medzhitov, 2002). PRRs are limited in number as compared to the vast range of pathogens, and a set of PRRs can recognize some particular PAMPs. Carbohydrate moieties are major components of microbial cell walls. These conserved carbohydrate PAMPs include lipopolysaccharides (LPS), peptidoglycans, lipoteichoic acids (LTA) and β-1,3-glucans. Non-carbohydrate PAMPs include proteins and nucleotides that are unique to pathogens, such as bacterial flagellin and virus nucleotides (Akira et al., 2006). In

mammals, Toll-like receptors are the best-studied family of PRRs (Beutler, 2009). In insects, a unique reservoir of PRRs occurs (Yu et al., 2002; Ferrandon et al., 2007): Two best-studied families are peptidoglycan recognition proteins (PGRPs) and β-1,3-glucan recognition proteins (βGRPs), also known as Gram-negative bacteria-binding proteins (GNBPs). How these PRRs recognize their cognate PAMPs remains largely unknown, and the subsequent activation mechanisms are poorly understood.

Carbohydrate PAMPs

Carbohydrates and their derivatives are the major surface components of bacterial and fungal cell walls. These molecules fulfill the criterion for PAMPs, as being constantly expressed and conserved only in microorganisms, but not in hosts. Carbohydrate PAMPs include peptidoglycan, β-1,3-glucan, LTA and LPS.

Peptidoglycans

In Gram-positive bacteria, peptidoglycan forms the outmost layer and is the largest component of the cell wall. In Gram-negative bacteria, an outermost membrane containing LPS covers the peptidoglycan layer. Peptidoglycan is a macromolecule formed by linear carbohydrate chains and cross-linking bridges. The linear carbohydrate chain is a repeat of disaccharide made of β-1,4 N-acetylglucosamine and N-acetylmuramic acid. The cross-linking bridges are peptides usually containing alternative L- and D- amino acids. While the linear chain of disaccharide repeat is conserved in all bacteria, the cross-linking bridges vary for different types (Vollmer & Seligman, 2010). In most Gram-positive bacteria, the cross-linking peptide between the carbohydrate chains has an L-lysine at the third residue (Lys-type peptidoglycan), whereas in most Gram-negative bacteria, the third residue is a meso-diaminopimelic acid (DAP). Insect PRRs for peptidoglycans include PGRPs and/or a set of βGRP/GNBP family members (Ferrandon et al., 2007; Jiang et al., 2010).

β-1,3-Glucans

β-1,3-glucan and chitin are the two major cell wall polysaccharides of fungi (Gow et al., 2012). β-1,3-glucan is a polysaccharide of D-glucose units linked by β-1,3 glycosidic bonds (Figure 1-1). The core structure of fungal cell wall is made by β-1,3-glucan covalently linked to β-1,6-glucan and chitin (poly-β-1,4-N-acetylglucosamine). Despite the simplicity of the D-

glucose unit, β-1,3-glucan can assemble into high-molecular weight complex and therefore complicate the molecular patterns. Furthermore, a linear β-1,3-glucan can undergo branching through β-1,4 or β-1,6 glycosidic bonds. Glycosidic linkages also introduce heterogeneity through different cross-linking patterns in polysaccharides, such as chitin and mannan (poly-β-1,4-mannose) on fungal surface. The structural characteristics of β-1,3-glucan vary from one species to another because of different kinds of modifications that occur in different species.

Three categories of β-1,3-glucan are widely used as experimental immune response elicitors, namely zymosan, curdlan, and laminarin. Zymosan, a particle preparation from yeast cell wall, is rich in β-1,3-glucan and mannan. Curdlan is an insoluble, linear β-1,3-glucan first isolated from a Gram-negative bacterium *Alcaligenes faecalis*. Laminarin, prepared from brown algae, is a soluble β-1,3-glucan with branches linked by β-1,6 glycosidic bonds. The number of branches in laminarin varies depending upon its source. For instance, laminarin isolated from *Laminaria digitata* has a 1,3:1,6 linkage ratio of ~7:1 (Hrmova & Fincher, 1993), while laminarin from *Eisenia bicyclis* has a ratio close to ~2:1 (Handa & Nisizawa, 1961). The cognate insect PRRs for β-1,3-glucan include βGRP/GNBP family members, immulectin family members (Yu et al., 2002), and/or PGRPs, as demonstrated in PGRP-1 and PGRP-2 from the beetle *Holotrichia diomphalia* (Lee et al., 2004).

Lipoteichoic Acids

LTA is a glycophosphate polymer linked to membrane glycolipids. LTA is the second major component of the cell wall of Gram-positive bacteria after peptidoglycan. One characteristic of LTA is amphiphilicity of its glycophosphate polymer caused by the negatively charged phosphate group and the positively charged amino group. LTAs from different Gram-positive bacteria are highly diverse (Weidenmaier & Peschel, 2008). LTA polymer is composed of different repeating units. In *Staphylococcus*, *Bacillus* and some streptococci, the repeating unit is a glycerophosphate chain modified with ester-linked D-alanine and α-D-acetylglucosamine. In *Streptococcus pneumoniae*, the repeating unit is a complex oligosaccharide combination. Insect PRRs for LTA include immulectin (a C-type lectin) family members (Yu et al., 2002) and hemolin, an immunoglobulin superfamily member (Yu & Kanost, 2002).

Lipopolysaccharide

LPS is a polysaccharide attached to a lipid called lipid A. Lipid A anchors LPS into the outer membrane of Gram-negative bacteria. The polysaccharide moiety is composed of a core region and an outermost region called O antigen. Compared with lipid A and the core region, the O antigen is highly variable (Raetz & Whitfield, 2002). The O antigen is a repeat of an oligosaccharide, with a large diversity of species sources, sugar unit number, linkage position, stereochemistry and modification of different monosaccharides. Insect PRRs for LPS include immulectin family members (Yu et al., 2002), hemolin (Yu & Kanost, 2002) and leureptin, a leucine-rich repeat protein (Zhu et al., 2010).

Insect PRR Categories

Immulectins are members of C-type lectin superfamily that contains carbohydrate-binding C-type lectin domains. Unlike most mammalian C-type lectin receptors such as Dectin-1 and DC-SIGN that have only one C-type lectin domain, immulectins have several such domains (Hardison & Brown, 2012; Yu et al., 2002). Immulectins are multi-ligand carbohydrate receptors that can recognize LPS as well as LTA and β-1,3-glucan, and the binding specificities of immulectins are under further investigation (Yu et al., 1999, 2005, 2006; Ling & Yu, 2006; Yu & Kanost, 2000).

A 9 kDa protein isolated from *Galleria mellonella* was characterized as a novel multi-ligand carbohydrate receptor (Kim et al., 2010). This protein was named GmCP8 because it is homologous to the cationic protein 8 (CP8) from *Manduca sexta* (Ling et al., 2009). CP8 is a cysteine-rich protein and its three-dimensional structure is yet to be determined.. MsCP8 was suggested to be a modulating protein for prophenoloxidase (proPO) activation (Ling et al., 2009). The discovery that GmCP8 was able to specifically bind to LPS, LTA and β-1,3-glucan would establish a new category of PRR if further research could confirm these findings.

Leureptin is an extracellular leucine-rich repeat protein found in the *M. sexta* hemolymph (Zhu et al., 2003). Sequence analysis suggests that it has 13 leucine-rich repeats and its homologs are distributed in *Lepidoptera*. Leureptin can bind LPS, and is involved in hemocyte immune responses (Zhu et al., 2010).

Hemolin, an immunoglobulin superfamily member, can recognize Gram-positive and Gram-negative bacteria by sensing LPS and LTA in the cell wall (Yu & Kanost, 2002). It was first found in *Lepidoptera* (Sun et al., 1990; Ladendorff & Kanost, 1991). The structure of hemolin from *Hyalophora cecropia* shows a horseshoe shape made by tandem immunoglobulin domains, and suggests that hemolin is involved in homophilic oligomerization (Su et al., 1998). Since this structure is an apo form, the proposed carbohydrate-binding site needs to be confirmed in order to elucidate the molecular basis for carbohydrate-protein interactions. Dscam, also an immunoglobulin superfamily member and initially found to be essential for nerve development, is a hypervariable PRR for a broad range of pathogens (Dong et al., 2006). Interestingly, the structure of Dscam (ecto-domains D1-D4) from *Drosophila melanogaster* shows a horseshoe shape similar to that of hemolin (Meijers et al., 2007). While the ligand-binding site of Dscam has not been explored, the homophilic binding site for immunoglobulin domains has been mapped out, which highlights the adhesion property of immunoglobulin superfamily members. Both hemolin and Dscam are involved in cellular immune responses (Schmidt et al., 2010).

PGRPs and βGRPs are the two best-known insect PRRs (Table 1-1), both first identified in *Lepidoptera* (Yoshida et al., 1996; Lee et al., 1996; Ochiai & Ashida, 2000; Ma & Kanost, 2000). Since then conclusive data have emerged showing that PGRPs and βGRPs can sense bacterial and fungal cell wall components (Ferrandon et al., 2007).

Peptidoglycan Recognition Proteins

Conserved from insects to human, PGRPs act as PRRs only in invertebrates (Royet & Dziarski, 2007). The first characterized PGRP was isolated from the hemolymph of *Bombyx mori* in the course of a study on melanization (Yoshida et al., 1986; 1996; Ochiai & Ashida, 1999). PGRPs have high affinities for peptidoglycans, in the presence of which they can trigger proPO activation. Later it was found that in *Drosophila*, a PGRP member (PGRP-SA) was involved in the Toll pathway activation (Michel et al., 2001). In human, the four PGRPs (PGLYRP1-4) catalyze peptidoglycan hydrolysis and hence are effectors against pathogens, besides being sensors (Kang et al., 1998; Liu et al., 2001). These findings were followed by extensive three-dimensional structure determinations of PGRP members from *Drosophila* and human (Dziarski & Gupta D, 2006). All PGRPs have a conserved carboxyl-terminal domain homologous to the bacteriophage T7 lysozyme, adopting a fold of a central five-stranded β-sheet

surrounded by three α-helices. The *D. melanogaster* genome has 15 PGRP family members and they can be grouped into short (S) and long (L) subfamilies based on protein length (Royet et al., 2011). The nine L members (PGRP-L) are generally DAP-type peptidoglycan receptors, and among the other six S members (PGRP-S) four are amidases, and two (PGRP-SA and PGRP-SD) are Lys-type peptidoglycan receptors.

PGRP-LB from *D. melanogaster* is the first member of the PGRP family whose three-dimensional structure was determined (Kim et al., 2003). Analysis of this structure led to the proposal of an oligomerization-induced proximity mechanism. This was later confirmed by the structure of PGRP-LE, which revealed DAP-type peptidoglycan-induced PGRP polymerization (Lim et al., 2006). Polymerization of PGRP-LE upon recognition of DAP-type peptidoglycan triggers subsequent molecular events by providing proximity contacts for signaling components. Structural studies of PGRP-LCa and -LCx revealed the specificity determinant for DAP-type peptidoglycan recognition (Chang et al., 2006). The DAP-type peptidoglycan-binding site is a long groove on the protein surface. Specificity arises from the electrostatic charge from an Arg side chain, the orientation of two aromatic residues, and extensive hydrogen bonds.

PGRP-SA and -SD recognize Lys-type peptidoglycans (Michel et al., 2001; Bischoff et al., 2004). The structure of *D. melanogaster* PGRP-SA reveals a Lys-type peptidoglycan-docking groove (Chang et al., 2004). The molecular details of PGRP-SA activation of Toll and melanization pathways are yet to be elucidated. For *D. melanogaster*, a pair of PRRs, PGRP-SA and GNBP1, collaborate to recognize Lys-type peptidoglycans and initiate signaling events (Gobert et al., 2003; Pili-Floury et al., 2004). A similar scenario utilizing a pair of PGRP-SA and GNBP1 to trigger immune responses to Lys-type peptidoglycans has been reported for the mealworm beetle *Tenebrio molitor* (Park et al., 2007; Kim et al., 2008). The structure of *D. melanogaster* PGRP-SD, in contrast, suggests a binding preference for DAP-type peptidoglycans (Leone et al., 2008). This is consistent with PGRP-SD's versatility for receptor complex formation (Wang et al., 2008). Details regarding such receptor complex formation (PGRP-SA and/or PGRP-SD plus GNBP1) remain unclear. The afore-mentioned PGRP structures indicate that PGRPs recognize only the peptide moiety of a peptidoglycan, thus offering a clue to understanding the specificity conferred on this PRR family. The requirement of GNBP1 for Lys-type peptidoglycan-induced immune responses suggests a common signaling strategy utilized by both PGRPs and βGRPs/GNBPs.

In *M. sexta*, PGRP1 was characterized recently (Sumathipala & Jiang, 2010). PGRP1 can bind to both DAP-type and Lys-type peptidoglycans, and can stimulate proPO activation. Sequence analysis indicates that it is orthologous to the first characterized PGRP (Yoshida et al., 1986; 1996; Ochiai & Ashida, 1999). Its binding specificity, however, could not be explained on the basis of known structures of *D. melanogaster* PGRP members.

Five *D. melanogaster* PGRPs demonstrate amidase activity (Royet et al., 2011). Among the 12 PGRPs annotated in the *B. mori* genome, four are possible amidases (Tanaka et al., 2008). Based on their functions, PGRPs can be categorized into three groups: DAP-type peptidoglycan receptors, components of the Lys-type peptidoglycan receptor complex, and effector proteins.

β-1,3-Glucan Recognition Proteins

βGRPs were first found in Lepidoptera insects (Ochiai & Ashida, 1986; 1988; 2000; Lee et al., 1996; Ma & Kanost. 2000) and many members have since been identified (Kim et al., 2000; Fabrick et al. 2003; Zhang et al., 2003; Jiang et al., 2004; Wang et al., 2006; Warr et al., 2008; Zheng & Xia, 2012; Wang et al., 2013). βGRPs and GNBPs are highly homologous, and therefore grouped to one family, the GNBP/βGRP family (Royet, 2004). All GNBP/βGRP family members share a conserved structure: a carbohydrate-recognition domain in their N-terminal region and a β-1,3-glucanase-like domain in their C-terminal region (Figure 1-1). The N-terminal carbohydrate recognition domain corresponds to a carbohydrate-binding module (CBM) that is assigned to family 39 (CBM39) based on its primary structure (Boraston et al. 2004). The C-terminal glucanase-like domain has no glucanase activity because the two catalytic Glu residues are replaced with non-charged residues. Sequence analysis indicates that the glucanase-like domain adopts a β-jelly roll fold and belongs to glycoside hydrolase family 16 (GH16) according to the CAZy database (Cantarel et al., 2009).

The carbohydrate-binding specificities of GNBP/βGRP family members are not well defined, despite the fact that the proteins are grouped together. For instance, the *D. melanogaster* genome has three GNBP/βGRP members: GNBP1, GNBP2 and GNBP3. The function of GNBP2 is not clear. GNBP1 participates in Lys-type peptidoglycan-induced immune responses, either as part of PGRP-SA-GNBP1 signaling complex (Gobert et al., 2003; Pili-Floury et al., 2004) or as an enzyme digesting Lys-type peptidoglycan for subsequent recognition by PGRP-SA (Wang et al., 2005). GNBP3 has been shown to detect the fungi cell wall through recognition

of β-1,3-glucan (Gottar et al., 2006). In *M. sexta*, two βGRPs have been found, namely βGRP1 and βGRP2 (Ma & Kanost, 2000; Jiang et al., 2004). Both can tightly bind curdlan. βGRP1 is expressed constitutively, while βGRP2 is an acute-phase protein. βGRP2 can aggregate fungi and both Gram-positive and Gram-negative bacteria. βGRP from the mosquito *Armigeres subalbatus* also binds different species of bacteria, irrespective of the Gram-type (Wang et al. 2006).

Four βGRPs, namely βGRP1-4, have been annotated in the *B. mori* genome (Tanaka et al., 2008). Unlike the afore-mentioned GNBP/βGRP family members, BmβGRP4 does not have the N-terminal CBM39 domain, and the catalytic Glu residues are retained. In the genome of the malaria vector, *Anopheles gambiae*, six GNBP/βGRP genes have been annotated, namely GNBPA1, GNBPA2, and GNBPB1-4 (Warr et al., 2008). The four B subgroup members (GNBPB1 to B4) have no CBM39 domain and retain glucanase catalytic residues. Such CBM39-free βGRPs have been found in *Helicoverpa armigera* (Pauchet et al., 2009), *Spodoptera frugiperda* (Bragatto et al., 2010) and termites (Bulmer et al., 2009), but not in the *Drosophila* genome. Discovery of glucanase members of GNBP/βGRP family has prompted a reassessment of this PRR family based on GNBP/βGRP/glucanase categories (Hughes, 2012).

Recently, a new β-1,3-glucanase-related protein named microbe binding protein (MBP) was identified in *M. sexta* hemolymph. MBP does not bind curdlan or laminarin, but prefers peptidoglycans and LTA (Wang et al., 2011). Sequence analysis indicates that MsMBP is more similar to BmβGRP2 than to MsβGRP1 and MsβGRP2. BmβGRP2 was originally named GNBP (Lee et al., 1996) and is more similar to DmGNBP1 than to DmGNBP3. Because DmGNBP1 is a component for PGRP-SA (-SD)-mediated Lys-type peptidoglycan recognition (Gobert et al., 2003; Pili-Floury et al., 2004) and DmGNBP3 is a β-1,3-glucan receptor, it is plausible that this group of GNBP/βGRP members (such as BmβGRP2, DmGNBP1, and MsMBP) has lost β-1,3-glucan-binding activity. Unlike their longer homologs (PGRP-Ls), PGRP-SA and PGRP-SD do not have a signaling domain and, therefore, GNBP1 might provide a link between Lys-type peptidoglycan recognition and immune signaling initiation. Such a scenario also applies to *T. molitor* (Park et al., 2007; Kim et al., 2008; Kan et al., 2008).

The prototypes of βGRPs (BmβGRP1 and MsβGRP1) are bona fide β-1,3-glucan receptors (Ochiai & Ashida, 2000; Ma & Kanost, 2000). These also include DmGNBP3 (Gottar

et al., 2006), PiβGRP from *Plodia interpunctella* (Fabrick et al., 2004), TmGNBP3 (renamed from TmGRP) from *Tenebrio molitor* (Zhang et al., 2003), MsβGRP2 (Jiang et al., 2004), and AsGRP from *A. subalbatus* (Wang et al., 2006). The N-terminal CBM39 domain alone can bind to β-1,3-glucan (Ochiai & Ashida, 2000; Fabrick et al., 2004; Lee et al., 2009). Furthermore, the N-terminal CBM39 domain can stimulate proPO in the presence of laminarin, thus suggesting a signaling role for CBM39 from *P. interpunctella* (Fabrick et al., 2004).

High Molecular Weight Complexes in Hemolymph

Formation of high molecular complexes against pathogens would localize immune responses on the intruders. Such complexes provide a platform for interactions between PRR and PAMP, between different PRRs, and between PRR and its corresponding signaling molecule, triggering downstream signaling events and recruiting effector proteins. In *M. sexta*, the hemolymph complexes containing βGRP, signaling molecules and effectors have been characterized by proteomics (Ragan et al., 2010; Christen et al., 2012).

PGRPs and GNBPs/βGRPs are components of a secreted recognition complex. PGRP-SA-GNBP1 complex is capable of Lys-type peptidoglycan recognition (Gobert et al., 2003; Pili-Floury et al., 2004; Park et al., 2007; Kim et al., 2008; Kan et al., 2008). The downstream factor, a modular serine protease, has been identified (Buchon et al., 2009). Components associated with the recognition complex include effector proteins. For instance, DmGNBP3 can assemble an effector complex by bringing proPO to the pathogen surface (Matskevich et al., 2010). Knowledge about how these complexes are formed would provide insight into the regulatory mechanism of insect immune responses.

Humoral Immune Responses in Insects

Cellular and humoral responses constitute the innate immune system. In insects, the cellular immune responses are mediated by adhesion molecules, such as hemolin and Dscam (Schmidt et al., 2010). Humoral immune responses include melanization and expression of antimicrobial peptides (Kanost et al., 2004; Lemaitre & Hoffmann, 2007; Jiang et al., 2010). The following summary focuses on the formation of PRR: PAMP complexes that initiate humoral immune responses, and mainly describes the initiation complexes of βGRPs and PGRPs for immune responeses.

Melanization Pathway

Melanization is an immediate immune response in insects and results in melanin deposition on the pathogen surface (Cerenius et al., 2008; 2010; Kanost et al., 2004; Jiang et al., 2010). Immediate immune responses including melanization kill the vast majority bacteria (Haine et al., 2008). Tight regulation of melanization ensures the targeting of pathogens. During melanization, the key enzyme phenoloxidase (PO) catalyzes melanin synthesis. Phenoloxidase exists in hemolymph as an inactive zymogen, proPO. When pathogen invasion occurs, proPO can be activated through a proteolytic cleavage of the prosegment. This reaction is mediated by a set of serine protease cascades, a strategy that has been found in a variety of signaling pathways (Krem & Di Cera, 2002). Two examples of a similar cascade theme are coagulation and complement system. Components of the serine protease cascade leading to proPO activation are being identified. In *M. sexta* and *T. molitor*, this is a three-step cascade, which is initiated by PRR:PAMP complexes such as PGRP-SA-GNBP1:Lys-type peptidoglycan and βGRP:β-1,3-glucan (Gorman et al., 2007; An et al., 2009; Roh et al., 2009). A modular serine protease containing LDLr (Low-density Lipoprotein Receptor) domain and CCP (complement control protein) is activated by formation of these PRR:PAMP complexes (Ji et al., 2004; Wang and Jiang, 2006; 2010).

Antimicrobial Peptide Response

Synthesis of antimicrobial peptides is controlled by Toll and immune deficiency (IMD) pathways (Figure 1-2). These two signaling pathways are similar with their intracellular components, as Toll and IMD pathways both utilize NF-κB homologs, named DIF (dorsal-related immunity factor) and Relish respectively (Lemaitre & Hoffmann, 2007). The transmembrane receptor Toll conveys extracellular signal that is amplified by a three-step cascade to intracellular NF-κB pathway, while in IMD pathway, transmembrane PGRP-Ls oligomerizes upon the recognition of DAP-type peptidoglycan and activates intracellular signaling.

Toll and proPO pathways share components of the serine protease cascade initiated by recognition of Lys-type peptidogylan and β-1,3-glucan via their cognate PRRs (Kim et al., 2008; An et al., 2009; Buchon et al., 2009). It remains a challenge to understand the molecular basis of how PRR:PAMP complexes activate the serine protease cascade. Knowledge of these initiation

complexes (PGRP-SA/GNBP1:Lys-type peptidoglycan and βGRP:β-1,3-glucan) is relatively poor, as compared to that of the IMD pathway (PGRP-Ls-DAP-type peptidoglycan). Structural studies have revealed an array of PGRP-Ls that could convey the extracellular recognition event to intracellular IMD signaling through their intracellular adaptor domains (Lim et al., 2006).

Carbohydrate Binding Modules

Carbohydrate binding module (CBM) is defined as a contiguous amino acid sequence within a carbohydrate-active enzyme with a discrete fold having carbohydrate-binding activity. The carbohydrate binding domain of PGRPs and GNBPs/βGRPs can be summarized as the following three categories: lysozyme-related, glucanase-related and CBM39. All PGRPs contain a conserved domain homologous to the bacteriophage T7 lysozyme. The two domains conserved in GNBPs/βGRPs are CBM39 at the N-terminal region and glucanase-like domain at the C-terminal region. The lysozyme-related and glucanase-related domains are related to an enzymatic activity, and thus not grouped as a CBM in the CAZy database (Cantarel et al., 2009).

Based on their binding specificity, CBMs were grouped into three functional types, type A, type B and type C (Boraston et al., 2004; Hashimoto 2006). Type A CBMs have a planar binding site made by the side chains of aromatic residues. Insoluble polysaccharides such as chitin and cellulose are the target ligands of type A CBMs. Stacking interactions between sugar rings and aromatic rings play a key role in target ligand-binding. Type B CBMs each have a binding groove on their surfaces, and individual glycan chains, rather than surfaces of insoluble polysaccharides, are their binding substrates. Orientations of aromatic residues around the binding groove, together with hydrogen bonds, are the structural determinant of ligand specificity for type B CBMs. Type C CBMs are less distinct than type A or B because of lack of a characteristic binding site. They preferably bind mono-, di- or trisaccharides, and, therefore, are defined as small sugar-binding CBMs.

Methods

Nuclear Magnetic Resonance (NMR) spectroscopy can be utilized for both solution structure determination and ligand-binding assay. A protein-carbohydrate complex formation can be monitored by biochemical and biophysical methods, such as calorimetry, electronic absorption spectroscopy, and analytical ultracentrifugation. Each method has its advantages as

well as limitations. A combination of different methods will provide a comprehensive view of interactions present in a protein-carbohydrate complex.

Nuclear Magnetic Resonance Spectroscopy

NMR spectroscopy has been widely used for structure determination of proteins and nucleic acids since 1980s (Wüthrich, 1989). By means of heteronuclear multidimensional spectroscopy, three-dimensional solution structures of proteins of 20-30 kDa can be determined, provided these molecules are isotopically enriched with ^{13}C and ^{15}N atoms (Cavanagh et al. 1996). NMR techniques are ideally suited not only for protein structure determination, but also for measurement of protein dynamics and identification of ligand-binding site(s). NMR methods have been developed over the years for studying protein-carbohydrate interactions to design new carbohydrate ligands, mapping of ligand epitopes and ligand-binding sites, determining protein-carbohydrate complex structures, and monitoring ligand-induced conformational changes (Kogelberg et al., 2003). NMR spectroscopy has also been applied for structural characterization of carbohydrate PAMPs like a peptidoglycan fragment (Meroueh et al. 2006). Limitations of NMR spectroscopy include requirement of high sample concentration (~0.5 mM), limiting molecular size (<50 kDa), and lack of side chain information in ligand binding-site mapping.

Isothermal Titration Calorimetry

Isothermal titration calorimetry is a powerful biophysical method to study protein-ligand binding (Ladbury & Chowdhry, 1996). It measures heat change when a ligand is added to its binding protein. All thermodynamic parameters associated with the binding process can be derived from heat change recorded periodically during a titration. An advantage of ITC is that ligands of various sizes and conformations can all be studied. Its limitation is that it cannot detect a binding process for which heat change or binding affinity is too small.

Analytical Ultracentrifugation

AUC allows characterization of macromolecular interactions in solution. Sedimentation coefficients determined from a sedimentation velocity experiment can be used to analyze the size and stoichiometry of macromolecular complexes (Lebowitz et al., 2002). New approaches for data analysis and developments of computational software have rendered AUC a handy tool for studying complexes of heterogeneous molecules (Philo, 2000). The derived apparent

sedimentation coefficient distribution function $g(s^*)$ provides important information required for evaluating molecular masses of complexes. A typical modern AUC instrument enables simultaneous monitoring of sedimentations of a protein and a carbohydrate, thus making it possible to determine the stoichiometry of a protein:carbohydrate complex.

Goals of Current Research

The Indian meal moth *P. interpunctella* (Hübner) is one of the major stored grain pests (Mohandass et al., 2007). The larvae of *P. interpunctella* feed on a variety of stored-products and processed food commodities. They are difficult to control because of their small size (9 mm) and because they feed within the grain. Studies on the mechanism of pathogen recognition by *P. interpunctella* may provide a basis for developing strategies to make the insect susceptible to pathogens and thus reduce insect-caused economic losses. Previous research shows that βGRP from *P. interpunctella* binds to the cell walls of bacteria and yeasts with variable affinity in different regions of the protein (Fabrick et al. 2004). Curdlan-binding studies of recombinant full-length *P. interpunctella* βGRP and its N- and C-terminal truncations reveal that the N-terminal domain contains a functionally sufficient β-1,3-glucan recognition domain (Fabrick et al., 2004). We are interested in characterizing the interaction between this domain and its carbohydrate ligands, and the role of the protein:carbohydrate complex in initiating immune responses. To characterize the structural basis of β-1,3-glucan recognition process, we have carried out multi-dimensional NMR studies of the N-terminal domain of βGRP. Studies carried out regarding β-1,3-glucan recognition by βGRP CBM39 domain will form the first necessary step toward delineating the mechanism of pathogen recognition.

The *Plasmodium* vector mosquito *Anopheles gambiae* transmits malaria, an infectious disease that renders nearly half of the world's population at risk according to World Health Organization (WHO: The World Malaria Report 2012). There are six GNBP members in *Anopheles* genome, and silencing of GNBPA2 caused the strongest effect of *Plasmodium* infection compared with other five GNBP members (Warr et al., 2008). Some key residues in the N-terminal domain of GNBPA2 are different from the *Lepidoptera* N-terminal domain of βGRP. We have characterized a multi-ligand specificity for this domain, which suggests a functional diversity of this protein family.

References

Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell* **124**, 783-801.

An, C., Ishibashi, J., Ragan, E. J., Jiang, H., & Kanost, M. R. (2009). Functions of *Manduca sexta* hemolymph proteinases HP6 and HP8 in two innate immune pathways. *J. Biol. Chem.* **284**, 19716-19726.

Beutler, B. A. (2009). TLRs and innate immunity. *Blood* **113**, 1399-1407.

Bischoff, V., Vignal, C., Boneca, I. G., Michel, T., Hoffmann, J. A., & Royet, J. (2004). Function of the *Drosophila* pattern-recognition receptor PGRP-SD in the detection of Gram-positive bacteria. *Nature Immunol*. **5**, 1175–1180.

Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769-781.

Bragatto, I., Genta, F. A., Ribeiro, A. F., Terra, W. R., & Ferreira, C. (2010). Characterization of a β-1,3-glucanase active in the alkaline midgut of *Spodoptera frugiperda* larvae and its relation to β-glucan-binding proteins. *Insect Biochem. Mol. Biol*. **40**, 861-872.

Bulmer, M. S., Bachelet, I., Raman, R., Rosengaus, R. B., Sasisekharan, R. (2009). Targeting an antimicrobial effector function in insect immunity as a pest control strategy. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12652-12657.

Buchon, N., Poidevin, M., Kwon, H. M., Guillou, A., Sottas, V., Lee, B. L., & Lemaitre, B. (2009). A single modular serine protease integrates signals from pattern-recognition receptors upstream of the *Drosophila* Toll pathway. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12442-12447.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*. **37**, D233-D238.

Cavanagh, C., Palmer, A. G. 3rd, & Skelton, N. J. (1996). *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego.

Cerenius, L., Kawabata, S., Lee, B. L., Nonaka, M., & Söderhäll, K. (2010). Proteolytic cascades and their involvement in invertebrate immunity. *Trends Biochem. Sci*. **35**, 575-583.

Cerenius, L., Lee, B. L., & Söderhäll, K. (2008). The proPO-system: pros and cons for its role in invertebrate immunity. *Trends Immunol.* **29**, 263-271.

Chang, C. I., Chelliah, Y., Borek, D., Mengin-Lecreulx, D., & Deisenhofer, J. (2006). Structure of tracheal cytotoxin in complex with a heterodimeric pattern-recognition receptor. *Science* **311**, 1761-1764.

Chang, C. I., Pili-Floury, S., Hervé, M., Parquet, C., Chelliah, Y., Lemaitre, B., Mengin-Lecreulx, D., & Deisenhofer, J. (2004). A *Drosophila* pattern recognition receptor contains a peptidoglycan docking groove and unusual L,D-carboxypeptidase activity. *PLoS Biol*. **2**, e277.

Christen, J. M., Hiromasa, Y., An, C., & Kanost, M. R. (2012). Identification of plasma proteinase complexes with serpin-3 in *Manduca sexta*. *Insect Biochem. Mol. Biol*. **42**, 946-955.

Dziarski, R., & Gupta, D. (2006). The peptidoglycan recognition proteins (PGRPs). *Genome Biol*. **7**, 232.

Dong, Y., Taylor, H. E., & Dimopoulos, G. (2006). AgDscam, a hypervariable immunoglobulin domain-containing receptor of the *Anopheles gambiae* innate immune system. *PLoS Biol*. **4**, e229.

Fabrick, J. A., Baker, J. E., & Kanost, M. R. (2003). cDNA cloning, purification, properties, and function of a beta-1,3-glucan recognition protein from a pyralid moth, *Plodia interpunctella*. *Insect Biochem. Mol. Biol*. **33**, 579-594.

Fabrick, J. A., Baker, J. E., & Kanost, M. R. (2004). Innate immunity in a pyralid moth: functional evaluation of domains from a β-1,3-glucan recognition protein. *J. Biol. Chem.* **279**, 26605-26611.

Ferrandon, D., Imler, J. L., Hetru, C., & Hoffmann, J.A. (2007). The *Drosophila* systemic immune response: sensing and signalling during bacterial and fungal infections. *Nat. Rev. Immunol*. **7**, 862-874.

Gobert, V., Gottar, M., Matskevich, A. A., Rutschmann, S., Royet, J., Belvin, M., Hoffmann, J. A., & Ferrandon, D. (2003). Dual activation of the *Drosophila* toll pathway by two pattern recognition receptors. *Science* **302**, 2126-2130.

González-Santoyo, I., & Córdoba-Aguilar, A. (2012). Phenoloxidase: a key component of the insect immune system. *Entomol. Exp. Appl*. **142**, 1-16.

Gorman, M. J., Wang, Y., Jiang, H., & Kanost, M. R. (2007). *Manduca sexta* hemolymph proteinase 21 activates prophenoloxidase-activating proteinase 3 in an insect innate immune response proteinase cascade. *J. Biol. Chem*. **282**, 11742-11749.

Gottar, M., Gobert, V., Matskevich, A. A., Reichhart, J. M., Wang, C., Butt, T. M., Belvin, M., Hoffmann, J. A., & Ferrandon, D. (2006). Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors. *Cell* **127**, 1425-1437.

Gow, N. A., van de Veerdonk, F. L., Brown, A. J., & Netea, M. G. (2012). *Candida albicans* morphogenesis and host defence: discriminating invasion from colonization. *Nat. Rev. Microbiol*. **10**, 112-122.

Haine, E. R., Moret, Y., Siva-Jothy, M. T., & Rolff, J. (2008). Antimicrobial defense and persistent infection in insects. *Science* **322**, 1257-1259.

Handa, N., & Nisizawa, K. (1961). Structural investigation of a laminaran isolated from *Eisenia bicyclis*. *Nature* **192**, 1078-1080.

Hardison, S. E., & Brown, G. D. (2012). C-type lectin receptors orchestrate antifungal immunity. *Nat. Immunol*. **13**, 817-822.

Hashimoto, H. (2006). Recent structural studies of carbohydrate-binding modules. *Cell Mol. Life Sci.* **63**, 2954-2967.

Hrmova, M., & Fincher, G. B. (1993). Purification and properties of three (1-3)-β-D-glucanase isoenzymes from young leaves of barley (*Hordeum vulgare*). *Biochem. J.* **289**, 453-461.

Hughes, A. L. (2012). Evolution of the βGRP/GNBP/β-1,3-glucanase family of insects. *Immunogenetics* **64**, 549-558.

Iwasaki, A., & Medzhitov, R. (2010). Regulation of adaptive immunity by the innate immune system. *Science* **327**, 291-295.

Janeway, C. A., Jr., & Medzhitov, R. (2002). Innate immune recognition. *Annu. Rev. Immunol.* **20**, 197-216.

Ji, C., Wang, Y., Guo, X., Hartson, S., & Jiang, H. (2004). A pattern recognition serine proteinase triggers the prophenoloxidase activation cascade in the tobacco hornworm, *Manduca sexta*. *J. Biol. Chem.* **279**, 34101-34106.

Jiang, H., Ma, C., Lu, Z. Q., & Kanost, M. R. (2004) β-1,3-glucan recognition protein-2 (βGRP-2) from *Manduca sexta*; an acute-phase protein that binds β-1,3-glucan and lipoteichoic acid to aggregate fungi and bacteria and stimulate prophenoloxidase activation. *Insect Biochem. Mol. Biol.* **34**, 89-100.

Jiang, H., Vilcinskas, A., & Kanost, M. R. (2010). Immunity in lepidopteran insects. In *Invertebrate Immunity*. Söderhäll K (ed) pp. 181-204. Landes Bioscience.

Kan, H., Kim, C.H., Kwon, H. M., Park, J. W., Roh, K. B., Lee, H., Park, B. J., Zhang, R., Zhang, J., Söderhäll, K., Ha, N. C., & Lee, B. L. (2008). Molecular control of phenoloxidase-induced melanin synthesis in an insect. *J. Biol. Chem*. **283**, 25316-25323.

Kanost, M. R., Jiang, H., & Yu, X.Q. (2004). Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunol. Rev.* **198**, 97-105.

Kang, D., Liu, G., Lundström, A., Gelius, E., & Steiner, H. (1998). A peptidoglycan recognition protein in innate immunity conserved from insects to humans. *Proc. Natl. Acad. Sci. U. S. A*. **95**, 10078-10082.

Kim, M. S., Byun, M. & Oh, B. H. (2003). Crystal structure of peptidoglycan recognition protein LB from *Drosophila melanogaster*. *Nature Immunol*. **4**, 787–793.

Kim, C. H., Kim, S. J., Kan, H., Kwon, H. M., Roh, K. B., Jiang, R., Yang, Y., Park, J. W., Lee, H. H., Ha, N. C., Kang, H. J., Nonaka, M., Söderhäll, K., & Lee, B. L. (2008). A three-step proteolytic cascade mediates the activation of the peptidoglycan-induced toll pathway in an insect. *J. Biol. Chem*. **283**, 7599-7607.

Kim, C. H., Shin, Y. P., Noh, M. Y., Jo, Y. H., Han, Y. S., Seong, Y. S., & Lee, I. H. (2010). An insect multiligand recognition protein functions as an opsonin for the phagocytosis of microorganisms. *J. Biol. Chem*. **285**, 25243-25250.

Kim, Y. S., Ryu, J. H., Han, S. J., Choi, K. H., Nam, K. B., Jang, I. H., Lemaitre, B., Brey, P. T., & Lee, W. J. (2000). Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and β-1,3-glucan that mediates the signaling for the induction of innate immune genes in *Drosophila melanogaster* cells. *J. Biol. Chem.* **275**, 32721-32727.

Kogelberg, H., Solís, D., & Jiménez-Barbero, J. (2003). New structural insights into carbohydrate-protein interactions from NMR spectroscopy. *Curr. Opin. Struct. Biol*. **13**, 646-653.

Krem, M. M., & Di Cera, E. (2002). Evolution of enzyme cascades from embryonic development to blood coagulation. *Trends Biochem. Sci*. **27**, 67-74.

Ladbury, J. E., & Chowdhry, B. Z. (1996). Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem. Biol.* **3**, 791-801.

Ladendorff, N. E. & Kanost, M. R. (1991). Bacteria-induced protein P4 (hemolin) from *Manduca sexta*: a member of the immunoglobulin superfamily which can inhibit hemocyte aggregation. *Arch. Insect Biochem. Physiol*. **18**, 285-300.

Lebowitz, J., Lewis, M. S., & Schuck, P. (2002). Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci*. **11**, 2067-2079.

Lee, M. H., Osaki, T., Lee, J. Y., Baek, M. J., Zhang, R., Park, J. W., Kawabata, S., Söderhäll, K., & Lee, B. L. (2004). Peptidoglycan recognition proteins involved in 1,3-β-D-glucan-dependent prophenoloxidase activation system of insect. *J. Biol. Chem*. **279**, 3218-3227.

Lee, W.J., Lee, J. D., Kravchenko, V. V., Ulevitch, R. J., & Brey, P. T. (1996). Purification and molecular cloning of an inducible gram-negative bacteria-binding protein from the silkworm, *Bombyx mori*. *Proc. Natl. Acad. Sci. U. S. A*. **93**, 7888-7893.

Lemaitre, B. & Hoffmann, J. (2007). The host defense of *Drosophila melanogaster*. *Annu. Rev. Immunol*. **25**, 697-743.

Lemaitre, B., Nicolas, E., Michaut, L., Reichhart, J. M., & Hoffmann, J. A. (1996). The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973-983.

Leone, P., Bischoff, V., Kellenberger, C., Hetru, C., Royet, J., & Roussel A. (2008). Crystal structure of *Drosophila* PGRP-SD suggests binding to DAP-type but not lysine-type peptidoglycan. *Mol. Immunol*. **45**, 2521-2530.

Lim, J. H., Kim, M. S., Kim, H. E., Yano, T., Oshima, Y., Aggarwal, K., Goldman, W. E., Silverman, N., Kurata, S., & Oh, B. H. (2006). Structural basis for preferential recognition of diaminopimelic acid-type peptidoglycan by a subset of peptidoglycan recognition proteins. *J. Biol. Chem.* **281**, 8286-8295.

Ling, E., Rao, X. J., Ao, J. Q., & Yu, X. Q. (2009). Purification and characterization of a small cationic protein from the tobacco hornworm *Manduca sexta*. *Insect Biochem. Mol. Biol.* **39**, 263-271.

Ling, E., & Yu, X. Q. (2006). Cellular encapsulation and melanization are enhanced by immulectins, pattern recognition receptors from the tobacco hornworm *Manduca sexta*. *Dev. Comp. Immunol*. **30**, 289-299.

Liu, C., Xu, Z., Gupta, D. & Dziarski, R. (2001). Peptidoglycan recognition proteins: a novel family of four human innate immunity pattern recognition molecules. *J. Biol. Chem*. **276**, 34686–34694.

Ma, C., & Kanost, M. R. (2000). A β1,3-glucan recognition protein from an insect, *Manduca sexta*, agglutinates microorganisms and activates the phenoloxidase cascade. *J. Biol. Chem.* **275**, 7505-7514.

Matskevich, A. A., Quintin, J., & Ferrandon, D. (2010). The *Drosophila* PRR GNBP3 assembles effector complexes involved in antifungal defenses independently of its Toll-pathway activation function. *Eur. J. Immunol.* **40**, 1244-1254.

Meijers, R., Puettmann-Holgado, R., Skiniotis, G., Liu, J. H., Walz, T., Wang, J. H., & Schmucker, D. (2007). Structural basis of Dscam isoform specificity. *Nature* **449**, 487-491.

Meroueh, S. O., Bencze, K. Z., Hesek, D., Lee, M., Fisher, J. F., Stemmler, T. L., & Mobashery, S. (2006). Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 4404-4409.

Michel, T., Reichhart, J. M., Hoffmann, J. A. & Royet, J. (2001). *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. *Nature* **414**, 756–759.

Mohandass, S., Arthur, F. H., Zhu, K. Y., & Throne, J. E. (2007). Biology and management of *Plodia interpunctella* (Lepidoptera: Pyralidae) in stored products. *J. Stored Products Res.* **43**, 302–311.

Ochiai, M., & Ashida, M. (1988). Purification of a β-1,3-glucan recognition protein in the prophenoloxidase activating system from hemolymph of the silkworm, *Bombyx mori*. *J. Biol. Chem*. **263**, 12056-12062.

Ochiai, M., & Ashida, M. (1999). A pattern recognition protein for peptidoglycan. Cloning the cDNA and the gene of the silkworm, *Bombyx mori*. *J. Biol. Chem.* **274**, 11854-11858.

Ochiai, M., & Ashida, M. (2000). A pattern-recognition protein for β-1,3-glucan. The binding domain and the cDNA cloning of β-1,3-glucan recognition protein from the silkworm, *Bombyx mori*. *J. Biol. Chem.* **275**, 4995-5002.

Park, J. W., Kim, C. H., Kim, J. H., Je, B. R., Roh, K. B., Kim, S. J., Lee, H. H., Ryu, J. H., Lim, J. H., Oh, B. H., Lee, W. J., Ha, N. C., & Lee, B. L. (2007). Clustering of peptidoglycan

recognition protein-SA is required for sensing lysine-type peptidoglycan in insects. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 6602-6607.

Pauchet, Y., Freitak, D., Heidel-Fischer, H. M., Heckel, D. G., & Vogel, H. (2009). Immunity or digestion: glucanase activity in a glucan-binding protein family from Lepidoptera. *J. Biol. Chem*. **284**, 2214-2224.

Pili-Floury, S., Leulier, F., Takahashi, K., Saigo, K., Samain, E., Ueda, R., & Lemaitre, B. (2004). In vivo RNA interference analysis reveals an unexpected role for GNBP1 in the defense against Gram-positive bacterial infection in *Drosophila* adults. *J. Biol. Chem*. **279**, 12848- 12853.

Philo, J. S. (2000). A method for directly fitting the time derivative of sedimentation velocity data and an alternative algorithm for calculating sedimentation coefficient distribution functions. *Anal. Biochem*. **279**, 151-163.

Raetz, C. R., & Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annu. Rev. Biochem*. **71**, 635-700.

Ragan, E. J., An, C., Yang, C. T., & Kanost, M. R. (2010). Analysis of mutually exclusive alternatively spliced serpin-1 isoforms and identification of serpin-1 proteinase complexes in *Manduca sexta* hemolymph. *J. Biol. Chem.* 285, 29642-29650.

Roh, K. B., Kim, C. H., Lee, H., Kwon, H. M., Park, J. W., Ryu, J. H., Kurokawa, K., Ha, N. C., Lee, W. J., Lemaitre, B., Söderhäll, K., & Lee, B. L. (2009). Proteolytic cascade for the activation of the insect toll pathway induced by the fungal cell wall component. *J. Biol. Chem.* **284**, 19474-19481.

Royet, J. (2004). Infectious non-self recognition in invertebrates: lessons from *Drosophila* and other insect models. *Mol. Immunol*. **41**, 1063-1075.

Royet, J., & Dziarski, R. (2007). Peptidoglycan recognition proteins: pleiotropic sensors and effectors of antimicrobial defences. *Nat. Rev. Microbiol*. **5**, 264-277.

Royet, J., Gupta, D., & Dziarski, R. (2011). Peptidoglycan recognition proteins: modulators of the microbiome and inflammation. *Nat. Rev. Immunol*. **11**, 837-851.

Schmidt, O., Söderhäll, K., Theopold, U., & Faye, I. (2010). Role of adhesion in arthropod immune recognition. *Annu. Rev. Entomol*. **55**, 485-504.

Schneider, D. S., & Chambers, M. C. (2008). Microbiology. Rogue insect immunity. *Science* **322**, 1199-1200.

Su, X. D., Gastinel, L. N., Vaughn, D. E., Faye, I., Poon, P., & Bjorkman, P.J. (1998). Crystal structure of hemolin: a horseshoe shape with implications for homophilic adhesion. *Science* **281**, 991-995.

Sumathipala, N. & Jiang, H. (2010). Involvement of *Manduca sexta* peptidoglycan recognition protein-1 in the recognition of bacteria and activation of prophenoloxidase system. *Insect Biochem. Mol. Biol.* **40**, 487-495.

Sun, S. C., Lindström, I., Boman, H. G., Faye, I., & Schmidt, O. (1990). Hemolin: an insect-immune protein belonging to the immunoglobulin superfamily. *Science* **250**, 1729-1732.

Tanaka, H., Ishibashi, J., Fujita, K., Nakajima, Y., Sagisaka, A., Tomimoto, K., Suzuki, N., Yoshiyama, M., Kaneko, Y., Iwasaki, T., Sunagawa, T., Yamaji, K., Asaoka, A., Mita, K., & Yamakawa, M. (2008). A genome-wide analysis of genes and gene families involved in innate immunity of *Bombyx mori*. *Insect Biochem. Mol. Biol*. **38**, 1087-1110.

Vollmer, W., & Seligman, S. J. (2010). Architecture of peptidoglycan: more data and more models. *Trends Microbiol*. **18**, 59-66.

Wang, L., Gilbert, R. J., Atilano, M. L., Filipe, S. R., Gay, N. J., & Ligoxygakis, P. (2008). Peptidoglycan recognition protein-SD provides versatility of receptor formation in *Drosophila* immunity. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11881-11886.

Wang, L., Weber, A. N., Atilano, M. L., Filipe, S. R., Gay, N. J., & Ligoxygakis, P. (2005). Sensing of Gram-positive bacteria in *Drosophila*: GNBP1 is needed to process and present peptidoglycan to PGRP-SA. *EMBO J*. **25**, 5005-5014.

Wang, X., Rocheleau, T.A., Fuchs, J. F., & Christensen, B. M. (2006). Beta 1, 3-glucan recognition protein from the mosquito, *Armigeres subalbatus*, is involved in the recognition of distinct types of bacteria in innate immune responses. *Cell Microbiol*. **8**, 1581-1590.

Wang, Y., & Jiang, H. (2006). Interaction of β-1,3-glucan with its recognition protein activates hemolymph proteinase 14, an initiation enzyme of the prophenoloxidase activation system in *Manduca sexta*. *J. Biol. Chem.* **281**, 9271-9278.

Wang, Y. & Jiang, H. (2010). Binding properties of the regulatory domains in *Manduca sexta* hemolymph proteinase-14, an initiation enzyme of the prophenoloxidase activation system. *Dev. Comp. Immunol*. **34**, 316-322.

Wang, Y., Sumathipala, N., Rayaprolu, S., & Jiang, H. (2011). Recognition of microbial molecular patterns and stimulation of prophenoloxidase activation by a β-1,3-glucanase-related protein in *Manduca sexta* larval plasma. *Insect Biochem. Mol. Biol*. **41**, 322-331.

Wang, Y., Yang, P., Cui, F., Kang, L. (2013) Altered immunity in crowded locust reduced fungal (*Metarhizium anisopliae*) pathogenesis. *PLoS Pathog*. **9**, e1003102.

Warr, E., Das, S., Dong, Y., & Dimopoulos, G. (2008). The Gram-negative bacteria-binding protein gene family: its role in the innate immune system of *Anopheles gambiae* and in anti-*Plasmodium* defence. *Insect Mol. Biol*. **17**, 39-51.

Warshakoon, H. J., Burns, M. R., & David, S. A. (2009). Structure-activity relationships of antimicrobial and lipoteichoic acid-sequestering properties in polyamine sulfonamides. *Antimicrob. Agents Chemother*. **53**, 57-62.

Weidenmaier, C., & Peschel, A. (2008). Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions. *Nat. Rev. Microbiol.* **6**, 276-287.

Wüthrich, K. (1989). Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science* **243**, 45-50.

Yoshida, H., Kinoshita, K., & Ashida, M. (1996). Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, *Bombyx mori*. *J. Biol. Chem*. **271**, 13854-13860.

Yoshida, H., Ochiai, M., & Ashida, M. (1986). β-1,3-glucan receptor and peptidoglycan receptor are present as separate entities within insect prophenoloxidase activating system. *Biochem. Biophys. Res. Commun*. **141**, 1177-1184.

Yu, X. Q., Gan, H., & Kanost, M. R. (1999). Immulectin, an inducible C-type lectin from an insect, *Manduca sexta*, stimulates activation of plasma prophenol oxidase. *Insect Biochem. Mol. Biol*. **29**, 585-597.

Yu, X. Q., & Kanost, M. R. (2000). Immulectin-2, a lipopolysaccharide-specific lectin from an insect, *Manduca sexta*, is induced in response to Gram-negative bacteria. *J. Biol. Chem*. **275**, 37373-37381.

Yu, X. Q., & Kanost, M. R. (2002). Binding of hemolin to bacterial lipopolysaccharide and lipoteichoic acid. An immunoglobulin superfamily member from insects as a pattern-recognition receptor. *Eur. J. Biochem*. **269**, 1827-1834.

Yu, X. Q., Ling, E., Tracy, M. E., & Zhu, Y. (2006). Immulectin-4 from the tobacco hornworm *Manduca sexta* binds to lipopolysaccharide and lipoteichoic acid. *Insect Mol. Biol*. **15**, 119-128.

Yu, X. Q., Tracy, M. E., Ling, E., Scholz, F. R., & Trenczek, T. (2005). A novel C-type immulectin-3 from *Manduca sexta* is translocated from hemolymph into the cytoplasm of hemocytes. *Insect Biochem. Mol. Biol*. **35**, 285-295.

Yu, X. Q., Zhu, Y. F., Ma, C., Fabrick, J. A., & Kanost, M. R. (2002). Pattern recognition proteins in *Manduca sexta* plasma. *Insect Biochem. Mol. Biol*. **32**, 1287-1293.

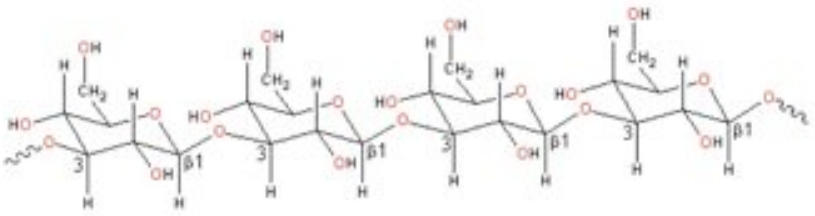
Zhang, R., Cho, H. Y., Kim, H. S., Ma, Y. G., Osaki, T., Kawabata, S., Söderhäll, K., & Lee, B. L. (2003). Characterization and properties of a 1,3-β-D-glucan pattern recognition protein of *Tenebrio molitor* larvae that is specifically degraded by serine protease during prophenoloxidase activation. *J. Biol. Chem*. **278**, 42072-42079.

Zheng, X., & Xia, Y. (2012). β-1,3-Glucan recognition protein (βGRP) is essential for resistance against fungal pathogen and opportunistic pathogenic gut bacteria in *Locusta migratoria* manilensis. *Dev. Comp. Immunol*. **36**, 602-609.

Zhu, Y., Johnson, T. J., Myers, A. A., & Kanost, M. R. (2003). Identification by subtractive suppression hybridization of bacteria-induced genes expressed in *Manduca sexta* fat body. *Insect Biochem. Mol. Biol*. **33**, 541-559.

Zhu, Y., Ragan, E. J., & Kanost, M. R. (2010). Leureptin: a soluble, extracellular leucine-rich repeat protein from *Manduca sexta* that binds lipopolysaccharide. *Insect Biochem. Mol. Biol*. **40**, 713-722.

Zou, Z., Evans, J. D., Lu, Z., Zhao, P., Williams, M., Sumathipala, N., Hetru, C., Hultmark, D., Jiang, H. (2007). Comparative genomic analysis of the *Tribolium* immune system. *Genome Biol*. **8**, R177.



β-1,3-glucan

βGRP/GNBP3

GNBP1

glucanase

Figure 1-1 Schematic structure of β-1,3-glucan and βGRP/GNBP categories.

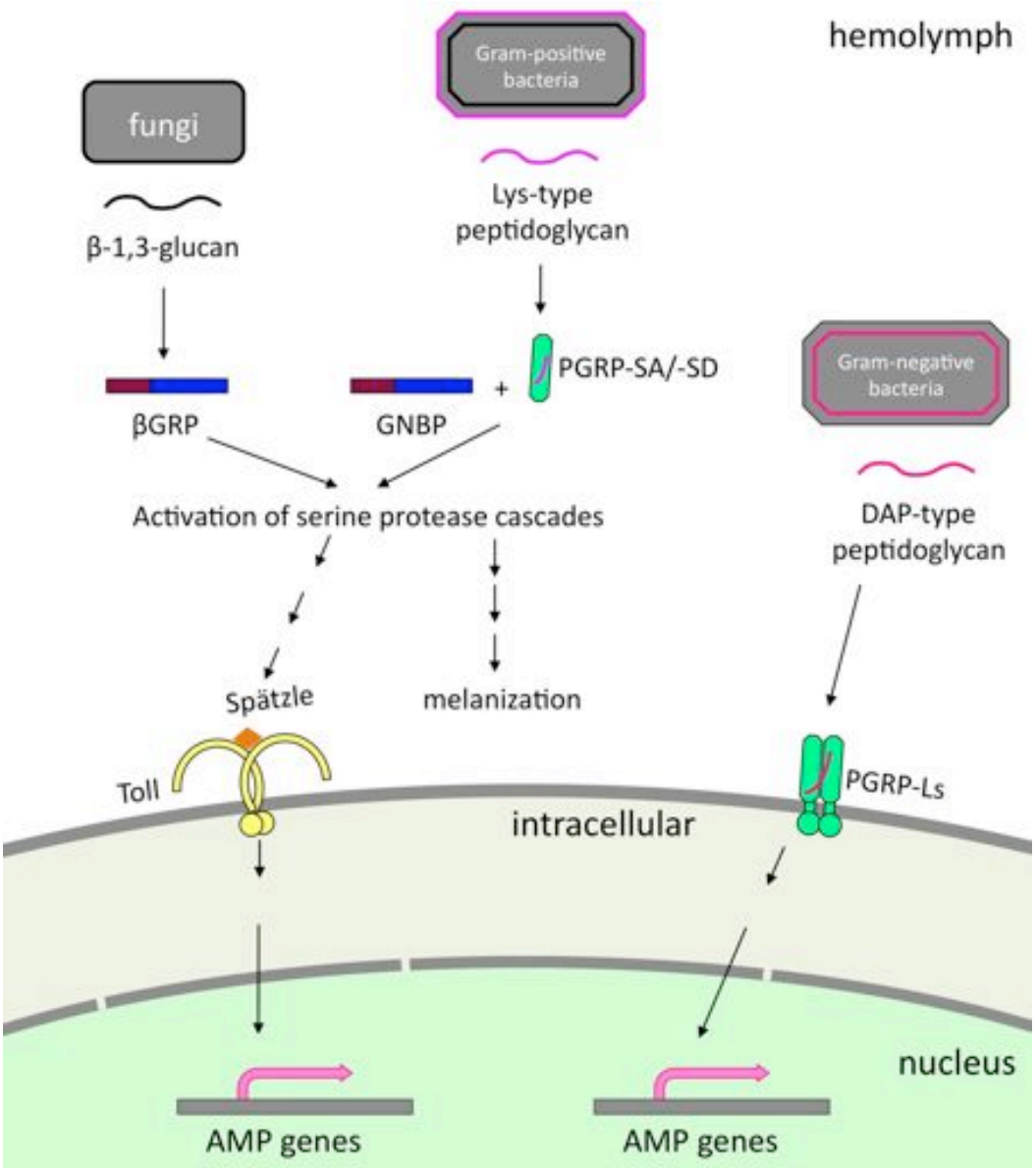


Figure 1-2 Pathogen recognition and activation of Toll and IMD pathways. The soluble pathogen recognition receptors (βGRPs/GNBPs and PGRP-Ss) recognize fungi and Gram-positive bacteria, activating extracellular serine protease cascades that lead to melanization and AMP expression through Toll receptor ligand Spätzle. The membrane receptor PGRP-Ls oligomerize upon DAP-type peptidoglycan recognition and trigger AMP expression through intracellular IMD signaling.

Table 1-1 PGRPs and βGRPs/GNBPs identified in five insects.

Genes	Subfamily	*D. melanogaster*	*A. gambiae*	*B. mori*	*M. sexta*	*T. castaneum*
PGRP	PGRP-L	6	3	6	?	2
	PGRP-S	9	4	6	1(?)	6
βGRP/ GNBP	βGRP/GNBP	GNBP1 GNBP2 GNBP3	GNBPA1 GNBPA2	βGRP1 βGRP2 βGRP3	βGRP1 βGRP2 MBP	βGRP2 βGRP3
	glucanase	\	GNBPB1 GNBPB2 GNBPB3 GNBPB4	βGRP4	?	βGRP1

The numbers of PGRPs and the names of βGRPs/GNBPs are listed. In *B. mori*, βGRP1 was also named βGRP/GNBP3; and βGRP2 was also named GNBP (Lee et al., 1996). In *M. sexta*, only PGRP1 has been characterized, which is a shorter form of PGRP (Sumathipala & Jiang, 2010). The annotations of *T. castaneum* PGRPs and βGRPs are from Zou et al. (2007).

Chapter 2 - NMR Solution Structure Determination of the N-terminal Domain of βGRP from *Plodia interpunctella*

Introduction

βGRPs possess in common a conserved amino-terminal carbohydrate-binding domain and a carboxyl-terminal β-1,3 glucanase-like domain. The amino-terminal domain is a carbohydrate-binding module (CBM) that is assigned to family CBM39 according to the Carbohydrate-Active Enzyme database (Cantarel et al., 2009). It is established that CBM39 can bind to β-1,3-glucan including curdlan and laminarin (Ochiai et al., 2000; Fabrick et al., 2004). So far, CBM39-containing proteins are exclusively found in βGRPs, and their sequences are highly similar, particularly among the well-characterized βGRPs from *P. interpunctella*, *B. mori*, *M. sexta* and *T. molitor*, and GNBP3 from *D. melanogaster* (Figure 2-1).

P. interpunctella N-βGRP mixed with laminarin generates a significant synergistic activation of the proPO pathway (Fabrick et al., 2004). N-βGRP also induces aggregation of microorganisms such as *S. cerevisiae* and *E. coli*, albeit less effectively than does the full-length protein (Fabrick et al., 2004). Aggregation of pathogens *in vivo* may create a superior trigger for activating biochemical cascades of cellular immunity (Fabrick et al., 2004) or may provide a platform to assemble effector complexes (Matskevich et al., 2010). To characterize the structural basis of carbohydrate recognition by βGRPs, we have first determined the solution structure of the amino-terminal CBM39 domain of βGRP from Indianmeal moth *Plodia interpunctella* by using heteronuclear multi-dimensional NMR spectroscopy. We denote the recombinant amino-terminal 118-residue polypeptide as N-βGRP.

Materials and Methods

Protein Expression and Purification

DNA sequence of N-βGRP was cloned via BamHI/HindIII sites into Invitrogen pPROEX HTb plasmid. The resulting recombinant protein therefore contains an N-terminal hexahistidine tag followed by a TEV protease cleavage site before N-βGRP. His_6-N-βGRP was expressed in *Escherichia coli* by induction with isopropyl-β-D-thiogalactopyranoside (IPTG) and purified

using a Ni^{2+}-affinity column. The hexahistidine tag was cleaved by mixing His_6-N-βGRP with His_6-TEV protease. The incubation mixture was then passed through the Ni^{2+} affinity column to remove the cleaved hexahistidine tag and the TEV protease.

NMR Spectroscopy

The solution sample for NMR spectroscopy was 1.7 - 2.0 mM uniformly $^{13}C/^{15}N$-labeled N-βGRP in 20 mM sodium phosphate buffer, pH 6.5. All spectra were recorded at 25°C on a Bruker Avance 800 spectrometer at The Structural Biology Center of the University of Kansas. All data sets were processed using NMRPipe (Delaglio et al., 1995) and analyzed using Sparky (Goddard & Kneller, 2001). 2D- and 3D-NMR spectra were collected for backbone and side-chain chemical shift assignments and structural constraints (Cavanagh et al., 1996). These include 2D ^{15}N-HSQC and ^{13}C-HSQC, 3D CBCA(CO)NH, HNCACB, HN(CO)CA, HNCA, CCCONH, HCCCONH, HCCH-TOCSY, ^{15}N-edited TOCSY, ^{15}N-edited NOESY, and ^{13}C-edited NOESY. An additional 3D ^{15}N-edited NOESY-HSQC data set was gathered from the ^{15}N-labeled sample in order to identify HN-HN contacts.

Sequence-specific backbone ^{1}H, ^{15}N, and ^{13}C resonance assignments were based on the 3D CBCA(CO)NH, HNCACB, HN(CO)CA, and HNCA. The side chain resonances were assigned using the data from 3D CCCONH, HCCCONH, HCCH-TOCSY, and ^{15}N-edited TOCSY.

Structure Calculation

Unambiguous NOE constraints were manually assigned using 3D ^{15}N-edited and ^{13}C-edited NOESY spectra. Redundant constraints were removed with Aqua (Laskowski et al., 1996). Dihedral constraints for the backbone torsion angles (φ, ψ) were obtained from TALOS (Cornilescu et al., 1999). Hydrogen bond constraints were employed for β-stand regions based on the NOE patterns and the chemical shift index prediction of the secondary structure (Wishart et al., 1994). Structures of N-βGRP were calculated by simulated annealing procedure using CNS (Brünger et al., 1998). A total of 100 structures were calculated from which 20 lowest-energy structures were chosen for the final structural ensemble.

Results and Discussion

Chemical Shift Assignments of N-βGRP

The ^{1}H-^{15}N HSQC spectrum of N-βGRP showed sharp signals in a well-dispersed pattern, which is typical for β strands (Figure 2-2). The resonance assignments for the backbone amide ^{1}H and ^{15}N nuclei of residues Tyr7-Glu117 were completed, except for Arg69 (Table 2-1). Figure 2-3 shows the backbone connectivity from residue Asn70 to Ile79 provided by CBCA(CO)NH and HNCACB. The chemical shifts of ^{1}H, ^{15}N, and ^{13}C were assigned using standard heteronuclear NMR methods (Clore & Gronenborn, 1994). Upfield Cγ chemical shift (24.36 ppm) and downfield Cβ chemical shift (33.84 ppm) of Pro19, together with the strong NOE correlation between its Hα and Hα of the preceding Tyr18, indicate that the peptide bond between Tyr18 and Pro19 is in the *cis* conformation (Figure 2-4). Chemical shift values were deposited into Biological Magnetic Resonance Bank (BMRB) with accession number 16231.

Solution Structure of N-βGRP

Table 2-2 summarizes the structural statistics for the 20 structures that have been deposited to the Protein Data Bank with accession number 2KHA. The root mean square deviation (rmsd) for residues Ser11- Val105 is 0.35 for backbone and 0.93 for non-hydrogen atoms, indicating a well-folded structure. The heteronuclear steady-state NOE values are indicative of the dynamic properties of the amide nitrogen-hydrogen internuclear vectors in a protein (Peng & Wagner, 1994). For residues Ser11- Val105, the average value of heteronuclear NOE is 0.83, confirming that this region is a well-folded domain. The decreased heteronuclear NOE values observed for some residues in the amino- and carboxyl-terminal regions are consistent with increased flexibility and lack of NOE constraints in the terminal regions (Figure 2-5).

The solution structure of N-βGRP consists of two anti-parallel β-sheets of total eight β-strands, representing an immunoglobulin-like β-sandwich fold (Figure 2-6). We named the β-strands according to the standard immunoglobulin nomenclature (Amzel & Poljak, 1979). The first β sheet is comprised of β-strand A (Lys13-Ile17), B (Gly21-Pro27) and E (Arg63-Asp68). The second β-sheet is comprised of β-strand C (Ser32-Leu40), C' (His51-Ile56), F (Lys78-Ile86), G (Gly91-Gln94) and G' (Gly97-Thr100). Strands G and G' are split by a kink (Asp95-Gln96). The hydrophobic core between two sheets is tightly packed (Figure 2-7). It is mainly

composed of phenylalanines (Phe31, Phe34, Phe36, Phe66 and Phe81) and aliphatic amino acid residues. Three tryptophans (Trp52, Trp64, Trp99) are buried within the interior of the protein and contribute to the hydrophobic core.

Based on the aromatic platform observed for other glucan binding proteins (Boraston et al., 2004), Mishima et al. (2009) proposed that loop C-C' could undergo a large conformational change to expose the interior Trp81. In contrast, we found abundant NOE constraints for this loop (Figure 2-8); all these NOE constraints have been deposited into PDB with accession number 2KHA. Furthermore, we measured the intensity of fluorescence emitted by tryptophans, among which Trp81 is shielded by loop C-C', before and after addition of laminarin to the protein sample, and found no change. This indicates that laminarin-binding does not cause any conformational changes for loop C-C'. Interestingly, the crystal structure of *P. interpunctella* N-βGRP complexed with laminarihexaose (Kanagawa et al., 2011) also shows no conformational alteration.

Surface of N-βGRP

The involvement of aromatic residues in protein-carbohydrate interactions has been shown in CBM-carbohydrate complex structures (Boraston et al., 2004) (Figure 2-9). In N-βGRP, aromatic residues not buried in the hydrophobic core are Tyr7, Tyr18, Tyr80, Trp82, Tyr84, Tyr92 and Phe104. Among these seven aromatic residues, Tyr7 and Tyr92 interact with each other; Tyr18 and Phe104 form a hydrophobic patch; Tyr80, Trp82, and Tyr84 are located on strand F and form a platform. However, a nearby conserved acidic loop (Asn41-Gly50) between strand C and C', herein referred as loop C-C', blocks full exposure of Tyr80 and Trp82 (Figure 2-10). The heteronuclear NOE values determined for this loop (Figure 2-5), together with abundant NOEs observed between Met44 and Tyr80/Trp82, indicate that this loop adopts rather a rigid conformation. Thus, the surface of N-βGRP has no carbohydrate-binding 'platforms' made of aromatic residues, and in contrast, presents hydrophilic patches (Figure 2-10).

Structural Relatives of N-βGRP

The structure of N-βGRP consists of an immunoglobulin-like β-sandwich fold as specified in the SCOP database (Andreeva et al., 2004). The immunoglobulin-like β-sandwich fold is the second most common fold in CBM families (the first being the β-jelly roll fold). The

immunoglobulin fold is one of the most prevalent folds among a wide variety of organisms, including *Drosophila* and mammals (Adams et al., 2000; Venter et al., 2001). Proteins with immunoglobulin folds include antibodies, immune receptors, transcriptional factors, and cell surface proteins, among others. Two interesting examples are hemolin and DSCAM in insect immunity (Su et al., 1998; Meijers et al., 2007). The most common structural feature of the immunoglobulin fold is a central hydrophobic core between its two β-sheets, with little sequence similarities among its various members (Clarke et al., 1999).

Three-dimensional structures have been previously described for three members of CBM39 family: N-βGRPs from *B. mori* (Takahasi et al., 2009; Kanagawa et al., 2011), *P. interpunctella* (Kanagawa et al., 2011) and *D. melanogaster* (Mishima et al., 2009). The overall structure of these three CBM39 members is highly conserved as indicated by their high Z scores (all > 10) from DALI search (Holm & Rosenström, 2010). We noticed that although N-βGRP shares 81% sequence identity to *B. mori* CBM39, its structure is more similar to *D. melanogaster* CBM39 despite the fact that the fruitfly protein has a lower sequence identity (64%) (Figure 2-11). The major differences between N-βGRP and the *B. mori* protein include the *cis*/*trans* peptide bond preceding the Pro residue between strands A and B (Pro19 in N-βGRP), backbone orientations of loop between strands E and F, and of the carboxyl-terminal region after strand G'. As this proline is conserved in CBM39 members (Figure 2-1) and the structures determined by the present NMR investigation and by X-ray crystallography earlier (Mishima et al., 2009; Kanagawa et al., 2011) all confirm its *cis* conformation, the *trans* conformation reported by Takahasi et al. (2009) is concluded to be in error.

CBM39 Is a Novel Functional Type of CBM Families

A summary of CBM members other than CBM39 with a Z score > 4 hit by DALI search (Holm & Rosenström, 2010) using N-βGRP as the search model is shown in Table 2-3. They are CBM9, CBM21, CBM31 and CBM34, and carbohydrate-binding 'platforms' made by aromatic residues have been observed in these four CBM families (Notenboom et al., 2001; Liu et al., 2007; Hashimoto et al., 2005; Kamitori et al., 1999). However, such a 'platform' is not observed in CBM39 (Figure 2-10). Boraston et al. (2004) categorized CBMs into three functional types, A, B and C, corresponding to surface-binding, glycan-chain-binding, and small-sugar-binding property, respectively. These four structural relatives of CBM39 are categorized into all three

functional types, thus indicating that there is no correlation between a CBM fold family and its functional type. Furthermore, the hydrophilic patches of CBM39 surface suggest that CBM39 is a novel functional type.

Since the appearance of the term CBM in 1999, sixty-five CBMs have been annotated, namely CBM1 to CBM66, with CBM7 being excluded later in CAZy. Among the sixty-five CBMs identified so far, fifty CBMs have known three-dimensional structures (Table 2-4), of which twelve are of the immunoglobulin-fold type. These twelve immunoglobulin-fold members can be further grouped into two types, I and II, based on their topology (Liu et al., 2007). Type I and type II differ in that the eighth β-strand (the last β-strand) in type I transits to the first β-strand in type II. Type I has five members and type II has seven. CBM39 belongs to type II group that also includes CBM 9, CBM21, CBM31, CBM33, CBM34 and CBM48. Among the six type II immunoglobulin-fold members except CBM39, four (CBM9, CBM21, CBM31, and CBM34) have a high Z score (> 4 to CBM39, Table 2-3). Five, including the afore-mentioned four members plus CBM48, the largest CBM family in CAZy, have surface-exposed aromatic patches, and can be identified with one of the three functional types. CBM33 is a chitin-binding module, and like CBM39, has no such patches. In contrast, the polar side chains, rather than the aromatic side chains, of CBM33 are implicated in chitin interactions (Vaaje-Kolstad et al., 2005). Such characteristics are similar to the CBM39 surface (Figure 2-10). However, it should be noted that CBM39 (N-βGRP) can bind to curdlan, an insoluble polysaccharide, and to laminarin, a soluble glycan (Ochiai & Ashida, 2000; Fabrick et al., 2004), which suggests CBM39 employs a functional mode different from that of chitin-binding CBM33.

In summary, there is no relatedness between the CBM fold and its function types. Surface analysis of the CBM39 solution structure indicates a novel functional type, because CBM39 has no planar patch formed by aromatic side chains as observed in type A CBMs, no groove that is distinctive in type B CBMs, or no pocket to accommodate small sugars as is characteristic of type C CBMs.

Discovery of the immunoglobulin fold adopted by CBM39, together with structural comparision with other related CBM families, implies that N-βGRP may have evolved into a new functional type. Proteins containing immunoglobulin folds participate in a variety of biological events. For instance, when we used DALI server (Holm & Rosenström, 2010) to search the structural homologs of N-βGRP, the two structures with the high Z scores (> 8) are

not CBM members. One is the sixth fibronectin type III (FnIII) domain of human netrin receptor DCC (PDB accession number 2EDE), and the other is the first FnIII domain of human integrin α6β4 (de Pereda, et al., 2009). Netrin receptor DCC is a membrane receptor for neuronal guide signal protein, netrin-1 (Rajasekharan & Kennedy, 2009); and integrin α6β4 is essential for the formation of the junctional adhesion complexes between epithelial cells and the epithelial basement membrane. Although neither of these two closest structural homologs to CBM39 can provide clues to the structure-function relationship for β-1,3-glucan recognition, their role in protein-protein interactions seems to hint at functions beyond protein-carbohydrate interaction for βGRPs

βGRP is composed of two conserved domains, and the C-terminal glucanase-like domain implies β-1,3-glucan-binding activity. Sequence analysis shows that this C-terminal domain is a homolog of glucanase, an enzyme belonging to Glycoside Hydrolase family 16 (GH16). A homology model of βGRP based on structures of CBM39 and glucanase built by SWISS-MODEL (Kiefer et al., 2009) suggests that the GH16 domain can recognize β-1,3-glucan through a groove on surface (Figure 2-12). The fold of GH16 is β-jelly roll, the most abundant fold in all known CBMs (Table 2-4). Considering the biological data that the N-terminal domain alone can trigger melanization while the C-terminal domain alone cannot (Fabrick et al., 2004), the immunoglobulin-like CBM39 should provide a molecular bridge between β-1,3-glucan recognition and immune signal intiation. This notion needs to be tested in future studies.

Conclusions

The solution structure of N-βGRP has been determined by NMR spectroscopy. It consists of an immunoglobulin fold. This is the first structure determined for a CMB39 family member (Table 2-5). The structure of N-βGRP differs from that of the silkworm protein (Takahashi et al., 2009) in that a Pro in a loop region exists in a *cis*, rather than *trans* conformation. The loop C-C' is rigid, in contrast to an earlier proposal (Mishima et al., 2009), and shields the aromatic residues that have been observed in other CBM structures to serve as a platform for carbohydrate binding. Hydrophilic patches instead occur as a characteristic feature of N-βGRP surface. Structural comparisons between N-βGRP and other related proteins suggest that N-βGRP is likely to present a novel type of carbohydrate-binding site.

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., et al. (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185-2195.

Amzel, L.M., & Poljak, R.J. (1979). Three-dimensional structure of immunoglobulins. *Annu. Rev. Biochem.* **48**, 961-997.

Andreeva, A., Howorth, D., Brenner, S. E., Hubbard, T. J. P., Chothia, C., & Murzin, A. G. (2004). SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res*. **32**, D226-D229.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., & Warren, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.* **54**, 905-921.

Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769-781.

Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*. **37**, D233-D238.

Cavanagh, C., Palmer, A. G. 3rd, & Skelton, N. J. (1996). *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego.

Clarke, J., Cota, E., Fowler, S. B., & Hamill, S. J. (1999). Folding studies of immunoglobulin-like β-sandwich proteins suggest that they share a common folding pathway. *Structure* **7**, 1145-1153.

Clore, G.M. & Gronenborn, A.M. (1994). Structures of larger proteins, protein-ligand and protein-DNA complexes by multi-dimensional heteronuclear NMR. *Protein Science* **3**, 372-390.

Cornilescu, G., Delaglio, F., & Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289-302.

Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293.

DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*. DeLano Scientific, San Carlos, CA, USA.

Fabrick, J. A., Baker, J. E., & Kanost, M. R. (2004). Innate immunity in a pyralid moth: functional evaluation of domains from a β-1,3-glucan recognition protein. *J. Biol. Chem.* **279**, 26605-26611.

Fibriansah, G., Masuda, S., Koizumi, N., Nakamura, S., & Kumasaka, T. (2007). The 1.3 A crystal structure of a novel endo-β-1,3-glucanase of glycoside hydrolase family 16 from alkaliphilic *Nocardiopsis* sp. strain F96. *Proteins* **69**, 683-690.

Goddard, T. D., & Kneller, D. G. (2001) SPARKY 3, University of California, San Francisco, CA.

Gouet, P., Robert, X., & Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res*. **31**, 3320-3323.

Hashimoto, H. (2006). Recent structural studies of carbohydrate-binding modules. *Cell Mol. Life Sci.* **63**, 2954-2967.

Hashimoto, H., Tamai, Y., Okazaki, F., Tamaru, Y., Shimizu, T., Araki, T., & Sato, M. (2005). The first crystal structure of a family 31 carbohydrate-binding module with affinity to β-1,3-xylan. *FEBS Lett*. **579**, 4324-4328.

Holm, L., & Rosenström, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res*. 38, W545-W549.

Jiang, H., Ma, C., Lu, Z. Q., & Kanost, M. R. (2004) β-1,3-glucan recognition protein-2 (βGRP-2) from *Manduca sexta*; an acute-phase protein that binds β-1,3-glucan and lipoteichoic acid to aggregate fungi and bacteria and stimulate prophenoloxidase activation. *Insect Biochem. Mol. Biol.* **34**, 89-100.

Kamitori, S., Kondo, S., Okuyama, K., Yokota, T., Shimura, Y., Tonozuka, T., & Sakano, Y. (1999). Crystal structure of *Thermoactinomyces vulgaris* R-47 alpha-amylase II (TVAII) hydrolyzing cyclodextrins and pullulan at 2.6 A resolution. *J. Mol. Biol*. **287**, 907-921.

Kanagawa, M., Satoh, T., Ikeda, A., Adachi, Y., Ohno, N., & Yamaguchi, Y. (2011). Structural insights into recognition of triple-helical β-glucans by an insect fungal receptor. *J. Biol. Chem.* **286**, 29158-29165.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucl. Acids Res*. **37**, D387-D392.

Kim, Y. S., Ryu, J. H., Han, S. J., Choi, K. H., Nam, K. B., Jang, I. H., Lemaitre, B., Brey, P. T., & Lee, W. J. (2000). Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and β-1,3-glucan that mediates the signaling for the induction of innate immune genes in *Drosophila melanogaster* cells. *J. Biol. Chem.* **275**, 32721-32727.

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR*. **8**, 477-486.

Liu, Y. N., Lai, Y. T., Chou, W. I., Chang, M. D., & Lyu, P. C. (2007). Solution structure of family 21 carbohydrate-binding module from *Rhizopus oryzae* glucoamylase. *Biochem. J.* **403**, 21-30.

Matskevich, A. A., Quintin, J., & Ferrandon, D. (2010). The *Drosophila* PRR GNBP3 assembles effector complexes involved in antifungal defenses independently of its Toll-pathway activation function. *Eur. J. Immunol.* **40**, 1244-1254.

Meijers, R., Puettmann-Holgado, R., Skiniotis, G., Liu, J. H., Walz, T., Wang, J. H., & Schmucker, D. (2007). Structural basis of Dscam isoform specificity. *Nature* **449**, 487-491.

Mishima, Y., Quintin, J., Aimanianda, V., Kellenberger, C., Coste, F., Clavaud, C., Hetru, C., Hoffmann, J.A., Latge, J.P., Ferrandon, D., & Roussel, A. (2009). The N-terminal domain of *Drosophila* Gram-negative binding protein 3 (GNBP3) defines a novel family of fungal pattern recognition receptors. *J. Biol. Chem.* **284**, 28687-28697.

Notenboom, V., Boraston, A. B., Kilburn, D. G., & Rose, D. R. (2001). Crystal structures of the family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. *Biochemistry* **40**, 6248-6256.

Ochiai, M., & Ashida, M. (2000). A pattern-recognition protein for β-1,3-glucan. The binding domain and the cDNA cloning of β-1,3-glucan recognition protein from the silkworm, *Bombyx mori*. *J. Biol. Chem.* **275**, 4995-5002.

Peng, J. W., & Wagner, G. (1994). Investigation of protein motions via relaxation measurements. *Methods Enzymol.* **239**, 563-596.

Rajasekharan, S., & Kennedy, T. E. (2009). The netrin protein family. *Genome Biol.* **10**, 239. doi: 10.1186/gb-2009-10-9-239.

Su, X. D., Gastinel, L. N., Vaughn, D. E., Faye, I., Poon, P., & Bjorkman, P.J. (1998). Crystal structure of hemolin: a horseshoe shape with implications for homophilic adhesion. *Science* **281**, 991-995.

Takahasi, K., Ochiai, M., Horiuchi, M., Kumeta, H., Ogura, K., Ashida, M., & Inagaki, F. (2009). Solution structure of the silkworm βGRP/GNBP3 N-terminal domain reveals the mechanism for β-1,3-glucan-specific recognition. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11679-11684.

Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr. Protoc. Bioinformatics* Chapter 2:Unit 2.3.

Vaaje-Kolstad, G., Houston, D. R., Riemen, A. H., Eijsink, V. G., & van Aalten, D.M. (2005). Crystal structure and binding properties of the *Serratia marcescens* chitin-binding protein CBP21. *J. Biol. Chem.* **280**, 11313-11319.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, et al. (2001). The sequence of the human genome. *Science* **291**, 1304-1351.

Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647-1651.

Zhang, R., Cho, H. Y., Kim, H. S., Ma, Y. G., Osaki, T., Kawabata, S., Söderhäll, K., & Lee, B. L. (2003). Characterization and properties of a 1,3-β-D-glucan pattern recognition protein of *Tenebrio molitor* larvae that is specifically degraded by serine protease during prophenoloxidase activation. *J. Biol. Chem*. **278**, 42072-42079.

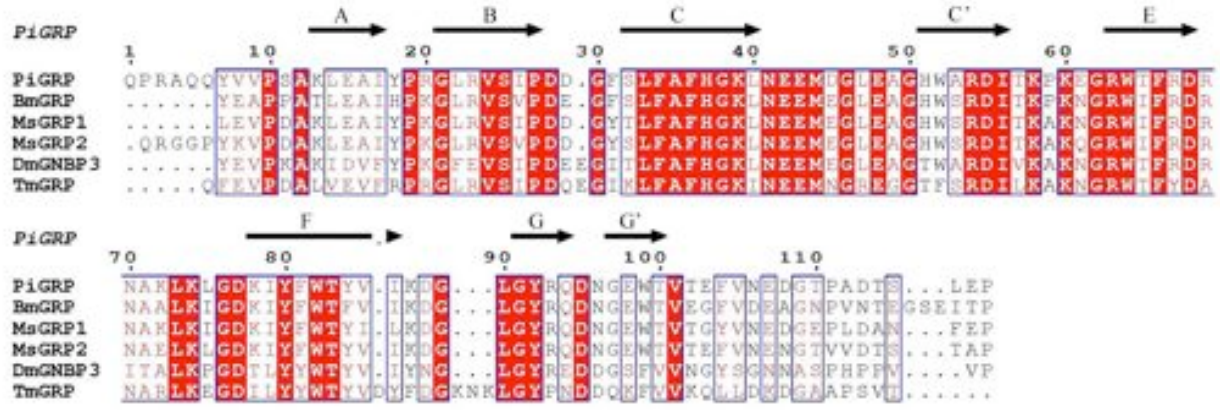


Figure 2-1 Alignment of amino acid sequence of N-βGRP from *P. interpunctella* (Fabrick et al., 2004) with those of N- domains of GNBP3/βGRP from *B. mori* (Ochiai & Ashida, 2000), *M. sexta* (Ma & Kanost, 2000; Jiang et al., 2004), *D. melanogaster* (Kim et al., 2000), and *T. molitor* (Zhang et al., 2003). Arrows correspond to β strands observed in the solution structure of N-βGRP. Sequence alignments were performed with ClustalW (Thompson et al., 2002) and ESPRIPT (Gouet et al., 2003).

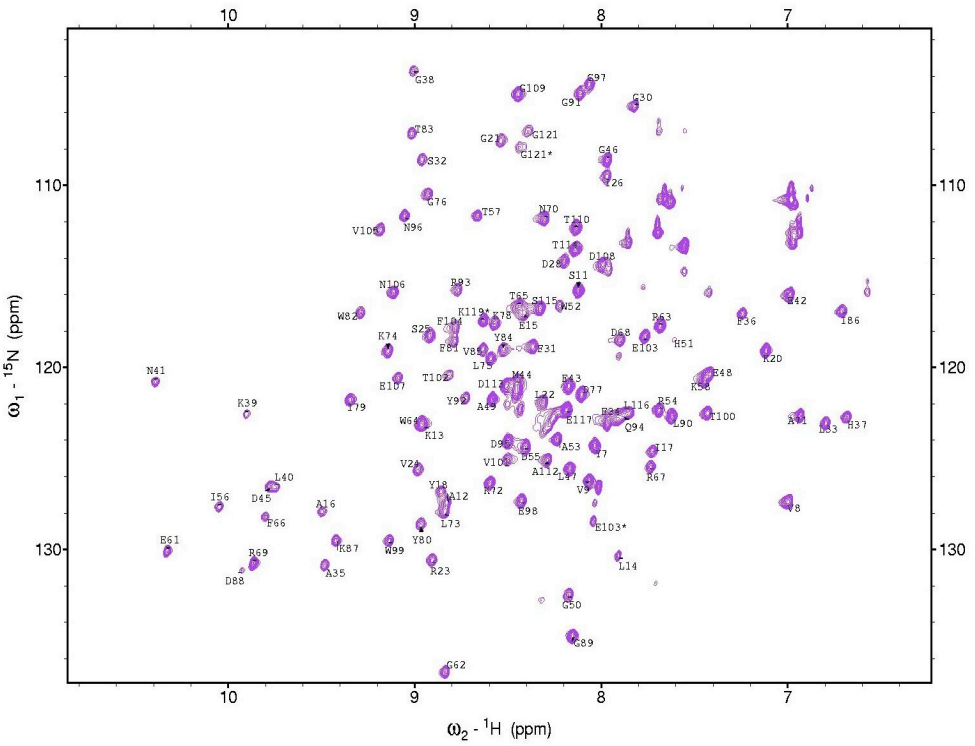


Figure 2-2 18.8 Tesla ^{1}H-^{15}N HSQC spectrum of N-βGRP at pH = 6.5, 25°C.

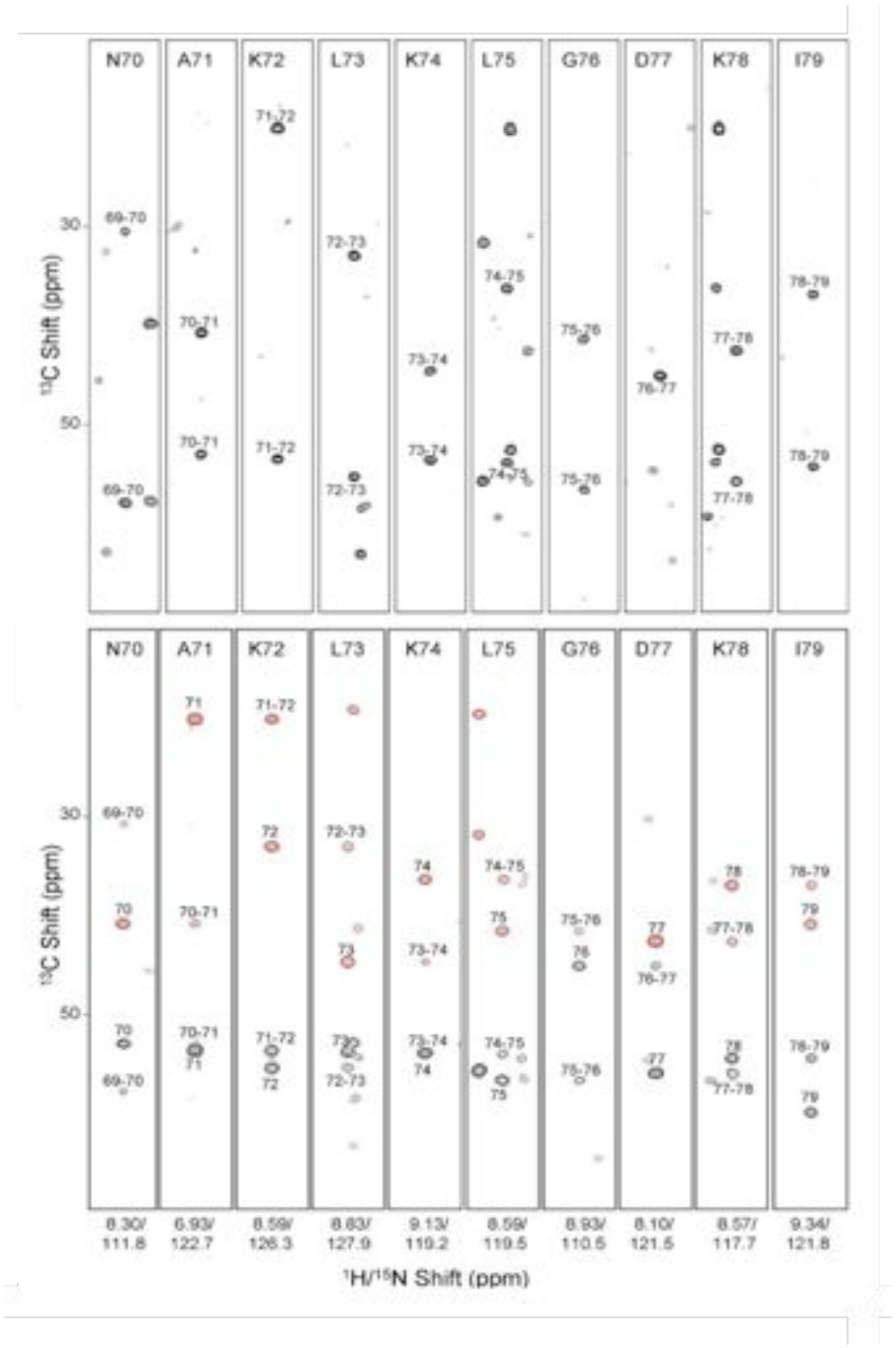


Figure 2-3 18.8 Tesla heteronuclear 3D spectra of N-βGRP. Backbone connectivities from Asn70 to Ile79 are shown as strip plots of CBCA(CO)NH (*upper*) and HNCACB (*lower*).

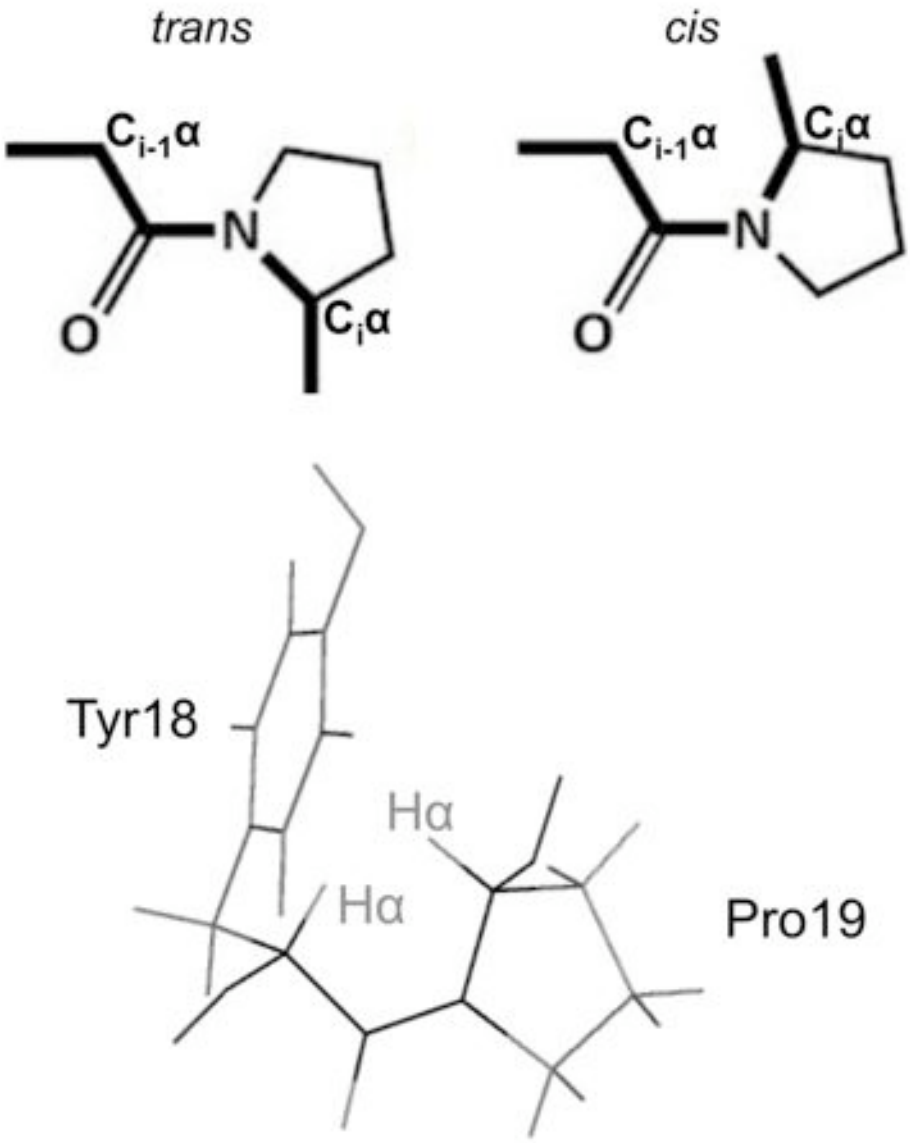


Figure 2-4 The *cis*/*trans* peptide bond between Pro and its preceding residue (*upper);* the *cis*-conformation of Pro19 (*lower*), as determined in this work.

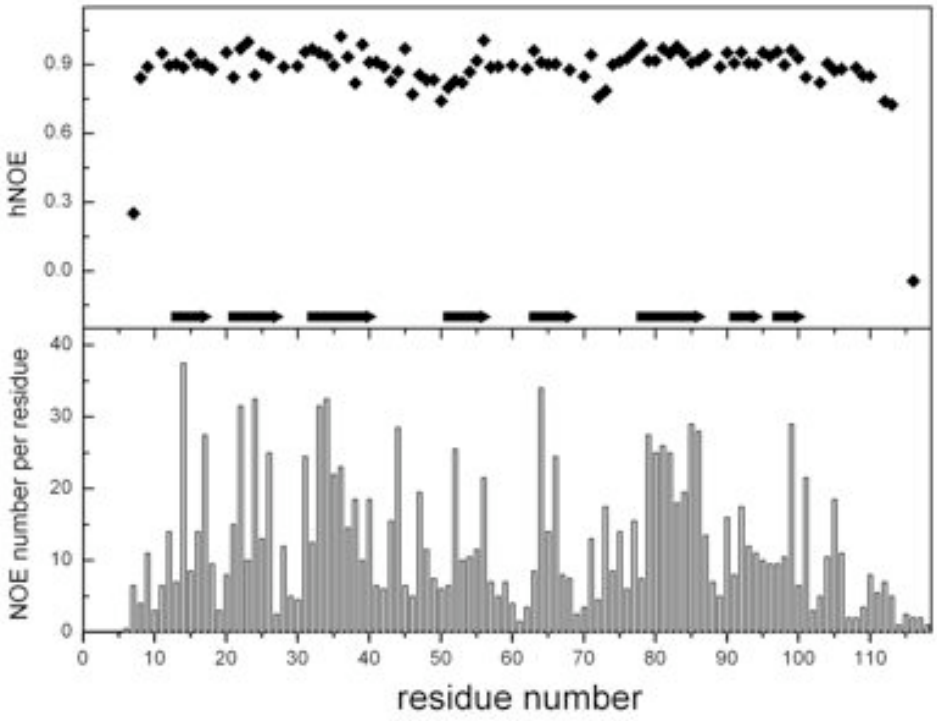


Figure 2-5 Heteronuclear ^{15}N-^{1}H NOEs (*upper*) and number of NOE constraints (*lower*) determined for the amino acid residues of N-βGRP. Heteronuclear NOEs corresponding to β strands are identified with arrows. The NOE constraints were used for structure calculation.

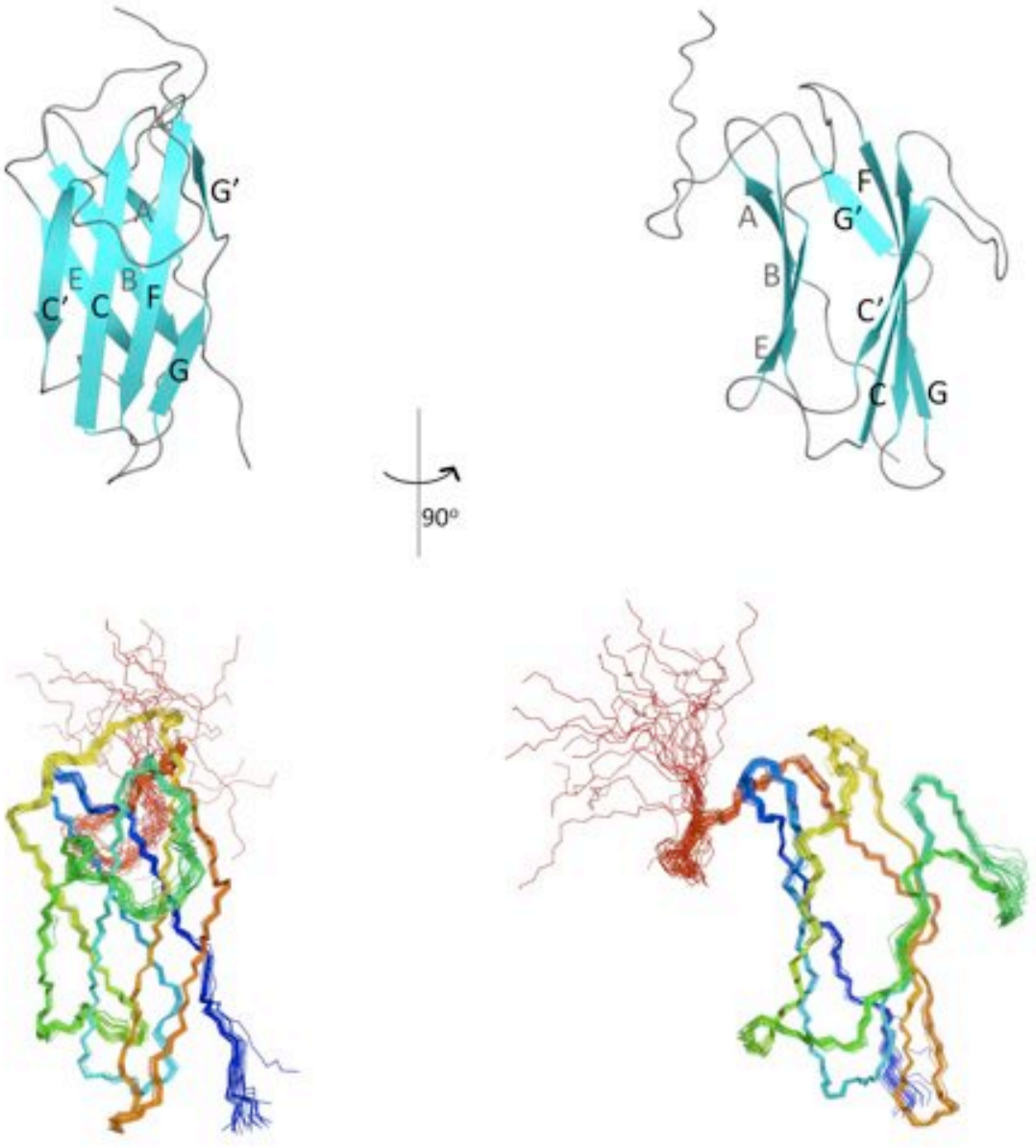


Figure 2-6 Solution structure of N-βGRP. Ribbon representation of the lowest-energy structure (*upper*); Twenty lowest-energy structure ensemble (*lower*). The backbones are painted with blue for the N-terminal and red for the C-terminal. The structures were drawn using PyMOL (DeLano, 2002).

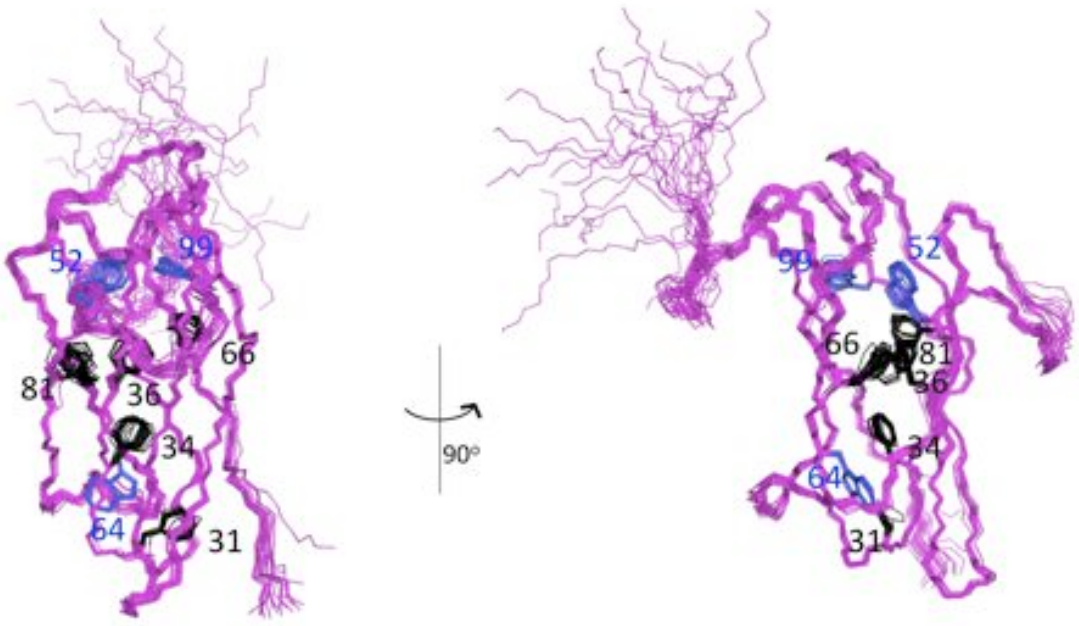


Figure 2-7 Hydrophobic core of N-βGRP. Phe31, Phe34, Phe36, Phe66 and Phe81 are shown in black and Trp52, Trp64, and Trp99 in blue.

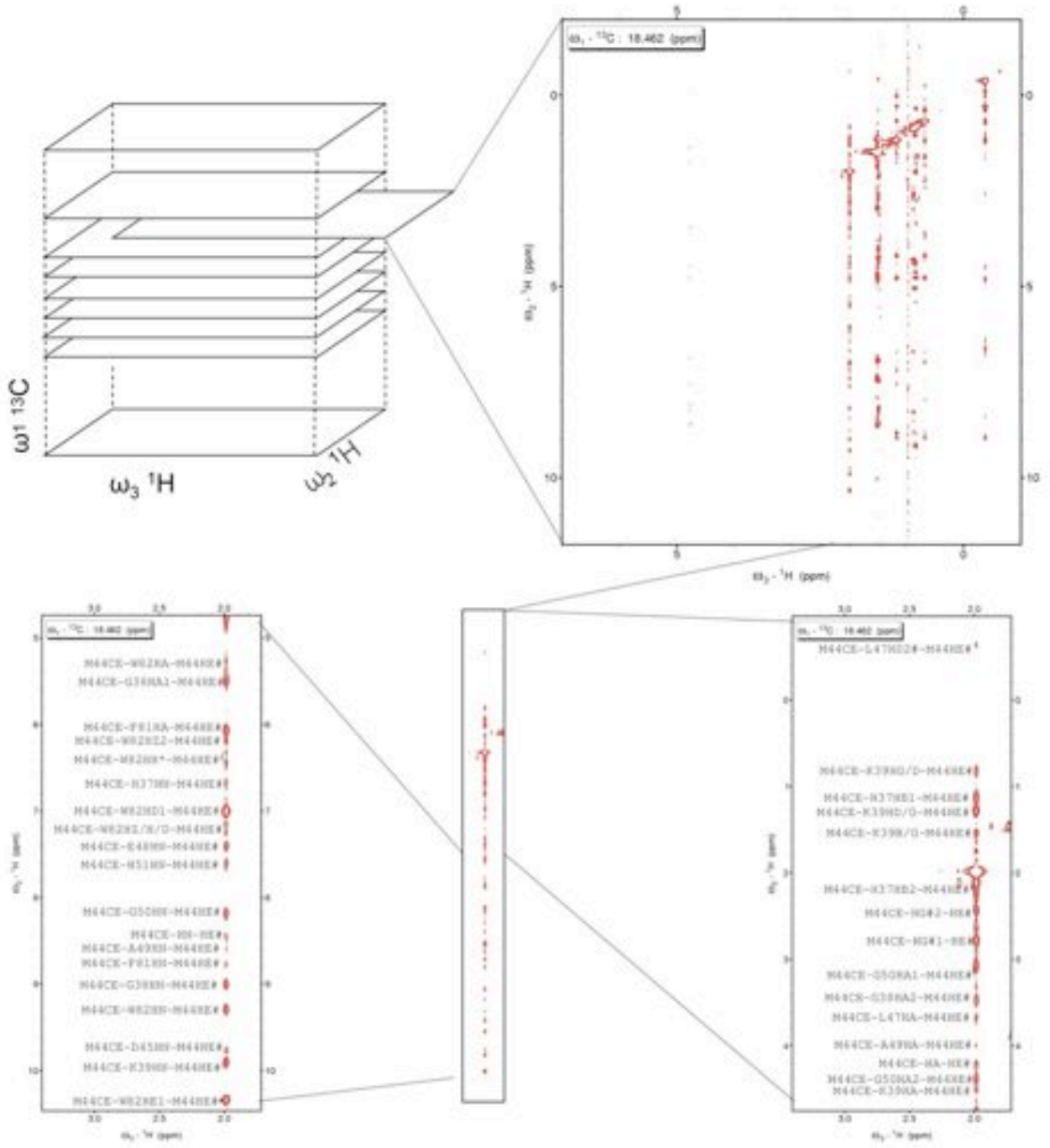


Figure 2-8 An example of the 3D ^{13}C-edited NOESY spectrum. A slice containing H_ε of Met44 is shown, and the strip plot is zoomed to show all the NOE connectivities observed for this hydrogen atom.

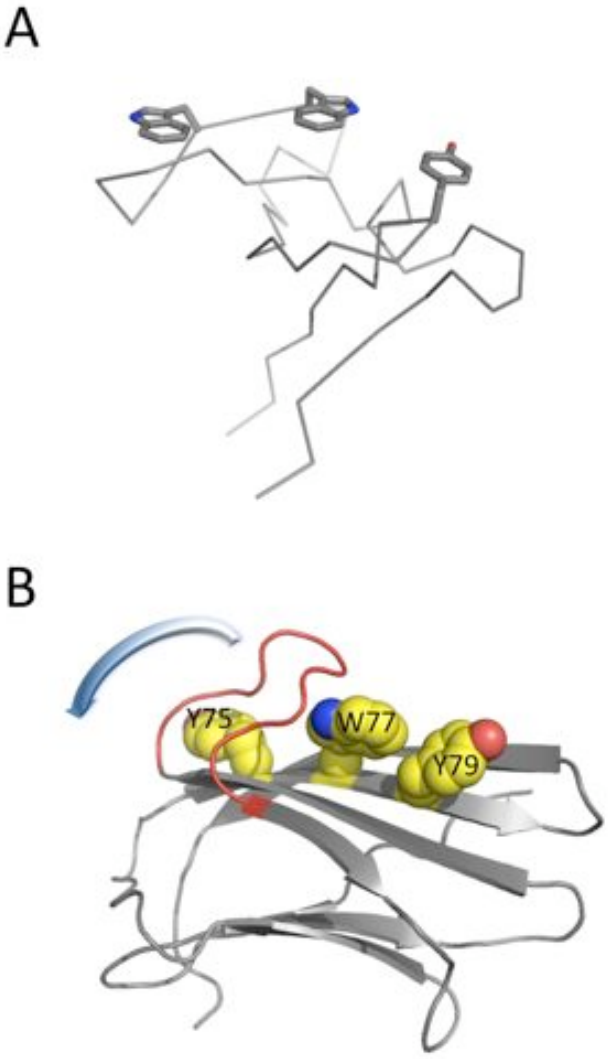


Figure 2-9 Aromatic platform for carbohydrate binding. (A) The 'planar' platform of a Type A CBM (CBM10; Boraston et al., 2004); (B) Conformational change proposed for loop C-C' to allow glucan-binding (Mishima et al., 2009).

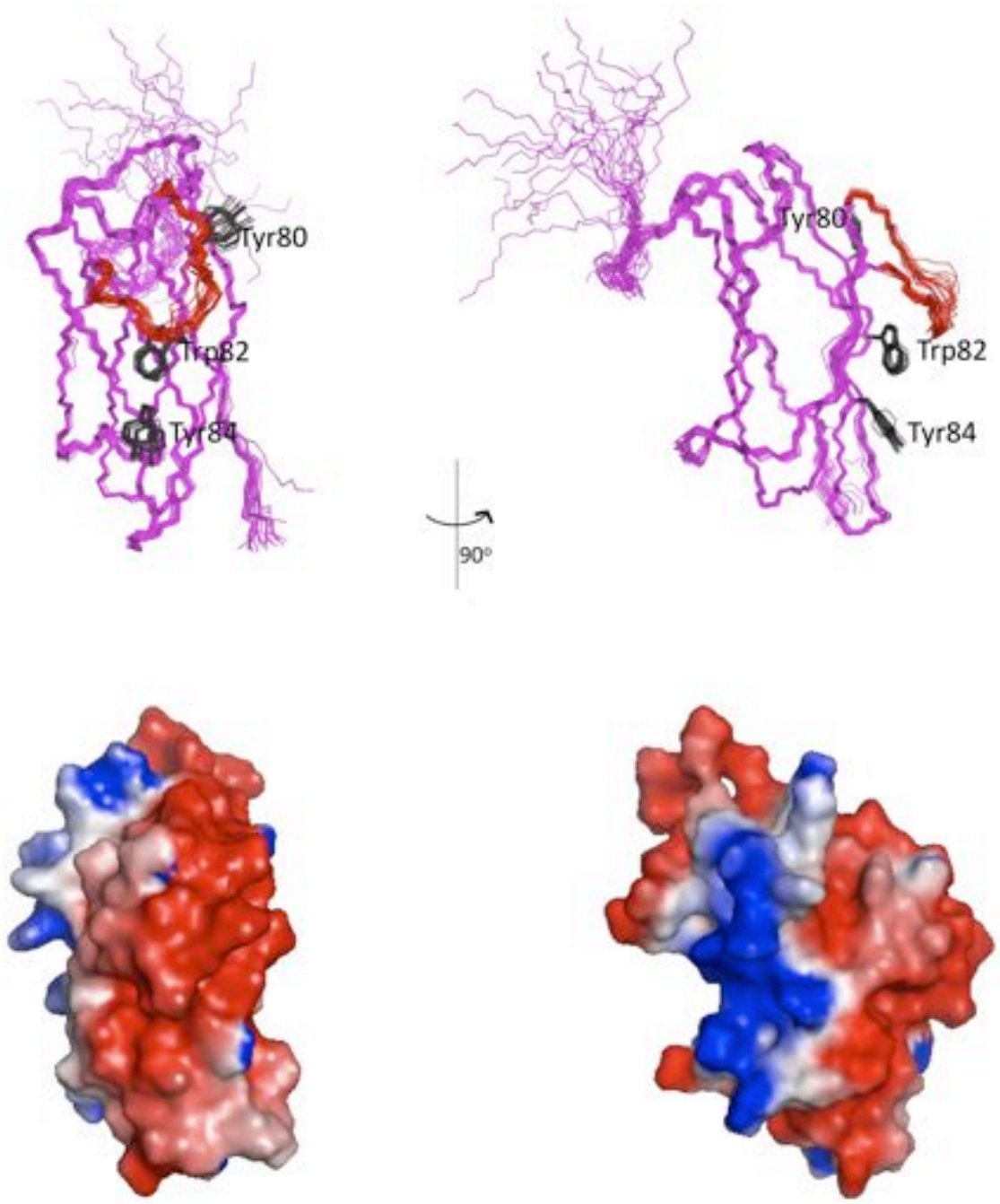


Figure 2-10 Surface characteristics of N-βGRP. Tyr80, Trp82, and Tyr84 are shown in black (*upper*), and the acidic loop in red (*lower*). The surface potential of the lowest-energy structure is colored red (negative) or blue (positive).

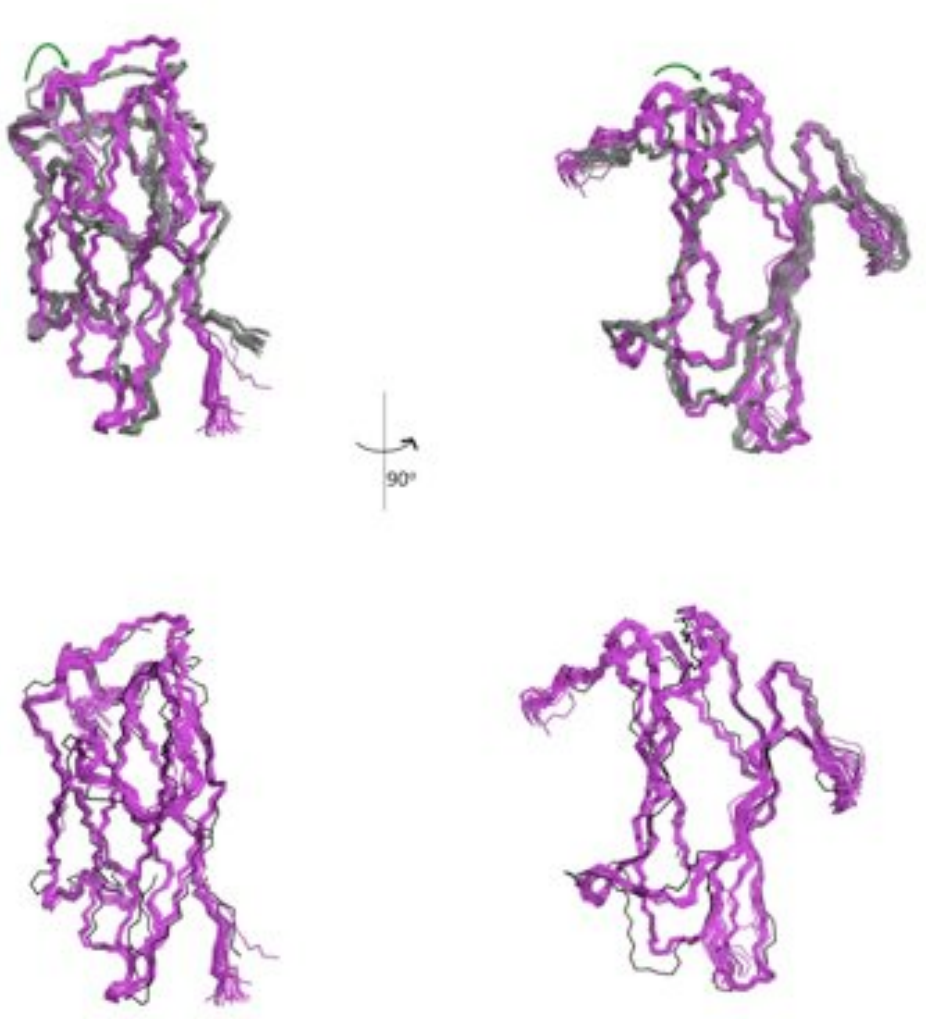


Figure 2-11 Comparison of N-βGRP with CBM39 of GNBP3 from *B. mori* and *D. melanogaster*. N-βGRP, *B. mori* CBM39 (*upper*) and *D. melanogaster* CBM39 (*lower*) are colored in purple, gray (Takahasi et al., 2009) and black (Mishima et al., 2009), respectively. For clarity, residues of N-βGRP after Asp108 are not shown. The different conformation of the loop for the *B.* mori protein (pointed with a green arrow) arises out of the wrong assignment of a *trans* conformation to a Pro (Takahasi et al., 2009).

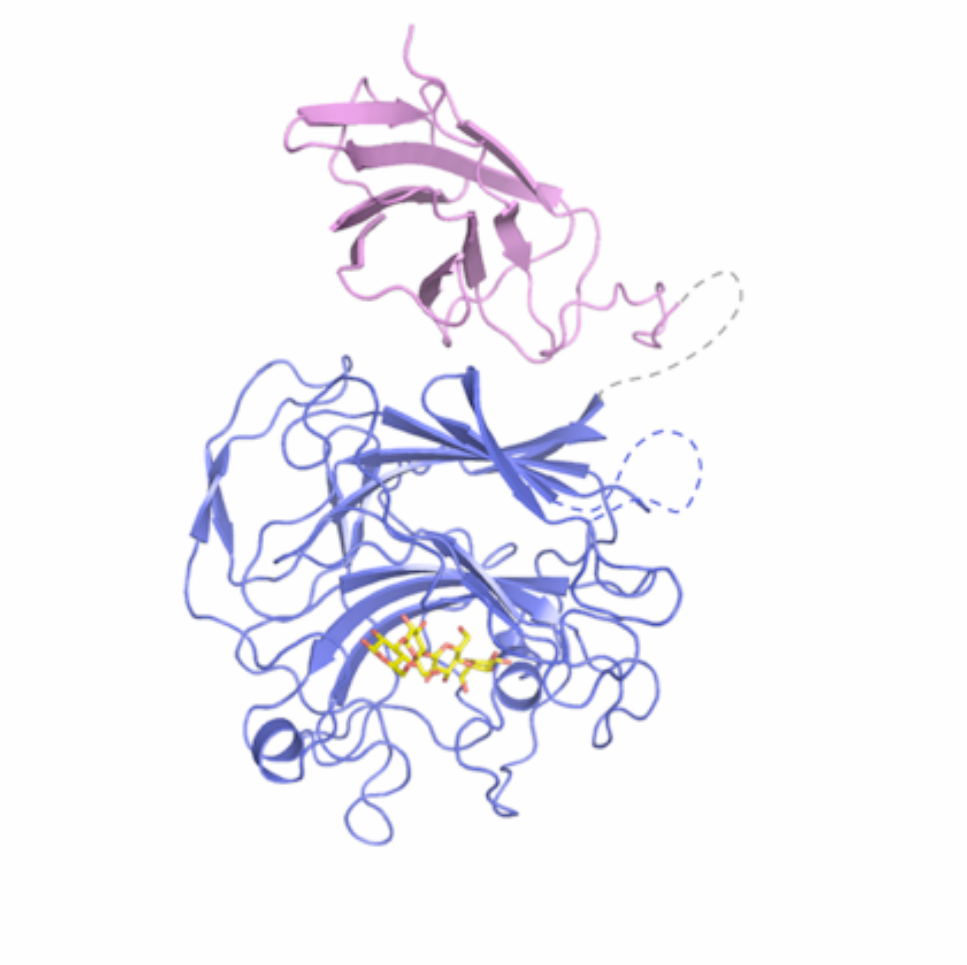


Figure 2-12 A model of βGRP. The N-domain is shown in purple. The C-domain is shown in blue. Dashed lines indicate regions that could not be homology-modeled. The template for the C-domain is a bacterial glucanase (Fibriansah et al., 2007). A laminaritetraose was manually docked into the glucan-binding site, and was shown as stick in yellow with oxygen atoms colored red using PyMOL (DeLano, 2002).

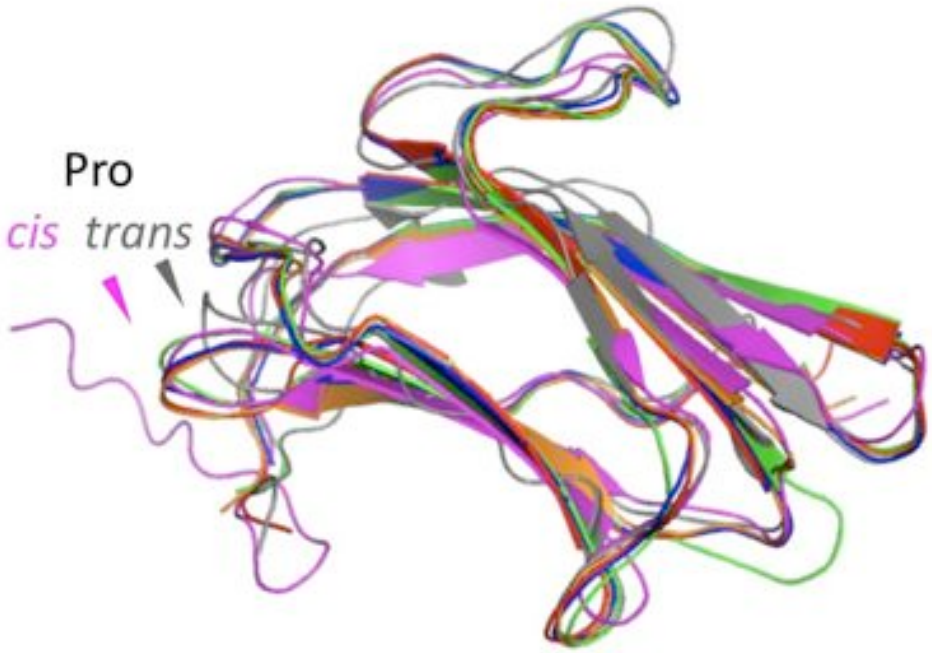


Figure 2-13 Superimposition of ribbon structures of N-βGRP from several insects: *P. interpunctella* - purple, 2KHA (NMR, this work); *B. mori* - gray, 2RQE [NMR; (Takahashi et al., 2009)]; *D. melanogaster* - green, 3IE4 [X-ray; (Mishima et al., 2009)]; *B. mori* (laminarihexaose-bound) – orange, 3AQX [X-ray; (Kanagawa et al., 2011)]; *P. interpunctella* - blue, 3AQY [X-ray; (Kanagawa et al., 2011)]; *P. interpunctella* (laminarihexaose-bound) - red, 3AQZ [X-ray; (Kanagawa et al., 2011)]; The NMR structure of *P. interpunctella* N-βGRP reported here differs from that of the *B. mori* protein (Takahashi et al., 2009) in the configuration of a Pro, as indicated.

Table 2-1 Resonance assignments of N-βGRP.

Residue	N (HN)	C	C^{α} (Hα)	C^{β} (Hβ)	Other Carbons (Hydrogens)
Q6		175.068	55.894 (4.455)	29.831 (2.120, 2.120)	C^{γ}, 33.777 (2.418, 2.418)
Y7	124.383 (8.034)	173.318	59.456 (4.123)	38.787 (1.796, 2.632)	$C^{\delta 1}$, (6.396); $C^{\delta 2}$, (6.396); $C^{\varepsilon 1}$, (6.536); $C^{\varepsilon 2}$, (6.536)
V8	127.390 (6.995)	174.014	60.851 (3.829)	33.172 (1.593)	$C^{\gamma 1}$, 20.844 (0.774); $C^{\gamma 2}$, 20.844 (0.774)
V9	126.277 (8.067)		(4.522)	32.687 (1.767)	$C^{\gamma 1}$, 21.244 (0.959); $C^{\gamma 2}$, 23.587 (0.834)
P10		175.074	62.826 (4.327)	32.293 (2.402, 1.830)	C^{γ}, 27.574 (2.239, 2.239); C^{δ}, 50.550 (3.884, 3.592)
S11	115.867 (8.123)	174.662	58.729 (4.174)	63.467 (3.792, 4.043)	
A12	127.434 (8.823)	176.680	52.856 (4.200)	18.275 (1.162)	
K13	123.228 (8.959)	175.079	54.944 (4.391)	33.511 (1.628, 1.662)	C^{γ}, 25.269; C^{δ}, 29.447; C^{ε}, 41.852
L14	132.849 (8.315)	174.127	56.384 (4.495)	39.021 (-0.833, -1.182)	C^{γ}, 30.325 (0.646); $C^{\delta 1}$, 25.396 (0.020); $C^{\delta 2}$, 26.527 (-0.108)
E15	117.118 (8.411)	173.193	54.925 (4.379)	35.969 (1.639, 1.710)	
A16	127.968 (9.500)	174.798	50.493 (4.815)	20.036 (1.409)	
I17	124.641 (7.725)	175.229	58.392 (4.621)	36.821 (1.573)	$C^{\gamma 1}$, 27.261 (1.065, 1.065); $C^{\gamma 2}$, 17.744 (0.778); $C^{\delta 1}$, 10.613 (0.344)
Y18	126.868 (8.855)		54.292 (4.597)	41.131 (2.541, 2.614)	$C^{\delta 1}$, (6.705); $C^{\delta 2}$, (6.705)
P19		175.439	63.775 (3.315)	33.854 (1.315, 1.697)	C^{γ}, 24.355 (1.642, 1.642); C^{δ}, 49.703 (3.200, 3.599)
R20	119.212 (7.113)		55.528 (4.814)	33.643 (1.551, 1.715)	C^{γ}, 26.176 (1.396, 1.476); C^{δ}, 43.544 (3.165, 3.165)
G21	107.747 (8.542)	172.725	44.654 (4.470, 4.034)		
L22	121.971 (8.306)	174.516	53.380 (5.113)	46.453 (1.229, 1.113)	C^{γ}, 27.697 (0.722); $C^{\delta 1}$, 21.938 (0.387); $C^{\delta 2}$, 24.699 (-0.428)
R23	130.606 (8.903)	172.507	55.022 (4.997)	35.399 (1.715, 1.715)	C^{γ}, 25.896; C^{δ}, 44.150
V24	125.641 (8.983)	174.434	60.309 (4.827)	35.078 (1.195)	$C^{\gamma 1}$, 20.272 (0.287); $C^{\gamma 2}$, 18.782 (-0.382)
S25	118.386 (8.921)	172.862	56.996 (5.872)	69.091 (3.823, 3.600)	
I26	109.563 (7.966)		58.127 (4.770)	41.558 (1.616)	$C^{\gamma 1}$, 24.022 (1.217, -0.257); $C^{\gamma 2}$, 18.956 (0.673); $C^{\delta 1}$, 16.698 (0.399)
P27		175.389	61.999 (4.723)	31.625 (1.398, 1.398)	C^{γ}, 26.531 (2.200, 2.200); C^{δ}, 50.392 (3.724, 3.601)
D28	114.288 (8.197)		55.383 (4.630)	41.786 (2.263, 2.521)	
D29	(5.360)	173.613	52.729 (4.618)	39.930 (2.103, 2.381)	
G30	105.789 (7.817)	176.275	46.518 (3.754, 3.849)		

F31	118.957 (8.363)	176.529	55.187 (5.465)	39.078 (2.628, 3.022)	$C^{\delta1}$, (6.793); $C^{\delta2}$, (6.793); $C^{\varepsilon1}$, (7.565); $C^{\varepsilon2}$, (7.565)
S32	108.686 (8.960)	172.910	58.218 (5.041)	64.661 (3.912, 3.912)	
L33	123.139 (6.796)	172.917	56.221 (4.279)	45.835 (1.084, 1.833)	C^{γ}, 27.468 (1.509); $C^{\delta1}$, 24.704 (0.941); $C^{\delta2}$, 26.043 (0.832)
F34	123.033 (7.917)	171.518	55.317 (5.327)	43.507 (2.716, 2.425)	$C^{\delta1}$, (6.982); $C^{\delta2}$, (6.982); $C^{\varepsilon1}$, (6.970); $C^{\varepsilon2}$, (6.970); C^{ζ}, (6.718)
A35	130.864 (9.484)	174.531	50.003 (5.107)	23.121 (1.225)	
F36	117.139 (7.243)	171.513	55.687 (4.167)	42.334 (1.681, 2.162)	$C^{\delta1}$, (6.684); $C^{\delta2}$, (6.684); $C^{\varepsilon1}$, (6.439); $C^{\varepsilon2}$, (6.439)
H37	122.772 (6.681)	173.226	52.936 (4.865)	28.327 (2.095, 1.117)	
G38	103.788 (9.009)	172.809	44.961 (3.481, 5.494)		
K39	122.616 (9.902)	173.882	56.083 (4.499)	38.420 (1.730, 1.951)	C^{γ}, 26.764; C^{δ}, 30.019
L40	126.614 (9.736)	176.641	54.303 (5.157)	43.508 (2.079, 1.631)	C^{γ}, 27.253 (1.651); $C^{\delta1}$, 23.775 (1.135); $C^{\delta2}$, 26.245 (1.058)
N41	120.854 (10.387)	172.937	55.868 (4.723)	37.328 (3.081, 3.293)	$N^{\delta2}$, (7.549)
E42	116.098 (6.985)	172.495	55.886 (4.357)	33.555 (2.017, 1.690)	C^{γ}, 36.902 (2.166, 2.371)
E43	121.095 (8.173)	178.148	56.993 (2.909)	29.949 (1.465, 1.465)	C^{γ}, 36.582 (1.735, 1.604)
M44	120.852 (8.448)	175.521	55.348 (4.216)	33.760 (2.227, 2.227)	C^{γ}, 33.486 (2.434, 2.785); C^{ε}, 18.772 (1.987)
D45	126.565 (9.772)	175.250	52.122 (4.696)	40.023 (2.339, 2.679)	
G46	108.690 (7.962)	174.923	44.916 (3.745, 3.489)		
L47	125.554 (8.168)	177.046	52.421 (3.675)	38.046 (1.122, 0.258)	C^{γ}, (0.902); $C^{\delta1}$, 24.699 (0.171); $C^{\delta2}$, 20.061 (-0.632)
E48	120.439 (7.412)	176.364	55.895 (4.235)	31.286 (2.128, 2.128)	C^{γ}, 37.037 (2.117, 2.117)
A49	121.821 (8.586)	179.717	55.547 (3.995)	18.809 (1.480)	
G50	96.699 (8.174)	173.450	43.564 (4.380, 3.087)		
H51	118.548 (7.604)	177.308	57.132 (4.450)	31.782 (2.606, 2.723)	
W52	116.743 (8.218)	173.823	58.294 (4.773)	30.342 (2.927, 2.927)	$C^{\varepsilon3}$, (7.232); $C^{\eta2}$, (7.130); $C^{\zeta2}$, (6.949); $C^{\zeta3}$, (6.959); $N^{\varepsilon1}$, (11.820)
A53	123.984 (8.233)	176.218	51.265 (4.785)	20.006 (1.172)	
R54	122.392 (7.691)	173.336	55.571 (4.436)	34.800 (1.637, 1.872)	C^{γ}, 27.194 (1.857, 1.790); C^{δ}, 42.992 (3.472, 3.472)
D55	124.551 (8.403)	174.806	53.587 (4.874)	41.262 (2.302, 2.655)	
I56	127.664 (10.041)	176.512	59.595 (4.684)	37.026 (2.961)	$C^{\gamma1}$, 27.444 (1.859, 1.859); $C^{\gamma2}$, 19.109 (1.512); $C^{\delta1}$, 11.563 (1.109)
T57	111.831 (8.658)	175.016	62.846 (4.881)	69.852 (4.700)	$C^{\gamma2}$, 22.165 (1.307)
K58	120.768 (7.440)		53.479 (4.636)	35.131 (1.638, 1.720)	

P59			61.332 (3.820)	30.626 (-0.249, 1.024)	C^{γ}, 24.606 (1.449, 1.449); C^{δ}, 49.711 (3.287, 2.717)
K60	117.564 (8.626)		56.410 (4.286)	35.357 (1.648, 1.701)	
E61	123.986 (9.436)	176.247	57.031 (3.800)	27.021 (1.945, 1.945)	C^{γ}, 36.795 (2.164, 2.164)
G62	100.933 (8.828)	174.381	45.527 (3.622, 4.091)		
R63	117.783 (7.683)	173.643	54.973 (5.098)	34.880 (1.672, 1.410)	C^{γ}, 27.820; C^{δ}, 42.951
W64	123.260 (8.964)	175.458	56.085 (4.506)	29.662 (2.918, 2.557)	$C^{\delta 1}$, (6.792); $C^{\varepsilon 3}$, (7.337); $C^{\eta 2}$, (6.815); $C^{\zeta 2}$, (7.359); $C^{\zeta 3}$, (6.902); $N^{\varepsilon 1}$, (11.634)
T65	116.703 (8.434)	172.878	62.106 (4.995)	70.939 (3.717)	$C^{\gamma 2}$, 22.331 (0.938)
F66	128.279 (9.801)	172.863	56.053 (4.770)	41.387 (2.586, 2.863)	$C^{\delta 1}$, (6.629); $C^{\delta 2}$, (6.629); $C^{\varepsilon 1}$, (6.801); $C^{\varepsilon 2}$, (6.801)
R67	125.454 (7.734)	174.287	54.231 (4.613)	33.975	C^{γ}, 28.237 (1.404, 1.162); C^{δ}, 43.248 (2.905, 3.056)
D68	118.650 (7.892)		53.158 (4.766)	42.524 (2.622, 2.622)	
R69*		175.485	58.123 (4.211)	30.040 (1.886, 1.886)	C^{γ}, 28.346 (1.587, 1.587); C^{δ}, 43.299 (3.127, 3.127)
N70	111.988 (8.297)	175.759	53.053 (4.773)	40.470 (2.862, 2.787)	$N^{\delta 2}$, 112.973 (7.852, 6.966)
A71	122.717 (6.930)	175.241	53.662 (4.253)	19.290 (1.491)	
K72	126.382 (8.596)	175.353	55.416 (4.793)	32.465 (1.632, 1.982)	C^{γ}, 24.862; C^{δ}, 29.335; C^{ε}, 42.261
L73	128.033 (8.834)	175.083	53.789 (4.923)	44.260 (2.096, 1.591)	C^{γ}, 27.623 (1.141); $C^{\delta 1}$, 26.572 (1.114); $C^{\delta 2}$, 24.173 (1.164)
K74	119.250 (9.140)	175.787	53.909 (4.926)	35.916 (1.790, 1.790)	C^{γ}, (1.344, 1.344)
L75	119.569 (8.592)	178.715	56.689 (3.626)	41.278 (1.353, 1.680)	C^{γ}, 26.892 (1.782); $C^{\delta 1}$, 25.324 (0.911); $C^{\delta 2}$, 23.208 (0.505)
G76	110.606 (8.930)	174.789	44.880 (3.290, 4.327)		
D77	121.518 (8.104)	174.552	55.952 (4.859)	42.265 (2.509, 2.986)	
K78	117.687 (8.572)	175.297	54.428 (5.307)	36.509 (1.305, 1.618)	
I79	121.836 (9.352)	174.825	60.004 (5.371)	40.534 (2.143)	$C^{\gamma 1}$, 28.951 (1.905, 1.501); $C^{\gamma 2}$, 17.565 (1.137); $C^{\delta 1}$, 14.873 (0.982)
Y80	128.716 (8.968)	176.272	56.753 (5.402)	39.808 (3.308, 3.055)	$C^{\delta 1}$, (7.167); $C^{\delta 2}$, (7.167); $C^{\varepsilon 1}$, (6.652); $C^{\varepsilon 2}$, (6.652)
F81	118.636 (8.780)	173.462	56.665 (6.062)	45.076 (2.766, 3.043)	$C^{\delta 1}$, (6.951); $C^{\delta 2}$, (6.951); $C^{\varepsilon 1}$, (7.207); $C^{\varepsilon 2}$, (7.207)
W82	117.068 (9.288)	174.321	56.251 (5.264)	31.716 (3.745, 3.449)	$C^{\delta 1}$, 49.861 (7.002); $C^{\zeta 2}$, 38.930 (6.182); $C^{\zeta 3}$, (6.585); $N^{\varepsilon 1}$, 130.130 (10.334)
T83	107.308 (9.017)	173.021	58.487 (5.309)	71.805 (4.540)	$C^{\gamma 2}$, 22.321 (1.506)
Y84	119.113 (8.528)	173.891	55.711 (5.772)	43.830 (3.191, 2.832)	$C^{\delta 1}$, (6.966); $C^{\delta 2}$, (6.966); $C^{\varepsilon 1}$, (6.591); $C^{\varepsilon 2}$, (6.591)
V85	119.047 (8.633)	174.128	58.050 (4.762)	35.993 (1.181)	$C^{\gamma 1}$, 20.412 (0.489); $C^{\gamma 2}$, 21.467 (0.103)

I86	117.042 (6.708)	175.275	59.457 (4.707)	38.501 (1.449)	$C^{\gamma 1}$, 27.241 (0.591, 1.277); $C^{\gamma 2}$, 17.182 (0.604); $C^{\delta 1}$, 13.153 (0.777)
K87	129.587 (9.427)	175.461	55.350 (4.782)	36.274 (1.888, 1.430)	C^{γ}, 25.002; C^{δ}, 28.775 (1.157, 0.752); C^{ϵ}, 42.279
D88	131.152 (9.926)	175.699	55.628 (4.282)	39.672 (2.929, 2.712)	
G89	98.898 (8.153)	173.176	45.319 (4.126, 3.437)		
L90	122.773 (7.620)	174.287	53.481 (4.567)	44.170 (1.749, 1.197)	C^{γ}, 27.141 (1.535); $C^{\delta 1}$, 25.285 (0.966); $C^{\delta 2}$, 23.360 (0.865)
G91	105.142 (8.114)	173.264	44.500 (3.340, 4.746)		
Y92	121.728 (8.715)	173.631	57.009 (4.755)	43.387 (2.299, 3.113)	$C^{\delta 1}$, (7.013); $C^{\delta 2}$, (7.013); $C^{\epsilon 1}$, (6.785); $C^{\epsilon 2}$, (6.785)
R93	115.798 (8.755)	178.571	54.874 (5.857)	35.218 (2.048, 1.696)	C^{γ}, 27.322; C^{δ}, 43.464 (3.136, 2.853)
Q94	122.932 (7.897)	174.598	56.593 (4.620)	27.938 (2.760, 2.760)	C^{γ}, 33.056 (2.052, 2.613); $N^{\epsilon 2}$, 107.011 (7.689, 7.554)
D95	124.149 (8.499)	176.907	53.465 (5.220)	44.135 (2.548, 2.406)	
N96	111.794	174.963	54.435 (4.535)	37.108 (2.807, 2.974)	$N^{\delta 2}$, 112.279 (7.698, 6.944)
G97	104.531 (8.065)	175.018	45.224 (1.357, 3.509)		
E98	126.981 (8.425)	174.557	56.170 (5.164)	33.366 (1.945, 1.945)	C^{γ}, 36.250 (2.025, 2.025); C^{δ}, 127.341
W99	129.593 (9.132)	173.552	57.506 (5.028)	32.975 (3.072, 3.442)	$C^{\delta 1}$, (7.514); $C^{\eta 2}$, (6.766); $C^{\zeta 2}$, (7.319); $C^{\zeta 3}$, (6.514); $N^{\epsilon 1}$, 130.780 (9.865)
T100	122.587 (7.423)	172.298	60.576 (4.791)	70.410 (3.511)	$C^{\gamma 2}$, 20.822 (0.825)
V101	125.112 (8.503)	176.594	63.027 (2.828)	30.694 (2.025)	$C^{\gamma 1}$, 21.279 (0.125); $C^{\gamma 2}$, 22.373 (0.584)
T102	120.476 (8.804)	173.685	61.872 (4.296)	70.173	$C^{\gamma 2}$, (0.933)
E103	118.408 (7.763)	172.862	55.243 (4.427)	31.207 (1.984, 1.984)	C^{γ}, 34.010 (1.944, 2.178)
F104	117.961 (8.775)	176.308	56.602 (5.414)	42.894 (2.685, 2.971)	$C^{\delta 1}$, (7.066); $C^{\delta 2}$, (7.066); $C^{\epsilon 1}$, (7.260); $C^{\epsilon 2}$, (7.260); C^{ζ}, (7.124)
V105	112.559 (9.176)	175.677	58.929 (5.053)	35.567 (2.012)	$C^{\gamma 1}$, 21.549 (0.792); $C^{\gamma 2}$, 18.354 (0.824)
N106	115.945 (9.113)	176.883	52.833 (5.034)	40.192 (2.774, 3.024)	$N^{\delta 2}$, 115.870 (7.409, 6.566)
E107	120.609 (9.088)	178.327	59.810 (3.905)	28.943 (2.116, 1.971)	
D108	114.417 (7.989)	177.269	53.983 (4.415)	39.864 (2.621, 3.068)	
G109	105.087 (8.443)	174.494	44.930 (3.352, 4.332)		
T110	112.459 (8.133)		60.507 (4.521)	69.057 (4.340)	$C^{\gamma 2}$, 22.123 (1.220)
P111		176.630	64.311 (4.398)	32.094	C^{γ}, 28.021; C^{δ}, (3.890, 3.595)
A112	125.149 (8.284)	176.910	51.517 (4.300)	19.469 (0.871)	
D113	121.045 (8.509)	176.722	53.437 (4.646)	40.721 (2.764, 2.534)	
T114	113.563 (8.140)	174.976	61.712 (3.935)	68.638 (4.245)	$C^{\gamma 2}$, 21.203 (0.769)

S115	116.873 (8.323)	174.562	59.416 (4.333)	63.777 (3.890, 3.890)	
L116	122.573 (7.847)	177.067	54.793 (4.355)	42.340	
E117	122.369 (8.188)		54.491 (4.528)	29.680	
P118			63.283 (4.382)	32.106	C^{δ}, (3.787, 3.660)

*The backbone NH of Arg69 could not be assigned, perhaps because of exchange-broadening and/or unusual ^{15}N and ^{1}H chemical shifts in a range outside the spectral width used. The homologous Arg in the *B. mori* protein was not assigned either (Takahasi et al., 2009).

Table 2-2 Statistics for a structural ensemble of 20 lowest-energy structures of N-βGRP.

Total constraints (residues 6-118)	1,584
NOEs	1,408
Intraresidue ($\|i-j\| = 0$)	98
Sequential ($\|i-j\| = 1$)	299
Medium-range ($2 \leq \|i-j\| \leq 4$)	167
Long-range ($\|i-j\| > 4$)	844
Dihedral constraints	106
Hydrogen bonds	35
Lennard-Jones potential energy (kcal/mol)	-192.0 ± 16.9
Number of violations	
NOE > 0.5Å	0
Dihedral > 5º	0
Coordinate precision (Å)	
Residues 6-118	
Backbone	1.62 ± 0.42
All non-hydrogen atoms	1.95 ± 0.37
Residues 7-106	
Backbone	0.51 ± 0.08
All non-hydrogen atoms	1.31 ± 0.11

Table 2-3 CBM relatives of CBM39.

CBM family	PDB code/ chain code	rmsd (Å)	Number of equivalent residues	Functional type (A/B/C)
9	1I82/A	2.7	90	C
21	2DJM/D	3.0	78	A
31	2COV/D	2.4	84	A
34	1BVZ/A	3.0	93	B
34	1J0H/A	2.8	87	B
34	1EA9/C	2.9	86	B

Table 2-4 CBM fold families. The fold classification is based on Boraston et al. (2004), and the update provided by Hashimoto (2006).

Fold family	Fold (sum of families)		CBM families
1	β-sandwich (40)		
	β-jelly roll (28)		2, 3, 4, 6, 11, 15, **16**, 17, 22, 27, 28, 29, 30, 32, 35, 36, **40**, 44, **47**, **51**, **57**, **59**, **60**, **61**, **62**, **63**, **65**, **66**
	Immuoglobulin (12)	I (5)	20, 25, 26, **41**, **58**
		II (7)	9, **21**, 31, 33, 34, ***39***, **48**
2	β-trefoil (2)		13, 42
3	Cysteine knot (2)		1, **43**
4	Unique, WW domain-like (2)		5, 12
5	OB fold (1)		10
6	Hevein fold (1)		18
7	Unique, hevein-like fold (1)		14
*	LysM fold (1)		**50**

* Unassigned

The nineteen novel structures since 2006 (Hashimoto, 2006) are highlighted in bold.

Updated December 20, 2012.

Table 2-5 Structures of CBM39 family members.

PDB code	Method	Experiment Details	PDB Deposition Date	Reference
2KHA	NMR	NMR Restraints, BMRB	2009-03-29	This work
2RQE	NMR	NMR Restraints	2009-04-02	Takahashi et al., 2009
3IE4	X-ray	Structure Factors, 1.45 Å	2009-07-22	Mishima et al., 2009
3AQX	X-ray	Structure Factors, 2.01 Å	2010-11-22	Kanagawa et al., 2011
3AQY	X-ray	Structure Factors, 1.58 Å	2010-11-22	Kanagawa et al., 2011
3AQZ	X-ray	Structure Factors, 2.20 Å	2010-11-22	Kanagawa et al., 2011

Chapter 3 - Structural and Biological Studies of the N-βGRP:β-1,3-Glucan Complex

Introduction

Recognition of PAMP by PRR and the subsequent formation of PAMP:PRR complex trigger innate immune responses (Janeway & Medzhitov, 2002). Biochemical and biophysical characterization of the formation of PAMP:PRR complex will provide insight at the molecular level into recognition mechanisms and specificities for different responses initiated by different recognition complexes. As for insect immunity, melanization is an immediate and effective response (Cerenius et al., 2008; Kanost et al., 2004). A misregulation of the proPO cascade can be indiscriminately harmful to both pathogens and their hosts. Thus, initiation of the proPO cascade must be tightly regulated. Herein we describe our studies at the molecular level relating to the formation of the βGRP:β-1,3-glucan complex.

Three-dimensional structures have been reported for N-βGRP from three insect species, *B. mori* (Takahasi et al., 2009; Kanagawa et al., 2011), *D. melanogaster* (Mishima et al., 2009) and *P. interpunctella* (Kanagawa et al., 2011): all three N-βGRPs adopt a common immunoglobulin-like fold with two sheets forming a β-sandwich (Figure 2-13). However, different models have been proposed for the binding of β-1,3-glucan by N-βGRP. In the model based on the solution NMR structure of *B. mori* N-βGRP, the β-1,3-glucan binding site is located on the concave sheet (strands A, B and E) (Takahasi et al., 2009), while in the models based on the X-ray crystal structure of ligand-free *D. melanogaster* GNBP3 N-domain (Mishima et al., 2009) and that of *P. interpunctella* N-βGRP complexed with laminarihexaose (Kanagawa et al., 2011), the binding site is confined to the convex sheet (strands C, C', F, G and G'). These three models, however, do not provide any clues for understanding the molecular basis of the synergistic activation of the proPO pathway by β-1,3-glucan and N-βGRP.

Characterization of the βGRP:β-1,3-glucan complex is hampered by the presence of heterogeneity in β-1,3-glucan isolated from difference sources. Whereas *in vitro* experiments have established that N-βGRP can bind to curdlan and to the cell wall of bacteria or yeast (Ochiai et al., 2000; Fabrick et al., 2004), the minimal molecular pattern recognized by βGRP

remains unknown. To date, no experimental data are available regarding the three dimensional structure at the atomic level of β-1,3-glucan, and the X-ray fiber diffraction data have been interpreted differently: the native curdlan has a single helical structure, and this native structure can be transformed into a triple-helical structure by heat treatment (Deslandes et al., 1980; Chuah et al., 1983; Okuyama et al., 1991). In this study, we have characterized the binding of laminarihexaose and that of laminarin to N-βGRP by NMR, isothermal titration calorimetry (ITC), analytical ultracentrifugation studies (AUC), site-directed mutagenesis, and prophenoloxidase activity measurements. Our results demonstrate, for the first time, ligand-induced self-association of N-βGRP:β-1,3-glucan complex and its biological significance.

Materials and Methods

Carbohydrates

Laminarin from *Laminaria digitata* was purchased from Sigma Aldrich (L9634, St. Louis, MO), and its β-1,3 to β-1,6 cross-link number ratio was 7 (Hrmova & Fincher, 1993). Laminarihexaose and curdlan were from Megazyme (Wicklow, Ireland). A much more branched laminarin from *Eisenia bicyclis* with a β-1,3 to β-1,6 cross-link number ratio of 3 (Handa & Nisizawa, 1961) was purchased from Tokyo Chemical Industry (Tokyo, Japan).

Site-directed Mutagenesis

N-βGRP containing mutation D45A or D45K was prepared according to the instructions of QuikChange mutagenesis kit (Stratagene Agilent Technologies, Santa Clara, CA). Protein expression and purification were carried out, as described for the wild type protein. In each case, the mutation was confirmed by DNA sequencing and mass spectrometry of the purified protein.

Analytical Ultracentrifugation

Sedimentation velocity experiments were conducted with an Optima XL-I ultracentrifuge (Beckman Coulter, Inc. Brea, CA) using an An-60 Ti rotor at 20°C with 50 mM Tris-HCl (pH 7.3) buffer containing 50 mM NaCl (Hiromasa et al., 2004). Sedimentation was monitored by absorbance or interference optics using double-sector aluminum cells with a final loading of 400 µl per sector. Sedimentation was performed at 49,000 rpm with scans made at 5 min intervals. Data were analyzed using DCDT+ software version 1.16 (www.jphilo.mailway.com). Sedimentation coefficients were calculated using $g(s^*)$ and dc/dt fitting functions in DCDT+

software. Buffer density and viscosity were calculated by SEDNTERP version 1.08 (www.jphilo.mailway.com). Bovine serum albumin (BSA) was used as a standard to account for effects of changes in solution density and viscosity when high concentrations of laminarin or laminarihexaose were used. The partial specific volume of a protein was calculated from its amino acid composition using SEDNTERP (0.7306 ml/g for N-βGRP at 20°C). The partial specific volume used for laminarin was 0.622 ml/g (Perkins et al., 1981).

Isothermal Titration Calorimetry

ITC measurements were carried out using MCS-ITC system (MicroCal, Northampton, MA) at 30°C (Wiseman et al., 1989). Recombinant N-βGRP and laminarin solutions were dialyzed overnight at 4°C against 50 mM Tris-HCl (pH 7.5) buffer containing 50 mM NaCl. Laminarihexaose was directly dissolved in the buffer. All solutions were degassed before use. A typical experiment consisted of 20 injections of 10 µl laminarin or laminarihexaose solution (1.67 mM) into 1.38 ml protein solution. The heat of dilution of laminarin or laminarihexaose solution upon injection into the buffer solution was subtracted from the experimental titration data. Baseline corrections and integration of the calorimeter signals were performed using the software, Origin (MicroCal/GE Healthcare, Piscataway, NJ).

NMR Titration

2D ^{1}H-^{15}N HSQC NMR spectra were collected of 0.5-1.0 mM ^{15}N-labeled N-βGRP in 20 mM sodium phosphate buffer (pH 6.5) at 25°C as a function of added ligand concentration.. Laminarihexaose or laminarin solutions (3 mM and 1mM, respectively) were prepared in the same buffer and then added stepwise to the protein sample (0.5 mM). The chemical shift difference between ligand-free and ligand-saturated protein was calculated for each backbone NH group as $[(\Delta H)^2+(\Delta N/5)^2]^{1/2}$.

Curdlan Pull-down Assay

The procedure described by Fabrick et al. (2004) was employed. Briefly, for each assay, 20 µg of purified protein (N-βGRP wide type, D45A, or D45K, respectively) was incubated with 0.5 mg of curdlan for 10 min. The protein-curdlan mixture was centrifuged at 10,000 × g for 5 min, and the supernatant corresponding to the unbound fraction was saved. The bound protein was eluted from curdlan by heating at 95°C for 5 min in SDS sample buffer. Equal volumes of

purified, unbound and bound proteins were analyzed by SDS-PAGE with Coomassie Blue staining.

Activation of the ProPO Pathway

A method as described by Ma and Kanost (2000) and Fabrick et al. (2004) was used, incorporating the modification recently developed by Laughton and Siva-Jothy (2011) to follow proPO activation in the absence or presence of laminarin. Briefly, 10 µl of recombinant proteins (0.4 mg/ml) was incubated with 5 µl of buffer or laminarin (10 mg/ml) and mixed with 5 µl of *M. sexta* plasma. The volume of each sample well was brought up to 130 µl with sodium phosphate buffer (pH 6.8). After incubation for 15 min at room temperature, 20 µl of 30 mM dopamine hydrochloride was added and phenoloxidase activity was determined by measuring absorbance at 490 nm. The phenoloxidase activity is presented as the change in milli-absorbance unit per minute. Statistical analysis was performed using Prism 5 (GraphPad Software).

Results and Discussion

Homogeneity of N-βGRP

We tested whether oligomerization of N-βGRP occurred in solution by AUC. The sedimentation analysis of N-βGRP exhibited a homogeneous species with an $S_{20,w}$ value of 1.90 S corresponding to a weight average molecular weight of 14,500 by *dc/dt* analysis--a value in good agreement with the calculated mass of 14,601 Da. N-βGRP sedimentation pattern remained the same for a 1.7 mM solution (Figure 3-1), thus indicating that N-βGRP existed as a monomer in solution at concentrations used for NMR experiments.

Mapping of β-1,3-Glucan Binding Site

Effect of β-1,3-glucan binding to N-βGRP was characterized by titrating laminarihexaose, a β-1,3-linked glucose hexamer, into a solution of N-βGRP and monitoring the NMR cross-peaks of the protein backbone ^{15}N-^{1}H groups (Figure 3-2). The perturbed residues are mainly localized on the convex sheet (Figure 3-3). This observation is consistent with the hexasaccharide-binding site identified in the crystal structure of N-βGRP: laminarihexaose complex (Kanagawa et al., 2011). The crystal structure shows the binding site interacting with three laminarihexaose molecules, which led the authors to propose that N-βGRP binds to a triple helical structure of laminarin (Kanagawa et al., 2011). However, to date no structural data are

available for laminarin which is a branched polysaccharide. The X-ray fiber diffraction data collected for the linear β-1,3 polysaccharide, curdlan, have been interpreted to be indicative of a single-helical or a triple-helical structure, depending upon the hydration state of the polysaccharide (Okuyama et al., 1991; Deslandes et al., 1980). In the presence of water, the diffraction data have been shown to be consistent only with a single-helix structure (Okuyama et al., 1991).

Heterogeneity of laminarin

Laminarin is a water-soluble oligosaccharide containing both β-1,3 and β-1,6 glycosidic bonds, and the number of β-1,6 glycosidic bonds varies depending upon the source. *L. digitata* laminarin used in this work has a β-1,3 / β-1,6 glycosidic bond ratio of 7 (Hrmova & Fincher, 1993). It shows a broad mass distribution (up to 7,505 Da) with major mass spectral peaks in the 3932 - 4580 Da region as obtained by MALDI-TOF mass spectrometry (Figure 3-4), consistent with the work of Barral et al. (2005). Sedimentation velocity analysis of laminarin gave an $S_{20,w}$ value of 1.02 S, which yielded an apparent molecular weight of 4.5 kDa by *dc/dt* analysis. At a high concentration (5mg/ml), laminarin had a similar S value, which suggests that laminarin does not associate/aggregate by itself. Young et al. (2000) determined an average molecular weight of 7,700 Da for laminarin from light-scattering experiments. For the purpose of calculating molarities of laminarin solutions used in our work, we adopted a value of 6 kDa for the molecular weight, as provided by the supplier (Sigma Aldrich).

Interactions between N-βGRP and Laminarin

Titration of laminarin into N-βGRP caused the disappearance of nearly all the cross-peaks in the ^{1}H-^{15}N HSQC spectrum due to their broadening to a noise level (Figure 3-5). Increase in temperature from 25 to 37°C did not improve the signal-to-noise ratio of the NMR spectrum. The NMR sample remained clear with no precipitation. These observations indicate that in the presence of laminarin, N-βGRP forms a higher molecular weight, water-soluble structure. The results also suggest that the binding-site identified for laminarihexaose could interact with the surface structure of a carbohydrate polymer. The few cross-peaks that still remained after addition of laminarin to N-βGRP arise from flexible side-chain NH groups and peptide NH groups of some terminal residues.

Binding of laminarin to N-βGRP was characterized using ITC (Figure 3-6). Because of the heterogeneity in laminarin as well as self-association of the resulting protein:carbohydrate complex, a likely cooperative process, the ITC data, although very similar to those of earlier studies (Kanagawa et al., 2011; Mishima et al., 2009), were not fit to any simple binding equilibrium. Qualitatively, the ITC data provided evidence for the binding of laminarin by N-βGRP, an exothermic process.

Interaction between N-βGRP and laminarin was characterized by sedimentation velocity analysis using absorption optics, which detects the distribution of the protein (Figures 3-7 and3-8): As the concentration of laminarin added to a fixed concentration of N-βGRP (26.2 μM) increased from 8.3 to 24.9 μM, the concentration of free N-βGRP at 1.9 S decreased and a broad peak appeared between 4 and 9 S. With a further increase in laminarin concentration (83.3 μM), nearly all of N-βGRP bound to laminarin and a major peak appeared at 5.5 S. A 20-fold excess of laminarin (1.87 mM) caused a slight decrease in S value (5.1 S) with a shoulder at ~8 S. A $g(s^*)$ fitting analysis yielded an average molecular weight of ~95 kDa for the 5.5 S species, a value that is considerably greater than would be expected for a complex that contains one molecule of N-βGRP and three molecules of laminarin. Laminarin contains at least 30 glucose units, and therefore, each laminarin molecule could bind multiple N-βGRP molecules. However, the sedimentation profile around 5.5 S region changed only in a limited way even after addition of 20-fold excess of laminarin (1.87 mM). This suggests that formation of the 5.5 S species involves not only protein-carbohydrate interactions, but also protein-protein interactions. As N-βGRP or laminarin alone does not undergo self-association in solution, these results indicate that binding of laminarin causes self-association of N-βGRP:laminarin complex. Based on the sedimentation coefficient rule (Schachman, 1959), which equates the ratio of average molecular weights of two proteins to the ratio of their sedimentation coefficients raised to the power of 3/2, the shoulder at ~8 S may be attributed to a higher order complex that is about twice the size of the 5.5 S macro complex.

AUC experiments were similarly performed to monitor the effect of laminarihexaose on N-βGRP (Figure 3-9): When 54 μM N-βGRP was mixed with 1.0 mM laminarihexaose, the $g(s^*)$ *vs* s^* plot of N-βGRP displayed no significant change, consistent with a weak-binding of the ligand. At a higher concentration of the mixture, 234 μM N-βGRP and 14.1 mM

laminarihexaose, a faster sedimenting species appeared in the 3-5 *s** region, and with 1.4 mM N-βGRP and 26 mM laminarihexaose, the sedimentation profile shifted further to a higher *s** region. The *s** value range (3 - 5 S) of the faster sedimenting species falls close to that *s** value of BSA (66.5 kDa, 4.3 S). These results suggest that high concentrations of laminarihexaose can mildly induce self-association of N-βGRP:laminarihexaose complex. It is of interest to note that *D. melanogaster* GNBP3 N-terminal domain does not bind to laminaritetraose (Gottar et al., 2006), laminariheptaose or laminarihexaose (Mishima et al., 2009). In contrast, several glucan-binding modules belonging to families other than CBM39, to which N-βGRP belongs, possess high binding affinities for hexasaccharides (Boraston et al., 2004).

Stability and Stoichiometry of N-βGRP:Laminarin Complex

The stability of the macro-assembly of N-βGRP:laminarin complex was assessed by dilution analysis: AUC experiments (Figure 3-10) with three different concentrations of N-βGRP at a constant N-βGRP:laminarin ratio of 1:22.9 reveal that the *s** profile of N-βGRP:laminarin complex was unchanged at 5 S even after 50-fold dilution (from 34.0 μM to 0.68 μM of the protein). This result indicates that the macro complex does not undergo dissociation at submicromolar concentrations due to strong protein-protein and protein-carbohydrate interactions. Sedimentation velocity experiments were also performed with the more branched *E. bicyclis* laminarin (Figure 3-11): While 83.3 μM *L. digitata* laminarin with a β-1,3 / β-1,6 ratio of 7 was sufficient for nearly complete conversion of N-βGRP (26 μM) into the 5.5 S protein:carbohydrate macro complex, a much higher concentration (0.8 mM) of *E. bicyclis* laminarin with a β-1,3 / β-1,6 ratio of 3 was required for the macro complex formation. We infer that increased β-1,6 branching reduces the carbohydrate-binding strength of N-βGRP and thus the stability of the protein:carbohydrate macro complex.

To determine the stoichiometry of N-βGRP:laminarin complex, we performed sedimentation velocity analysis using interference optics. The sedimentation of both N-βGRP and laminarin could be monitored, as these water-soluble molecules affect the refractive index of the solution. Fringe displacements caused by sedimentation provide a measure of weight concentration of sedimenting species (Hiromasa & Roche, 2003). Sedimentation boundaries were compared between 75 μM N-βGRP and a mixture of 75 μM N-βGRP and 125 μM laminarin (Figure 3-12). Fringe displacement of N-βGRP alone was 2.5 arbitrary units, and that

of the N-βGRP:laminarin complex (the faster species than either N-βGRP or laminarin alone) was 2.9. By using refractive indices of protein [(*dn/dc*) x 103 = 0.186 ml/mg] and dextran [(*dn/dc*) x 103 = 0.151 ml/mg], the weight ratio of N-βGRP:laminarin in the complex was calculated to be approximately 1:0.20, which equals 2:0.92 as a molar ratio. This is consistent with each laminarin (L) molecule binding two N-βGRP (P) molecules to form a P_2L complex. Thus, the 5.5 S species is likely to be a trimer of P_2L (~ 102 kDa; Figure 3-13).

Biological Significance of Self-association of N-βGRP:Laminarin Complex

Formation of a macro assembly of the protein:carbohydrate complex in solution, as characterized in this work, likely allows for ready recruitment of circulating βGRP molecules in the hemolymph to initiate protease cascades that function in defense against invading pathogens. The nature of protein-protein interactions present in the complex may be gleaned from a pseudoquadruplex observed in the crystal structure of N-βGRP:laminarihexaose complex (Kanagawa et al., 2011). In the unit cell N-βGRP molecules pack in a side-by-side fashion due to strong electrostatic attractions such that the tip of the loop connecting strands C and C' of one molecule fits into a cleft formed between strands C' and E of another molecule. This structural arrangement is stabilized predominantly by the hydrogen bonds and the salt bridge formed between Leu47 and Asp45 of one protein molecule and Asp55, Arg54, and Asp68 of another (Figure 3-14).

To test the hypothesis that these electrostatic interactions play a role in the formation of protein:carbohydrate macro complex, we selected Asp45 for mutagenesis because it makes a hydrogen bond with Asp68 and a salt bridge with Arg54 of another protein molecule through its side-chain carboxyl group (Figure 3-14, inset). The sedimentation profiles of N-βGRP and mutants, D45A and D45K, reveal that the complex formed with D45K shifted to 4 S compared to the 5 S complex formed with the wild type protein (Figure 3-15). On the other hand, the protein:laminarin macro complex formed by the D45A mutant does not show any change relative to the wild type complex. These results indicate that in the D45K mutant, electrostatic repulsion between positively charged Lys45 and Arg54 on one hand, and absence of a hydrogen bond between Lys45 and Asp68 on the other, perturb the protein:carbohydrate macro complex formation. Clearly, other protein-protein and protein-carbohydrate interactions contribute significantly to the stability of the macro protein:carbohydrate complex, as the D45A mutant

shows no alteration in its ability to interact with laminarin and make the protein:carbohydrate macro complex, as compared to the wild type protein. It is of interest to note that while Arg54 is conserved in the three-dimensional structures of N-βGRPs from *P. interpunctella*, *B. mori*, and *D. melanogaster*, Asp45 in *P. interpunctella* N-βGRP is replaced by a Glu in the other two species.

β-1,3-glucan-binding activities of D45A and D45K mutants were compared with that of the wild-type protein by curdlan pull-down assay and ITC with laminarin (Figure 3-16). Neither of these mutations has any effect on carbohydrate-binding activity. Thus, it is concluded that the electrostatic attractions involving Asp45 (Figure 3-14), as deduced from the crystal structure of N-βGRP:laminarihexaose (Kanagawa et al., 2011), contribute to the self-association of N-βGRP:laminarin complex.

Functional significance of the protein:carbohydrate macro complex formation was assessed by measuring proPO activation in insect blood plasma by the wild-type and mutant N-βGRPs in the absence and presence of laminarin (Figure 3-17). The D45K mutation decreased proPO activation both in the absence of, and, to a smaller extent, in the presence of, laminarin, while the D45A mutation showed no change, consistent with the AUC data (Figure 3-15).

Formation of a macro structure of βGRP:β-1,3-glucan complex, a likely cooperative process, might present a repeating pattern and thus provide an effective platform to initiate innate immune responses (Wang & Jiang, 2006; Buchon et al., 2009). Such an arrangement is reminiscent of peptidoglycan recognition by insect peptidoglycan recognition proteins (PGRPs) (Lim et al., 2006). It is also of interest to note that some lectins which have a low binding affinity for monosaccharides increase their affinities significantly for oligosaccharides by forming a cluster of lectin-oligosaccharide complexes (Weis & Drickamer, 1996).

Conclusions and Perspectives

The present study characterizes the laminarihexaose-binding site on *P. interpunctella* N-βGRP in solution, which is consistent with the crystal structure of *P. interpunctella* N-βGRP:laminarihexaose complex (Kanagawa et al., 2011).

Biophysical characterization of interactions between laminarin, a β-1,3-glucan, and N-terminal domains of insect βGRPs in solution has led to the novel finding that a stable macro

complex results from self-association of the initially formed N-βGRP:laminarin complex. Electrostatic interactions between bound protein molecules in the macro complex contribute to its stability and ability to influence the rate of activation of the prophenoloxidase pathway. An increase in β-1,6 branching of laminarin reduces the carbohydrate's βGRP-binding affinity.

Macro protein:carbohydrate complexes appear to provide an efficient means for recruiting immune response proteins and thus amplifying the initial response. The assembly of a βGRP:β-1,3-glucan complex, as suggested by the N-βGRP:laminarin complex characterized in this study, and the GNBP:PGRP:Lys-type peptidolgycan complex suggested by other studies (Gobert et al., 2003; Park et al., 2007; Kim et al., 2008; Kan et al., 2008) permit us to infer that such macro protein:carbohydrate complexes may provide a platform for binding a modular serine protease (MSP, Buchon et al., 2009; Wang & Jiang, 2010). Characterization of interactions between the PRR:PAMP complexes and MSP will yield insight into the mechanism of self-activation of MSP (Figure 3-18).

References

Barral, P., Suarez, C., Batanero, E., Alfonso, C., Alche Jde, D., Rodriguez-Garcia, M. I., Villalba, M., Rivas, G., & Rodriguez, R. (2005). An olive pollen protein with allergenic activity, Ole e 10, defines a novel family of carbohydrate-binding modules and is potentially implicated in pollen germination. *Biochem. J.* **390**, 77-84.

Boraston, A. B., Bolam, D. N., Gilbert, H. J., & Davies, G. J. (2004). Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* **382**, 769-781.

Buchon, N., Poidevin, M., Kwon, H. M., Guillou, A., Sottas, V., Lee, B. L., & Lemaitre, B. (2009). A single modular serine protease integrates signals from pattern-recognition receptors upstream of the *Drosophila* Toll pathway. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12442-12447.

Cerenius, L., Lee, B. L., & Söderhäll, K. (2008). The proPO-system: pros and cons for its role in invertebrate immunity. *Trends Immunol.* **29**, 263-271.

Chuah, C. T., Sarko, A., Deslandes, Y. & Marchessault, R. H. (1983). Triple-helical crystalline structure of curdlan and paramylon. *Macromolecules* **16**, 1375-1382.

Deslandes, Y., Marchessault, R. H., & Sarko, A. (1980). Triple-helical structure of (1-3)-β-D-glucan. *Macromolecules* **13**, 1466-1471.

Fabrick, J. A., Baker, J. E., & Kanost, M. R. (2004). Innate immunity in a pyralid moth: functional evaluation of domains from a β-1,3-glucan recognition protein. *J. Biol. Chem.* **279**, 26605-26611.

Gobert, V., Gottar, M., Matskevich, A. A., Rutschmann, S., Royet, J., Belvin, M., Hoffmann, J. A., & Ferrandon, D. (2003). Dual activation of the *Drosophila* toll pathway by two pattern recognition receptors. *Science* **302**, 2126-2130.

Gottar, M., Gobert, V., Matskevich, A. A., Reichhart, J. M., Wang, C., Butt, T. M., Belvin, M., Hoffmann, J. A., & Ferrandon, D. (2006). Dual detection of fungal infections in *Drosophila* via recognition of glucans and sensing of virulence factors. *Cell* **127**, 1425-1437.

Handa, N., & Nisizawa, K. (1961). Structural investigation of a laminaran isolated from *Eisenia bicyclis*. *Nature* **192**, 1078-1080.

Hiromasa, Y., Fujisawa, F., Aso, Y., & Roche T. E. (2004). Organization of the cores of the mammalian pyruvate dehydrogenase complex formed by E2 and E2 plus the E3-binding protein and their capacities to bind the E1 and E3 components. *J. Biol. Chem.* **279**, 6921-6933.

Hiromasa, Y., & Roche T. E. (2003). Facilitated interaction between the pyruvate dehydrogenase kinase isoform 2 and the dihydrolipoyl acetyltransferase. *J. Biol. Chem.* **278**, 33681-33693.

Hrmova, M., & Fincher, G. B. (1993). Purification and properties of three (1-3)-β-D-glucanase isoenzymes from young leaves of barley (*Hordeum vulgare*). *Biochem. J.* **289**, 453-461.

Kan, H., Kim, C.H., Kwon, H. M., Park, J. W., Roh, K. B., Lee, H., Park, B. J., Zhang, R., Zhang, J., Söderhäll, K., Ha, N. C., & Lee, B. L. (2008). Molecular control of phenoloxidase-induced melanin synthesis in an insect. *J. Biol. Chem*. **283**, 25316-25323.

Kanagawa, M., Satoh, T., Ikeda, A., Adachi, Y., Ohno, N., & Yamaguchi, Y. (2011). Structural insights into recognition of triple-helical β-glucans by an insect fungal receptor. *J. Biol. Chem.* **286**, 29158-29165.

Kanost, M. R., Jiang, H., & Yu, X.Q. (2004). Innate immune responses of a lepidopteran insect, *Manduca sexta*. *Immunol. Rev.* **198**, 97-105.

Kim, C. H., Kim, S. J., Kan, H., Kwon, H. M., Roh, K. B., Jiang, R., Yang, Y., Park, J. W., Lee, H. H., Ha, N. C., Kang, H. J., Nonaka, M., Söderhäll, K., & Lee, B. L. (2008). A three-step proteolytic cascade mediates the activation of the peptidoglycan-induced toll pathway in an insect. *J. Biol. Chem*. **283**, 7599-7607.

Janeway, C. A., Jr., & Medzhitov, R. (2002). Innate immune recognition. *Annu. Rev. Immunol.* **20**, 197-216.

Laughton, A. M., & Siva-Jothy, M. T. (2011). A standardised protocol for measuring phenoloxidase and prophenoloxidase in the honey bee, *Apis mellifera*. *Apidologie* **42**, 140-149.

Lim, J. H., Kim, M. S., Kim, H. E., Yano, T., Oshima, Y., Aggarwal, K., Goldman, W. E., Silverman, N., Kurata, S., & Oh, B. H. (2006). Structural basis for preferential recognition of diaminopimelic acid-type peptidoglycan by a subset of peptidoglycan recognition proteins. *J. Biol. Chem.* **281**, 8286-8295.

Ma, C., & Kanost, M. R. (2000). A β1,3-glucan recognition protein from an insect, *Manduca sexta*, agglutinates microorganisms and activates the phenoloxidase cascade. *J. Biol. Chem.* **275**, 7505-7514.

Mishima, Y., Quintin, J., Aimanianda, V., Kellenberger, C., Coste, F., Clavaud, C., Hetru, C., Hoffmann, J.A., Latge, J.P., Ferrandon, D., & Roussel, A. (2009). The N-terminal domain of *Drosophila* Gram-negative binding protein 3 (GNBP3) defines a novel family of fungal pattern recognition receptors. *J. Biol. Chem.* **284**, 28687-28697.

Ochiai, M., & Ashida, M. (2000). A pattern-recognition protein for β-1,3-glucan. The binding domain and the cDNA cloning of β-1,3-glucan recognition protein from the silkworm, *Bombyx mori*. *J. Biol. Chem.* **275**, 4995-5002.

Okuyama, K., Otsubo, A., Fukuzawa, Y., Ozawa, M., Harada, T. & Kasai, N. (1991). Single helical structure of native curdlan and its aggregation state. *Carbohydr. Chem.* **10**, 645-656.

Park, J. W., Kim, C. H., Kim, J. H., Je, B. R., Roh, K. B., Kim, S. J., Lee, H. H., Ryu, J. H., Lim, J. H., Oh, B. H., Lee, W. J., Ha, N. C., & Lee, B. L. (2007). Clustering of peptidoglycan

recognition protein-SA is required for sensing lysine-type peptidoglycan in insects. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 6602-6607.

Perkins, S. J., Miller, A., Hardingham, T. E., & Muir, H. (1981). Physical properties of the hyaluronate binding region of proteoglycan from pig laryngeal cartilage: Densitometric and small-angle neutron scattering studies of carbohydrates and carbohydrate-protein macromolecules. *J. Mol. Biol.* **150**, 69-95.

Schachman, H. K. (1959). *Ultracentrifugation in Biochemistry*. Academic Press, New York.

Schneider, D. S., & Chambers, M. C. (2008). Microbiology. Rogue insect immunity. *Science* **322**, 1199-1200.

Takahasi, K., Ochiai, M., Horiuchi, M., Kumeta, H., Ogura, K., Ashida, M., & Inagaki, F. (2009). Solution structure of the silkworm βGRP/GNBP3 N-terminal domain reveals the mechanism for β-1,3-glucan-specific recognition. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11679-11684.

Wang, Y., & Jiang, H. (2006). Interaction of β-1,3-glucan with its recognition protein activates hemolymph proteinase 14, an initiation enzyme of the prophenoloxidase activation system in *Manduca sexta*. *J. Biol. Chem.* **281**, *9271*-9278.

Wang, Y. & Jiang, H. (2010). Binding properties of the regulatory domains in *Manduca sexta* hemolymph proteinase-14, an initiation enzyme of the prophenoloxidase activation system. *Dev. Comp. Immunol*. **34**, 316-322.

Weis, W. I., & Drickamer, K. (1996). Structural basis of lectin-carbohydrate recognition. *Annu. Rev. Biochem.* **65**, 441-473.

Wiseman, T., Williston, S., Brandts, J. F., & Lin, L. N. (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal. Biochem.* **179**, 131-137.

Young, S. H., Dong, W. J, & Jacobs, R. R. (2000). Observation of a partially opened triple-helix conformation in 1->3-β-glucan by fluorescence resonance energy transfer spectroscopy. *J. Biol. Chem.* **275**, 11874-11879.

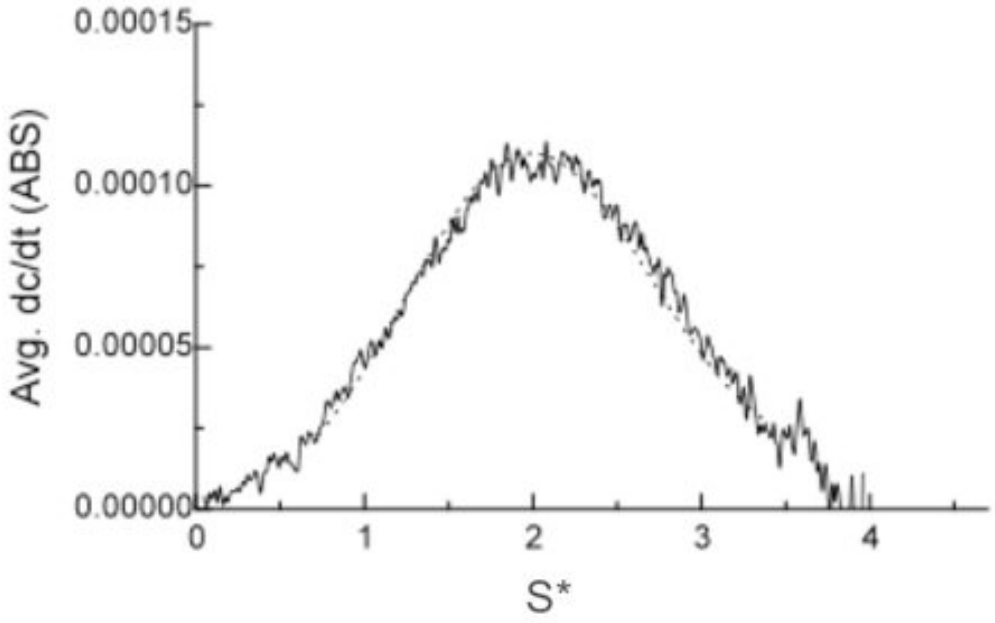


Figure 3-1 Sedimentation velocity analysis for N-βGRP from *P. interpunctella*. Dotted line is the fitting result by DCDT+ software.

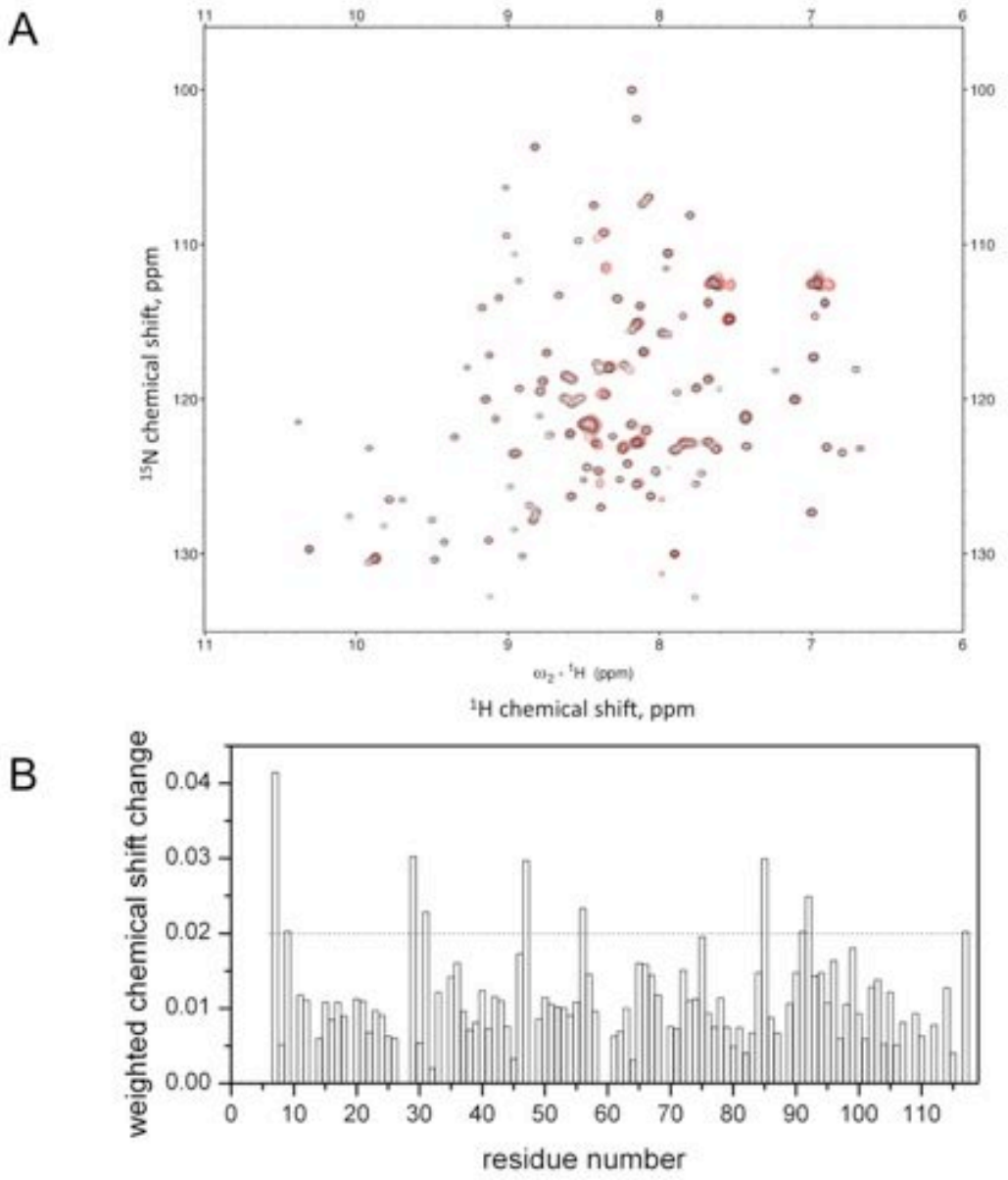


Figure 3-2 Mapping of ligand-binding site on *P. interpunctella* N-βGRP by NMR titration of laminarihexaose at 25°C, pH 6.5: (A) ^{1}H-^{15}N HSQC spectra of N-βGRP (0.5 mM) in the absence (*black*) or presence (*red*) of laminarihexaose (3 mM); (B) Chemical shift changes undergone by the backbone ^{15}N-^{1}H groups; the weighted average of chemical shift changes of an ^{15}N-^{1}H group was calculated by using the formula, $[(\Delta H^2+(\Delta N/5)^2)/2]^{1/2}$, where ΔH and ΔN represent ^{1}H and ^{15}N chemical shift change, respectively.

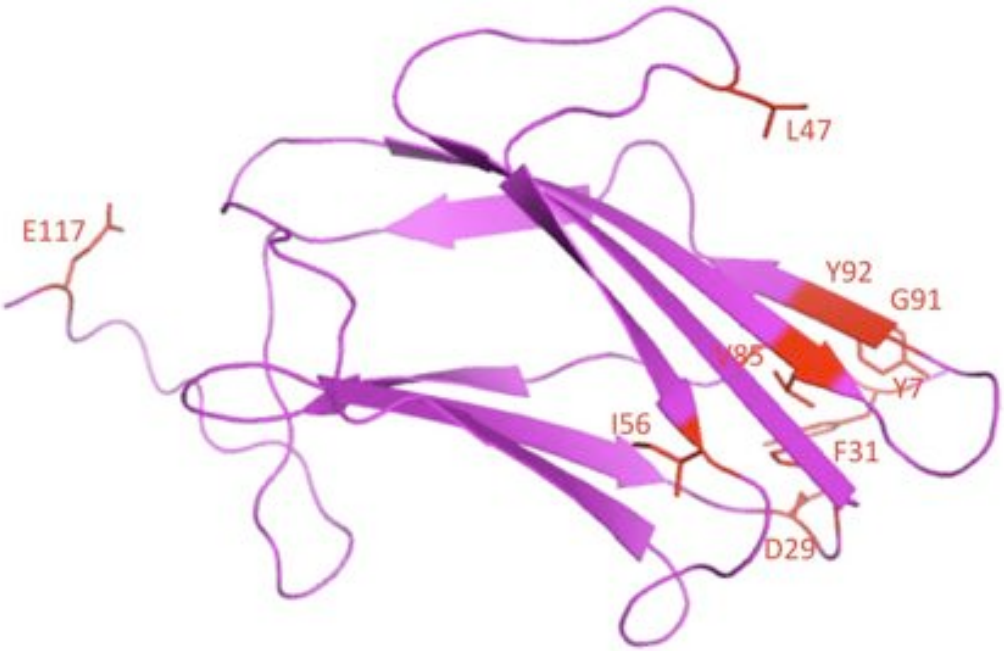


Figure 3-3 Laminarihexaose-binding site on *P. interpunctella* N-βGRP. Residues undergoing a chemical shift perturbation of > 0.02 ppm upon laminarihexaose titration are shown in red.

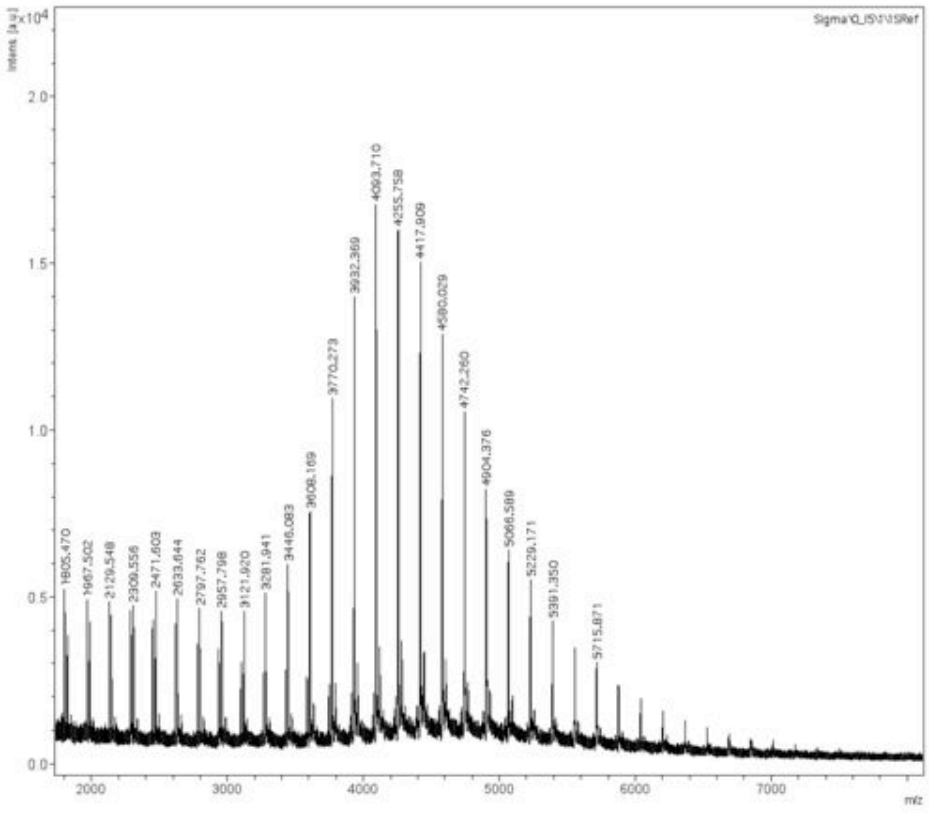


Figure 3-4 MALDI mass spectrum of laminarin. The data were acquired on a Bruker Ultraflex II MALDI-TOF mass spectrometer (Bremen, Germany). 2,5-dihydroxybenzonic acid (DHB) was used as a matrix. The instrument was operated in a positive ion mode.

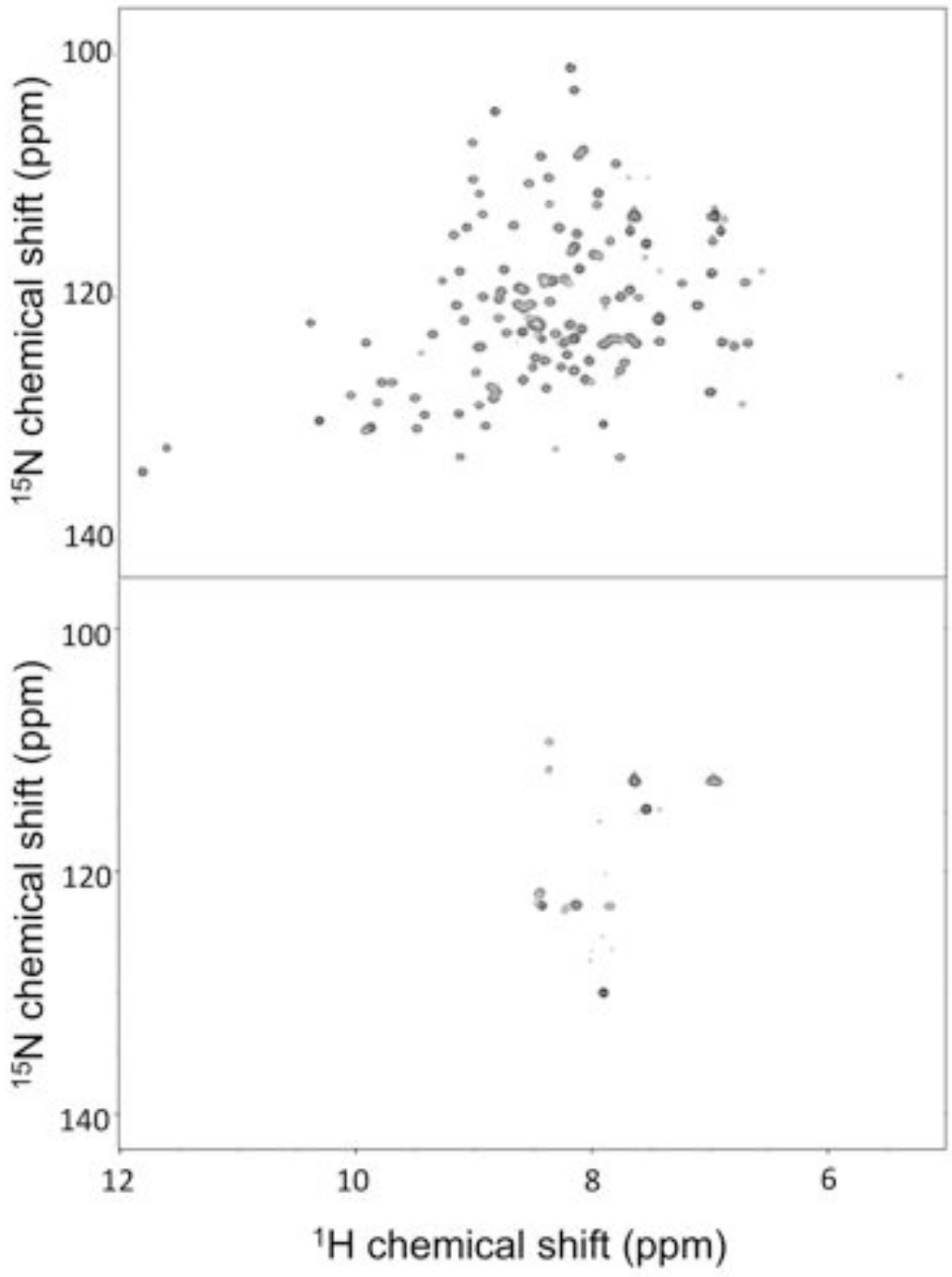


Figure 3-5 ^{1}H-^{15}N HSQC spectra of N-βGRP (0.5 mM) in the absence (*upper*) or presence (*lower*) of laminarin (1 mM) at 25°C, pH 6.5.

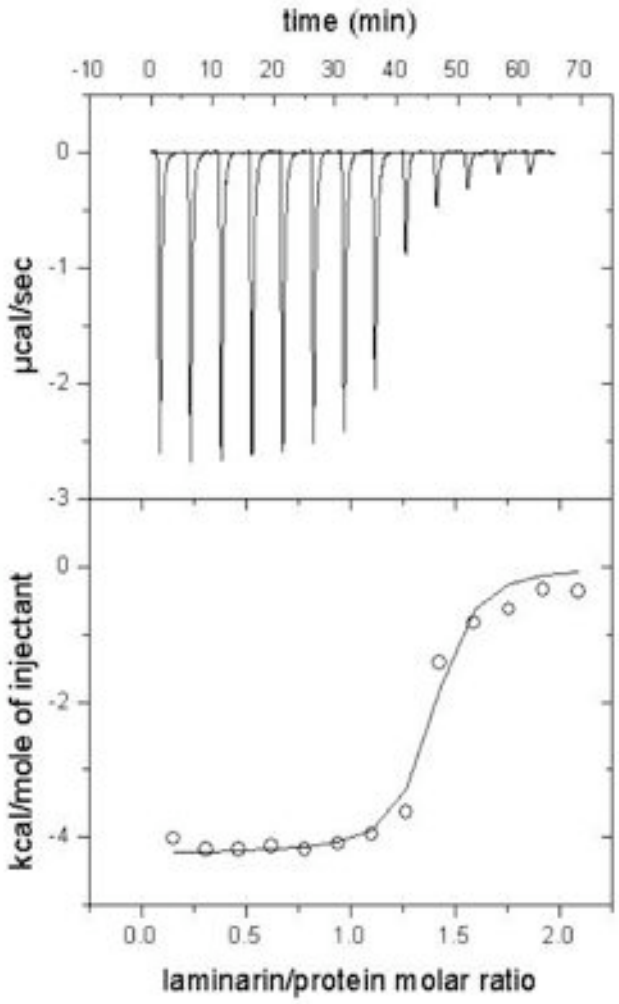


Figure 3-6 Titration of laminarin with N-βGRP as monitored by ITC. The protein concentration was 78 μM and the ligand (injectant) concentration was 1.67 mM.

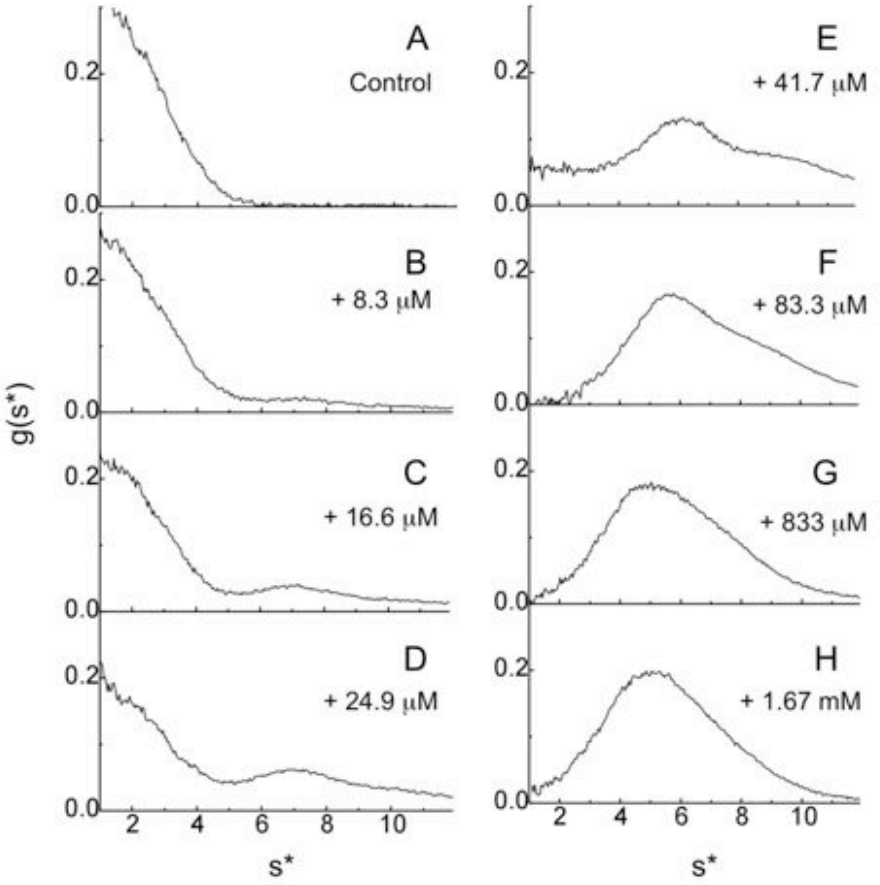


Figure 3-7 Effect of addition of varying concentrations of laminarin on N-βGRP, as monitored by sedimentation velocity experiments: Sedimentation profiles of N-βGRP (26.2 μM) in the absence (A) or presence of different levels of laminarin (B-H). Sedimentation at 49,000 rpm and 20°C was monitored by measuring absorbance at 280 nm.

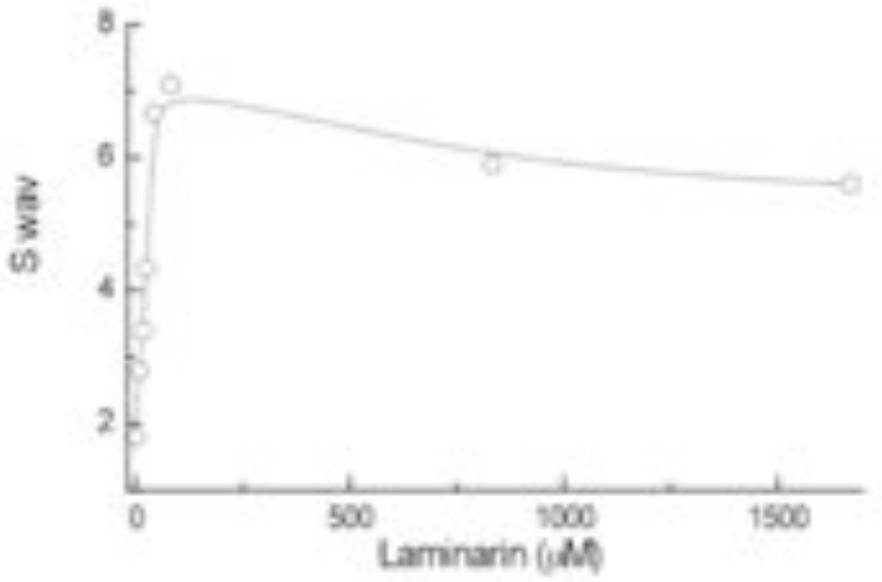


Figure 3-8 The weight average sedimentation coefficient for the mixture of N-βGRP and laminarin. The *Swav* values for Figure 3-6 (A-H) are shown as open circle.

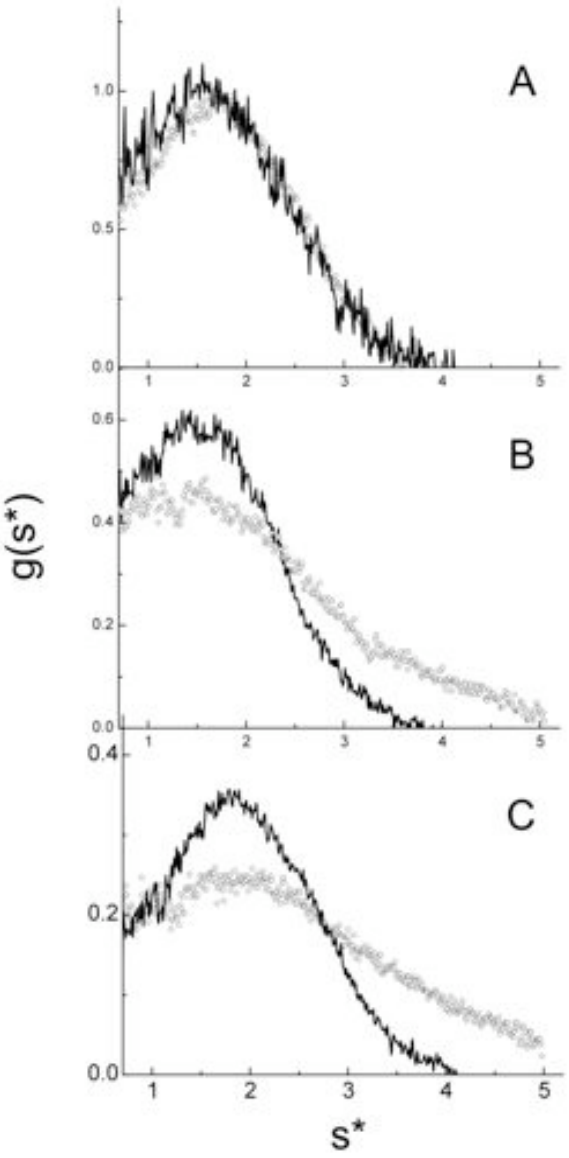


Figure 3-9 $g(s^*)$ profiles of sedimentation velocity studies of N-βGRP in the presence of varying amounts of laminarihexaose. Panel A corresponds to 54 μM N-βGRP alone (solid line) and in the presence of 1mM laminarihexaose (dotted line); panel B to 234 μM N-βGRP alone (solid line) and in the presence of 14.1 mM laminalihexaose (dotted line); and panel C to 1.4 mM N-βGRP alone (solid line) and in the presence of 26 mM laminalihexaose (dotted line). Sedimentation experiments were performed at 49,000 rpm and 20°C, using absorption optics at 280 nm (A), 300 nm (B) and 308 nm (C). The shift of the dotted line toward higher s^* values with increasing concentration of laminarihexaose strongly suggests the formation of a weak macro complex of the protein and the hexasaccharide.

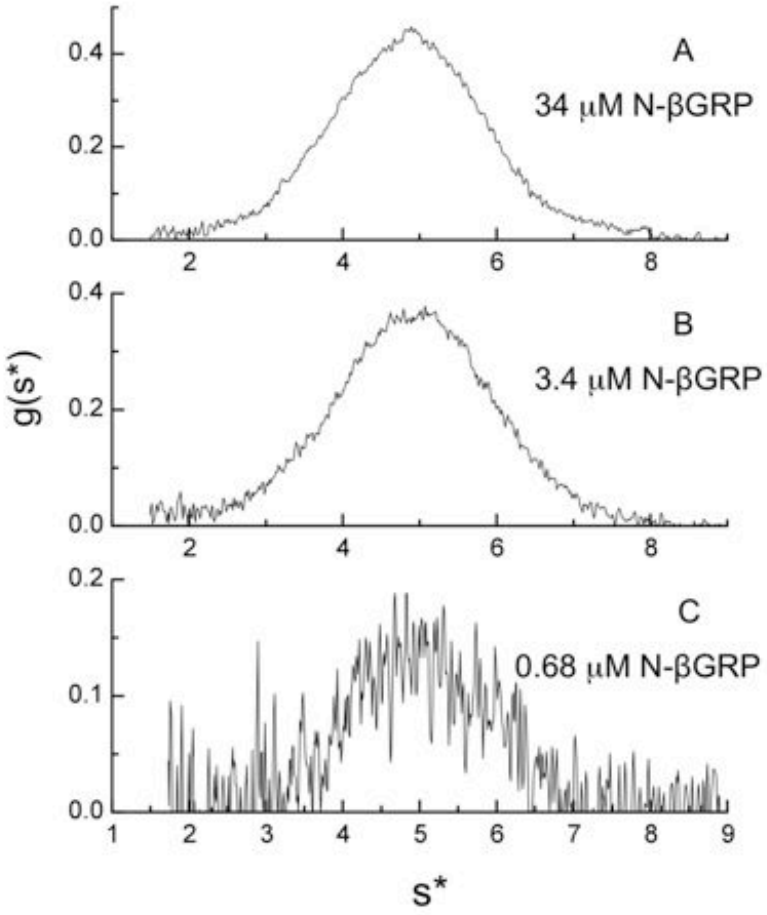


Figure 3-10 Effect of dilution on N-βGRP:laminarin complex as monitored by sedimentation velocity profiles: (A) 34 μM N-βGRP and 780 μM laminarin at 280 nm; (B) 3.4 μM N-βGRP and 78 μM laminarin at 220 nm; (C) 0.68 μM N-βGRP and 15.6 μM laminarin at 205 nm. The concentration ratio of N-βGRP:laminarin was maintained at 1:22.9. Sedimentation was carried out at 49,000 rpm and 20°C.

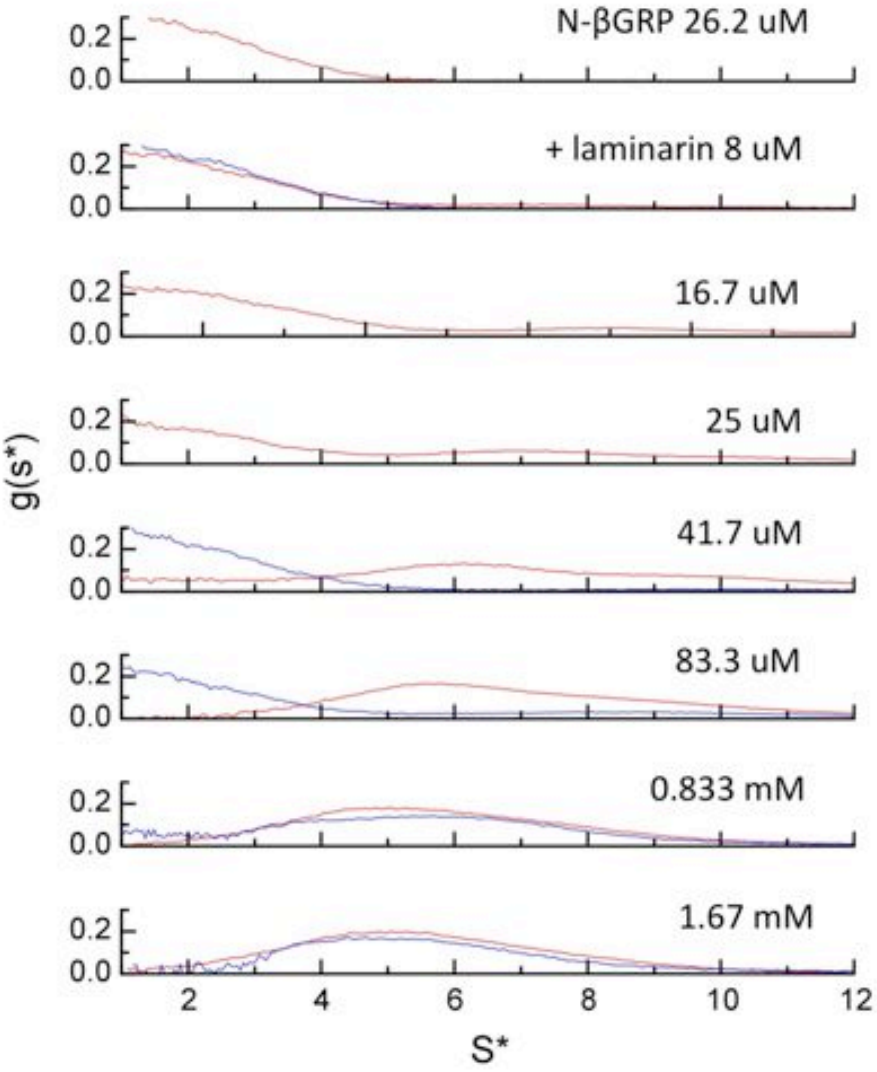


Figure 3-11 Sedimentation velocity profiles of N-βGRP:laminarin complex with increasing amounts of laminarin: *L. digitata* laminarin (red); *E. bicyclis* laminarin (blue) added to a constant amount of the protein (26.2 μM). *E. bicyclis* laminarin is more branched with β(1-3)/ β(1−6) ratio of 3 than is *L. digitata* laminarin that has a corresponding value of 7.

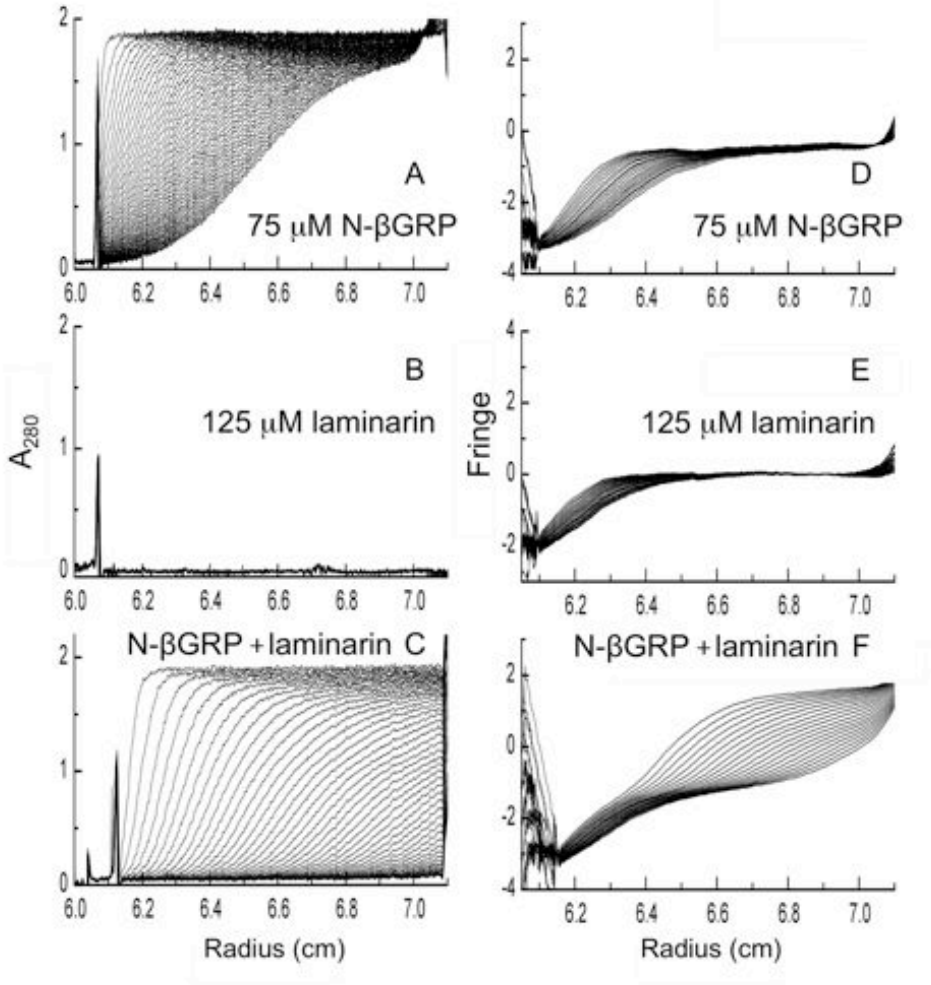


Figure 3-12 Sedimentation velocity profiles for N-βGRP, laminarin and their mixtures, as monitored at 49,000 rpm and 20°C by absorbance at 280 nm (A-C) and interference optics (D-F) at 5 min. intervals: 75 μM N-βGRP (A & D) and 125 μM laminarin (B & E) were sedimented separately or together (C & F). Fringe displacements in D-F are given in arbitrary units.

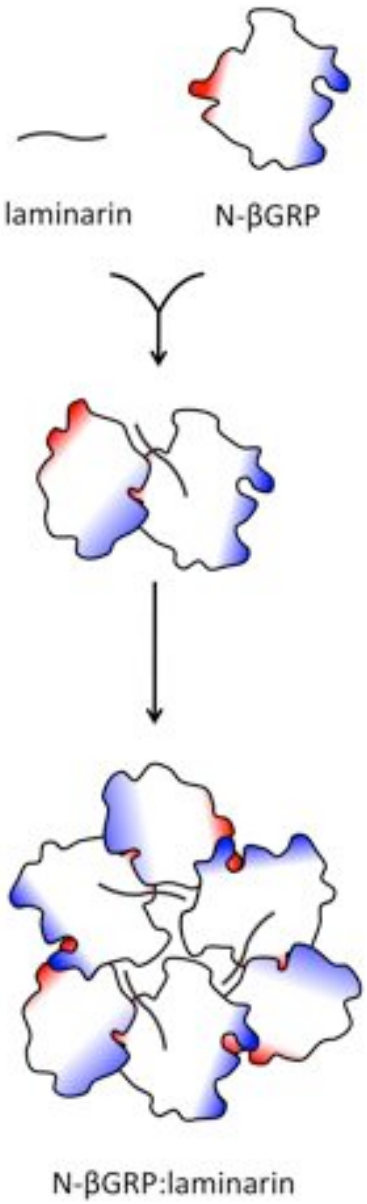


Figure 3-13 A schematic representation of formation of N-βGRP: β-1,3-glucan macro complex. Laminarin binding to N-βGRP causes self-association of N-βGRP:laminarin complex.

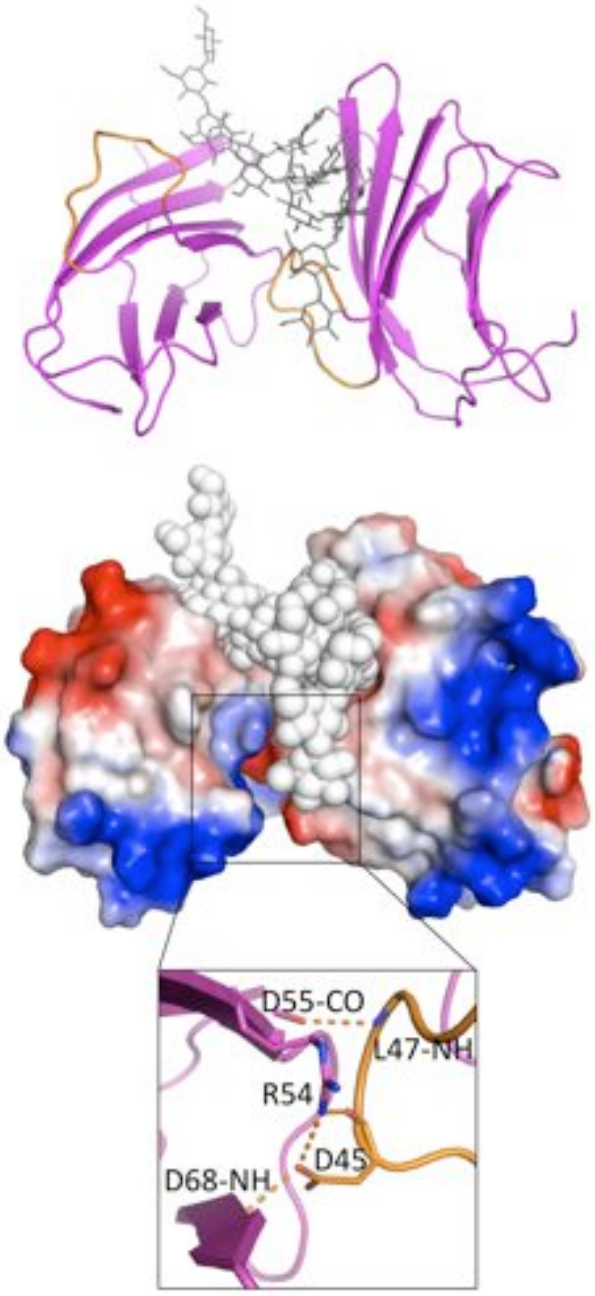


Figure 3-14 N-βGRP packing and concomitant electrostatic interactions as can be observed in the crystal structure of N-βGRP:laminarihexaose (Kanagawa et al., 2011). Negative and positive potential surfaces are shown in red and blue, respectively.

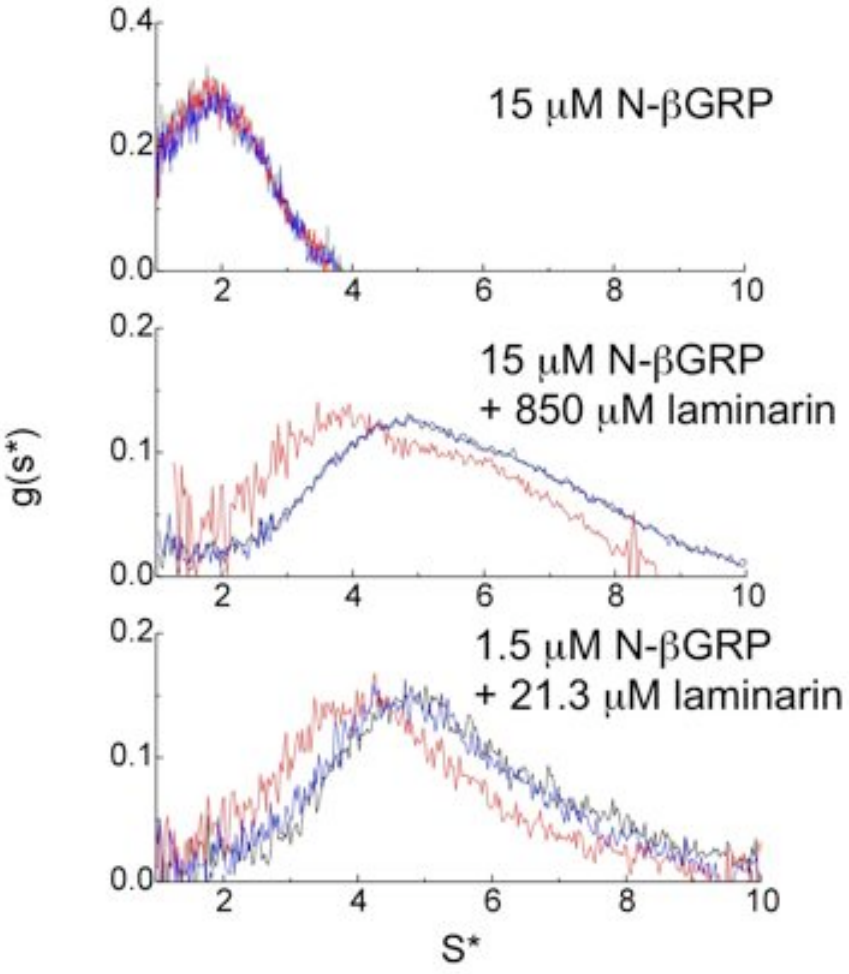


Figure 3-15 Sedimentation velocity profiles for N-βGRP (wild type, black; D45A, blue; and D45K, red) in the presence of laminarin.

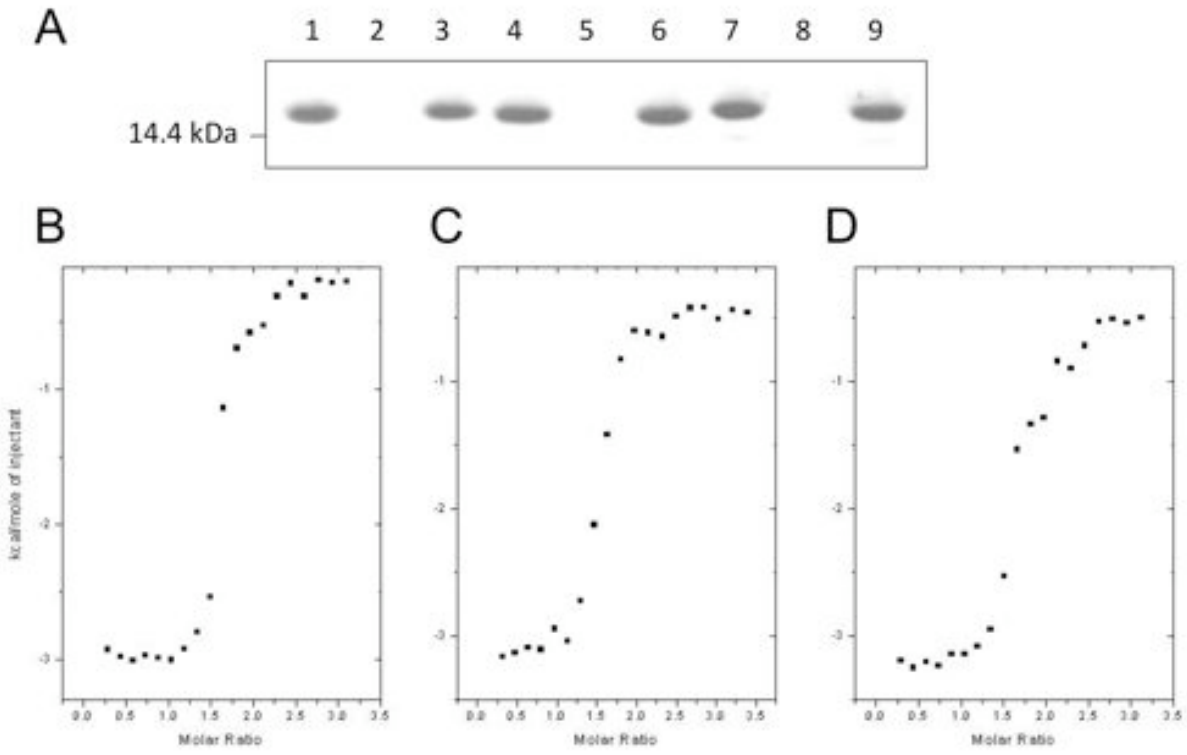


Figure 3-16 β-1,3-glucan-binding activities of N-βGRP and mutants as measured by curdlan pull-down assay and isothermal titration calorimetry. (A) SDS-PAGE of purified recombinant N-βGRP proteins before and after co-precipitation with curdlan: purified wide type (lane 1), unbound (lane 2), and bound (lane 3); purified D45A mutant (lane 4), unbound (lane 5), and bound (lane 6); purified D45K mutant (lane 7), unbound (lane 8), and bound (lane 9). (B-D) Isothermal titration of laminarin with N-βGRP wild type (B), D45A (C) and D45K (D). The protein concentrations were ~78 μM and the ligand (injectant) concentration was 1.67 mM.

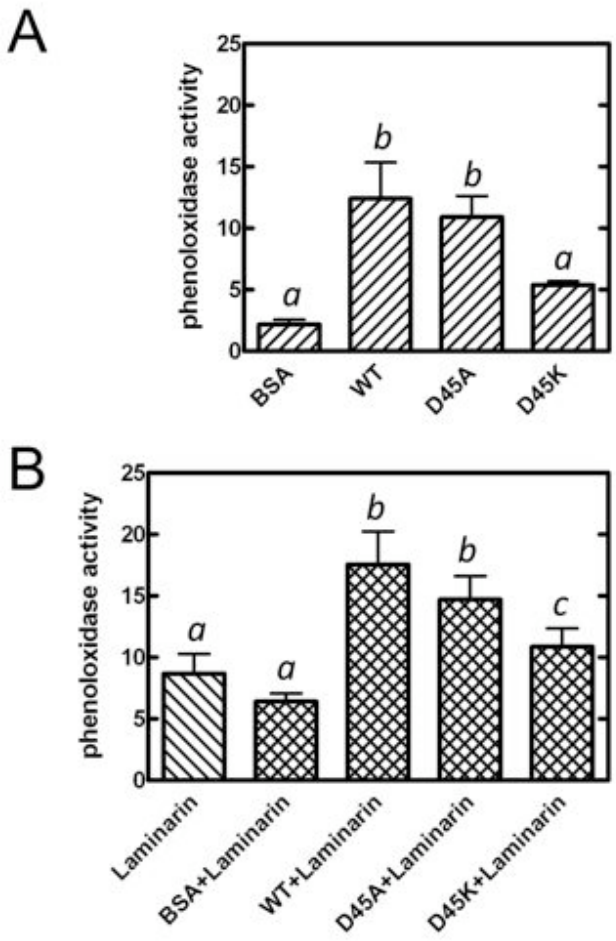


Figure 3-17 Activation of the prophenoloxidase pathway by N-βGRP and mutants without (A) or with laminarin (B). Samples of plasma (5 μl) were mixed with protein alone or with protein and laminarin. After incubation at room temperature for 15 min, phenoloxidase activity was measured using dopamine hydrochloride as a substrate, as described under "Materials and Methods". The bars represent the means ± S.E. of data from three sets of measurements on a pooled plasma sample. Bars labeled with different letters (*a*, *b*, and *c*) are significantly different [analysis of variance (ANOVA), $p < 0.05$].

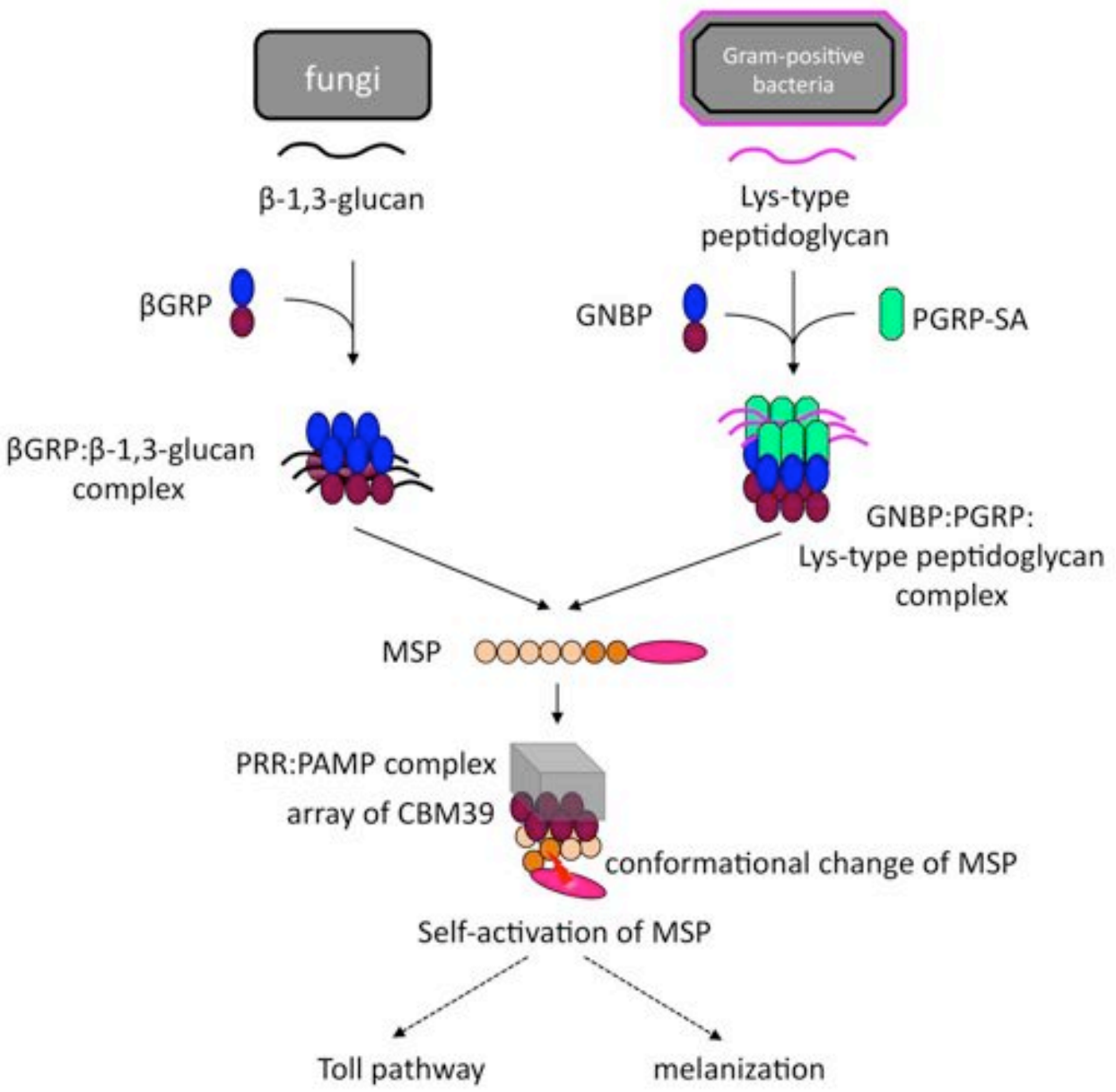


Figure 3-18 Model of βGRP-mediated activation of Toll and melanization pathways. Formation of βGRP:β-1,3-glucan complex or GNBP:PGRP:Lys-type peptidoglycan complex activates MSP (modular serine protease), possibly through the interactions between N-βGRP and the LDLr domain (Low Density Lipoprotein receptor, salmon circle). A conformational change is proposed to trigger the self-activation of MSP by cleavage of its CCP (complement control protein, orange circle) domain. The gray box indicates the βGRP:β-1,3-glucan complex or GNBP:PGRP:Lys-type peptidolgycan complex. Activated MSP then initiates downstream cascades that lead to Toll and melanization pathways.

Chapter 4 - Characterization of the N-terminal Domain of GNBPA2 from *Anopheles gambiae*

Introduction

Six βGRP/GNBP members are found in the genome of the malaria vector mosquito *Anopheles gambiae*, namely GNBPA1, GNBPA2 and GNBPB1-B4 (Christophides et al., 2002). GNBPA1 and GNBPA2 are homologs to *Lepidoptera* βGRP, and the four GNBPBs are homologs of glucanases whose N-terminal region does not contain the CBM39 domain. *Anopheles* GNBPs may participate in anti-malaria responses through unknown mechanisms. As the first mosquito GNBP member identified, GNBPB1 can be used as an anti-malaria reaction marker of the mosquito immunity (Dimopoulos et al., 1997). The mRNA level of GNBPB1 increases upon *Plasmodium berghei* infection (Richman et al., 1997) and *Plasmodium falciparum* infection (Tahar et al., 2002; Lehmann et al., 2009). These studies suggested a relationship between glucan digestion and anti-malaria responses. However, the surface of malaria parasites differs greatly from that of bacteria or fungi, and the molecular pattern that mosquito GNBPs can recognize is unknown. An RNAi screen for anti-bacterial and anti-*Plasmodium* gene revealed that these six GNBPs are involved in immune responses in different ways (Warr et al., 2008). Among the six GNBPs, silencing of GNBPA2 caused the strongest effect of *Plasmodium* infection. GNBPA2 is the closest homolog to *Lepidoptera* βGRP. The surface on *Plasmodium* is rich with glycosylphosphatidylinositol (GPI), which is important for all stages (Yassine & Osta, 2010). Whether GNBPA2 recognizes the carbohydrate moiety of the GPI may help to understand its anti-*Plasmodium* role. As the C-terminal domain is conserved among GNBPs, the N-terminal domain of GNBPA2 might provide a switch for creating its uniqueness as observed in the RNAi screen (Warr et al., 2008). We have characterized the binding specificity of the N-terminal domain of GNBPA2, which is denoted as N-GNBPA2 in this study.

Materials and Methods

The experimental steps for ITC, curdlan pull-down assay and AUC were as described in the Materials and Methods section of Chapter 3.

Reverse Transcriptase-Polymerase Chain Reaction

Because GNBPA2 is incompletely annotated in ENSEMBL and thus lacks the N-terminal end, reverse transcription was carried out. First-strand cDNA synthesis was completed using SuperScript First-Strand Synthesis System (Invitrogen). Primers used are: 5' RACE primer, 5'-GGCTGTAAGTACCGGCGGAATCATGTCA-3'; 5' RACE primer (Nested): 5'-CGCGAACTGCTATGGCCCGTGCACTCTT-3'.

Recombinant Protein Expression and Purification

The cDNA sequence of N-GNBPA2 from *A. gambiae* was cloned via NcoI/XhoI sites into Invitrogen pPROEX HTb plasmid. His_6-GNBPA2 was expressed in *E. coli* DH5α by induction with IPTG, and the expressed protein was found in the inclusion body. In an attempt to get soluble N-GNBPA2, its DNA sequence was cloned *via* NcoI/HindIII into pET28 Novagen plasmid. GNBPA2-His_6 was expressed in *E. coli* BL21 (DE3) by induction with IPTG, and still most of the expressed protein accumulated in the inclusion body. The GNBPA2-His_6-pET28 plasmid was co-transformed with a plasmid encoding GroEL and GroES into *E. coli* BL21 (DE3). Selection for the presence of both plasmids was made by μg/ml kanamycin and 50 μg/ml chloramphenicol. Soluble GNBPA2-His_6 was obtained and purified through a Ni^{2+}-affinity column. For site-directed mutagenesis, N-GNBPA2 containing the triple mutation, Q45E/Q46E/Q49E or Q45K/Q46K/Q49K,was prepared according to the instructions of QuikChange mutagenesis kit (Stratagene Agilent Technologies, Santa Clara, CA).

Solubility Test

Protein concentrations at different Arg concentrations (50, 100, 150, 200 and 250 mM), at pH 8.0, were determined by the Bradford method (Bradford, 1976) by measuring absorbance at 595 nm.

Results and Discussion

Sequence of N-GNBPA2

The sequencing result of the 5'-end (N-terminal region of GNBPA2) indicated that a sequence corresponding to 30 amino acid residues was missing in the VectorBase/ENSEMBL (Figure 4-1). This could be due to an incorrected gene prediction. Because the missing sequence

is after the CBM39 domain and before the C-terminal glucanse-like domain of GNBPA2, these 30 residues are part of the linker region between these two domains. The sequence excluding the putative signal sequence was cloned for recombinant protein expression (Figure 4-2A). A homology model of N-GNBPA2 was built based on the template structure of N-GNBP3 from *Drosophila* (PDB entry: 3IE4; Mishima et al., 2009) using SWISS-MODEL (Kiefer et al., 2009) (Figure 4-2B). N-GNBP3 has the highest similarity with N-GNBPA2 among all known CBM39 structures. The major differences between N-GNBPA2 and N-βGRPs/GNBPs from *M. sexta*, *P. interpunctella*, *B. mori*, *D. melanogaster*, and *T. molitor* are as follows: introduction of a Trp on strand B, amino acid sequence of the C-C'loop, and loss of a Trp on strand F. The Trp on strand B is probably solvent-exposed . As derived from known CBM39 structures (Takahashi et al., 2009; Mishima et al., 2009; Kanagawa et al., 2009), the loop C-C' shields the Trp on strand F. In N-GNBPA2, loss of this Trp appears to cause conformational changes for the loop C-C'. A different conformation for the loop C-C', if indeed observed, would suggest a new or altered function for GNBPA2 (Figure 4-3).

Expression and Purification of N-GNBPA2

Expressed N-GNBPA2 ended in the inclusion body; however, co-expression with GroEL and GroES (Baker et al., 2000) yielded soluble recombinant proteins (Figure 4-4A). When the eluting imidazole was removed, the protein precipitated. The protein precipitate could be re-dissolved in the eluting imidazole (pH 8.0), 1.5 M sodium chloride, or 0.25 M Arg (pH 8.0). Lowering Arg concentration resulted in protein precipitation and, therefore, a concentration-dependent solubility test was performed (Figure 4-4B). To test the role of the three Gln residues on loop C-C', two mutants (Q45E/Q46E/Q49E and Q45K/Q46K/Q49K) were made. The mutations were confirmed by DNA sequencing and mass spectrometry. However, neither of the mutants showed increased solubility.

Characterization of N-GNBPA2

N-GNBPA2 shows a decreased curdlan-binding activity, as compared with N-βGRP (Figure 4-5). Upon titration with laminarin, N-GNBPA2 solution became turbid. The ITC signals were unusual and not representative of a typical titration (Figure 4-6).

To characterize the solution state of N-GNBPA2, we did AUC sedimentation velocity experiments. Unlike any βGRP/GNBP, a sample of N-GNBPA2 alone yeilded a 4.5 S peak

together with the expected peak (Figure 4-7A). This behavior indicated the presence of a high molecular weight aggregate in the clear solution of N-GNBPA2. The ratio of OD260 to OD280 was ~1.5, suggesting the presence of nucleic acid(s) in the sample solution. Indeed, when the clear solution was assayed by agarose gel electrophoresis, a band of less than 100 bp appeared (The arrow in Figure 4-7B). Treatment of the sample by RNase did not affect the band, while DNase treatment (DNase I, 15 min) resulted in the disappearance of this band, thus indicating that the protein was associated with DNA.

DNA binding by an immunoglobulin fold is not uncommon. Usually the loop region is associated with the binding specificity for a DNA sequence (Rudolph & Gergen, 2001). It is possible that the loop C-C' whose length is conserved in CBM39 (Figure 4-2) play a role in DNA interactions. The precipitate observed in the solubility test, and the macro complex in the AUC experiment, are likely due to the presence of protein:DNA complex(es). The biological implication of the putative DNA-binding property of N-GNBPA2 may be gleaned from the fact that neutrophil extracellular traps utilize extracellular nucleic acids to enhance immune responses (Urban et al., 2006; Altincicek et al., 2008). Further experiments are needed to elucidate the structure-function relationship for N-GNBPA2.

Summary and Perspectives

The N-terminal region of *A. gambiae* GNBPA2 was sequenced. Three-dimensional homology modeling suggests that the C-C' loop of N-GNBPA2 is likely to have an altered conformation, as compared to those of other βGRP/GNBP members. The altered conformation of the loop C-C' may have biological implications: Unlike other βGRPs/GNBPs characterized to date, N-GNBPA2 binds DNA, in addition to β-1,3-glucan, thus suggesting a new role for CBM39. The loop C-C' of N-GNBPA2 could be swapped with that of *Plodia* N-βGRP to characterize the structural and functional properties of this loop as well as for comparison of these two βGRPs/GNBPs from two different orders of insects. Site-directed mutagenesis of the Trp residue on strand B and introduction of the missing Trp on strand F may provide clues about the role of these aromatic residues in ligand-binding.

Lack of particular ligand specificity of βGRP/GNBP proteins is consistent with the notion a limited number of PRRs serves as receptors for a variety of pathogen surfaces. To deal with multiple pathogens, a host requires an effective signaling system, triggering melanization

and AMP expression in insect hemolymph. However, such multi-ligand specificity should be tuned to adapt to the evolution of insects and their pathogens. Understanding the role of mosquito βGRP/GNBP proteins in anti-*Plasmodium* responses may help in developing strategies to control transmission of malaria.

References

Altincicek, B., Stötzel, S., Wygrecka, M., Preissner, K. T., & Vilcinskas, A. (2008). Host-derived extracellular nucleic acids enhance innate immune responses, induce coagulation, and prolong survival upon infection in insects. *J. Immunol*. **181**, 2705-2712.

Bradford, M.M. (1976). Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem*. **72**, 248-254.

Christophides, G. K., Zdobnov, E., Barillas-Mury, C., Birney, E., Blandin, S., Blass, C., Brey, P. T., Collins, F. H., Danielli, A., Dimopoulos, G., Hetru, C., Hoa, N. T., Hoffmann, J. A., Kanzok, S. M., Letunic, I., Levashina, E. A., Loukeris, T. G., Lycett, G., Meister, S., Michel, K., Moita, L. F., Muller, H. M., Osta, M. A., Paskewitz, S. M., Reichhart, J. M., Rzhetsky, A., Troxler, L., Vernick, K. D., Vlachou, D., Volz, J., von Mering, C., Xu, J., Zheng, L., Bork, P., & Kafatos, F. C. (2002). Immunity-related genes and gene families in *Anopheles gambiae*. *Science* **298**, 159-165.

Baker, J. C., Yan, X., Peng, T., Kasten, S., & Roche, T. E. (2000). Marked differences between two isoforms of human pyruvate dehydrogenase kinase. *J. Biol. Chem*. **275**, 15773-15781.

Debierre-Grockiego, F., & Schwarz, R. T. (2010). Immunological reactions in response to apicomplexan glycosylphosphatidylinositols. *Glycobiology* **20**, 801-811.

Dimopoulos, G., Richman, A., Müller, H. M., & Kafatos, F. C. (1997). Molecular immune responses of the mosquito *Anopheles gambiae* to bacteria and malaria parasites. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 11508-11513.

Kanagawa, M., Satoh, T., Ikeda, A., Adachi, Y., Ohno, N., & Yamaguchi, Y. (2011). Structural insights into recognition of triple-helical β-glucans by an insect fungal receptor. *J. Biol. Chem.* **286**, 29158-29165.

Kiefer, F., Arnold, K., Künzli, M., Bordoli, L., & Schwede, T. (2009). The SWISS-MODEL Repository and associated resources. *Nucl. Acids Res*. **37**, D387-D392.

Lehmann T, Hume JC, Licht M, Burns CS, Wollenberg K, Simard F, & Ribeiro JM. (2009). Molecular evolution of immune genes in the malaria mosquito *Anopheles gambiae*. *PLoS One* **4**, e4549.

Mishima, Y., Quintin, J., Aimanianda, V., Kellenberger, C., Coste, F., Clavaud, C., Hetru, C., Hoffmann, J.A., Latge, J.P., Ferrandon, D., & Roussel, A. (2009). The N-terminal domain of *Drosophila* Gram-negative binding protein 3 (GNBP3) defines a novel family of fungal pattern recognition receptors. *J. Biol. Chem.* **284**, 28687-28697.

Richman, A. M., Dimopoulos, G., Seeley, D., & Kafatos, F. C. (1997). *Plasmodium* activates the innate immune response of *Anopheles gambiae* mosquitoes. *EMBO J*. **16**, 6114-6119.

Rudolph, M. J, & Gergen, J. P. (2001). DNA-binding by Ig-fold proteins. *Nat. Struct. Biol*. **8**, 384-386.

Tahar, R., Boudin, C., Thiery, I., & Bourgouin, C. (2002). Immune response of *Anopheles gambiae* to the early sporogonic stages of the human malaria parasite *Plasmodium falciparum*. *EMBO J*. **21**, 6673-6680.

Takahasi, K., Ochiai, M., Horiuchi, M., Kumeta, H., Ogura, K., Ashida, M., & Inagaki, F. (2009). Solution structure of the silkworm βGRP/GNBP3 N-terminal domain reveals the mechanism for β-1,3-glucan-specific recognition. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 11679-11684.

Urban, C. F., Reichard, U., Brinkmann, V., & Zychlinsky, A. (2006). Neutrophil extracellular traps capture and kill *Candida albicans* yeast and hyphal forms. *Cell Microbiol*. **8**, 668-676.

Warr, E., Das, S., Dong, Y., & Dimopoulos, G. (2008). The Gram-negative bacteria-binding protein gene family: its role in the innate immune system of *Anopheles gambiae* and in anti-*Plasmodium* defence. *Insect Mol. Biol*. **17**, 39-51.

Yassine, H., & Osta, M. A. (2010). *Anopheles gambiae* innate immunity. *Cell Microbiol*. **12**, 1-9.

acgccaattatgtgaggtagctcactcattaggcacccccaggctttacactttatgctc
T P I M - G S S L I R H P Q A L H F M L
ggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatga
G S Y V V W N C E R I T I S H R K Q L -
ccatgattacgccaagctcagaattaaccctcactaaagggactagtctgcaggtttaaa
P - L R Q A Q N - P S L K G L V C R F K
cgaattcgcccttggacactgacatggactgaaggagtagaaaagtgaacagctacatct
R I R P W T L T W T E G V E K - T A T S
acaagcagaatgaatcgttttctgagactctttatttttacttttgtgtttccctttca
T S R **M** N R F L R L F I L L F V F P F S
tatagtgatcctcgtaaaagtagtcgctatcagccgccaaagccccgatttgaagtgttt
Y S D P R K S S R Y Q P P K P R F E V F
gatcccaagggattgattgtgtggattaacgctgatccaggaattagttcgtttacattt
D P K G L I V W I N A D P G I S S F T F
cacgggaaacttaatcagcagtttgtacagaattacgatgttggacgatgggctcagaca
H G K L N Q Q F V Q N Y D V G R W A Q T
ataattaaaataaaaaacggccgatacctttcatcgatcgcgaagcaaaactcgtacca
I I K I K N G R Y L F I D R E A K L V P
ggtgataccatattttatcgtacagtgattgttcgtaacggacagacctaccgtacgaat
G D T I F Y R T V I V R N G Q T Y R T N
tccggagcgtttaccgtggaagagctgcgtccggccgctacaccttcacccacatccact
S G A F T V E E L R P A A T P S P T S T
tccggtatcgtgtcgcgctctgctacgtccgagctatatccatatgtgctcaccgcaaat
S G I V S R S A T S E L Y P Y V L T A N
gaacgacgaactaaggacgtgaggtcaaccgctacccaaactgatgattacgagggaaat
E R R T K D V R S T A T Q T D D Y E G N
tcagctgaacattgcgctaatgcacaaaccattgtaaatggacgaaaagtgtgtgctggt
S A E H C A N A Q T I V N G R K V C A G
aaactgttgtttgaggataattttaatggtcgttcgatagatttgcgcaagtggcgcatt
K L L F E D N F N G R S I D L R K W R I
gagaaccgttttgcatctgaccccgacaatgaatttgtggtttatgctgatttcccagaa
E N R F A S D P D N E F V V Y A D F P E
aatataatgatacaaaatggtctattagctatccgccccactctgttcgaggaaaaattt
N I M I Q N G L L A I R P T L F E E K F
ggaccaggtgcgacgactcagcagtttaggtttggcgaagagtgcacgggccatagcagt
G P G A T T Q Q F R F G E E C T G H S S

Figure 4-1 The N-terminal sequence of GNBPA2. The starting Met is in bold. The sequence underlined with zigzag corresponds to the region matching AGAP012409 (VectorBase) (GNBPA2); the sequence underlined with line doesn't match AGAP012409.

A

```
                          [-A-]   [--B--]     [---C---]           [--C'-]
N-GNBPA2        PRKSSRYQPPKPRFEVFDPKGLIVWINADPGISSFTFHGKLNQQFVQNYDVGRWAQTIIK 60
N-PiGRP         QPRAQQYVVPSAKLEAIYPRGLRVSIPD-DGFSLFAFHGKLNEEMDG-LEAGHWARDITK 58
N-BmGRP         ------YEAPPATLEAIHPKGLRVSVPD-EGFSLFAFHGKLNEEMEG-LEAGHWSRDITK 52
N-MsGRP1        ------LEVPDAKLEAIYPKGLRVSIPD-DGYTLFAFHGKLNEEMEG-LEAGHWSRDITK 52
N-MsGRP2        -QRGGPYKVPDAKLEAIYPKGLRVSVPD-DGYSLFAFHGKLNEEMEG-LEAGHWSRDITK 57
N-DmGNBP3       ------YEVPKAKIDVFYPKGFEVSIPDEEGITLFAFHGKLNEEMEG-LEAGTWARDIVK 53
N-TmGRP         -----QFEVPDALVEVFRPRGLRVSIPDQEGIKLFAFHGKINEEMNG-REGGTFSRDILK 54
                       * . .:.: *:*: * :    * . *:****:*:::    : * ::: * *

                    [--E-]         [---F---]        [-G-][-G']
N-GNBPA2        IKNGRYLFIDREAKLVPGDTIFYRTVIVR----NGQTYRTNSGAFTVEELR--------- 107
N-PiGRP         PKEGRWTFRDRNAKLKLGDKIYFWTYVIK----DGLGYRQDNGEWTVTEFVNEDGTPADT 114
N-BmGRP         PKNGRWIFRDRNAALKIGDKIYFWTFVIK----DGLGYRQDNGEWTVEGFVDEAGNPVNT 108
N-MsGRP1        AKNGRWIFRDRNAKLKIGDKIYFWTYILK----DGLGYRQDNGEWTVTGYVNEDGEPLDA 108
N-MsGRP2        AKQGRWIFRDRNAELKLGDKIYFWTYVIK----DGLGYRQDNGEWTVTEFVNENGTVVDT 113
N-DmGNBP3       AKNGRWTFRDRITALKPGDTLYYWTYVIY----NGLGYREDDGSFVVNGYSGNNASPHPP 109
N-TmGRP         AKNGRWTFYDANARLKEGDILYYWTYVDYFDGKNKLGYPNDDQKFVVKQLLDKDG-AAPS 113
                 *:**: * *  : *  ** ::: * :      :   *  :.  :.*
```

B

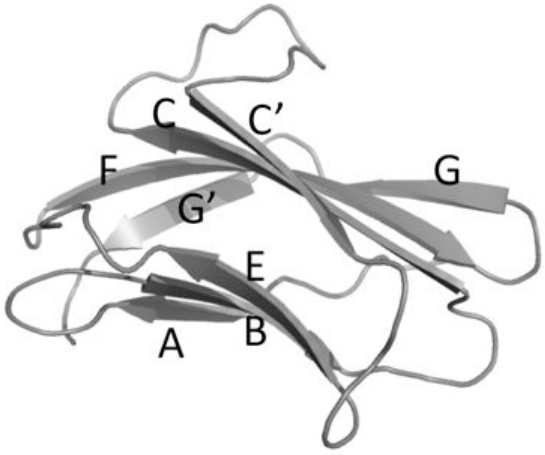


Figure 4-2 A homology-based structural model of N-GNBPA2. (A) Alignment of amino acid sequences of N-GNBPA2 and N-βGRPs/GNBPs from *M. sexta*, *P. interpunctella*, *B. mori*, *D. melanogaster*, and *T. molitor*. The β-strands of the immunoglobulin fold are labeled on top. (B) A homology SWISS-MODEL (Kiefer et al., 2009) structure of N-GNBPA2.

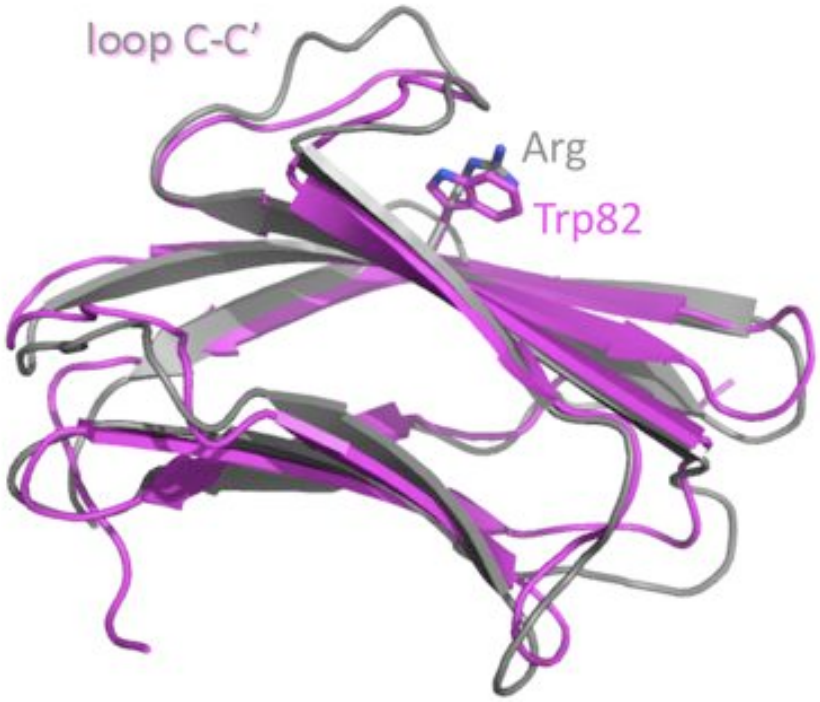


Figure 4-3 Overlay of the homology model of N-GNBPA2 (gray) with the average structure of N-βGRP (PDB code 2KHA, purple).

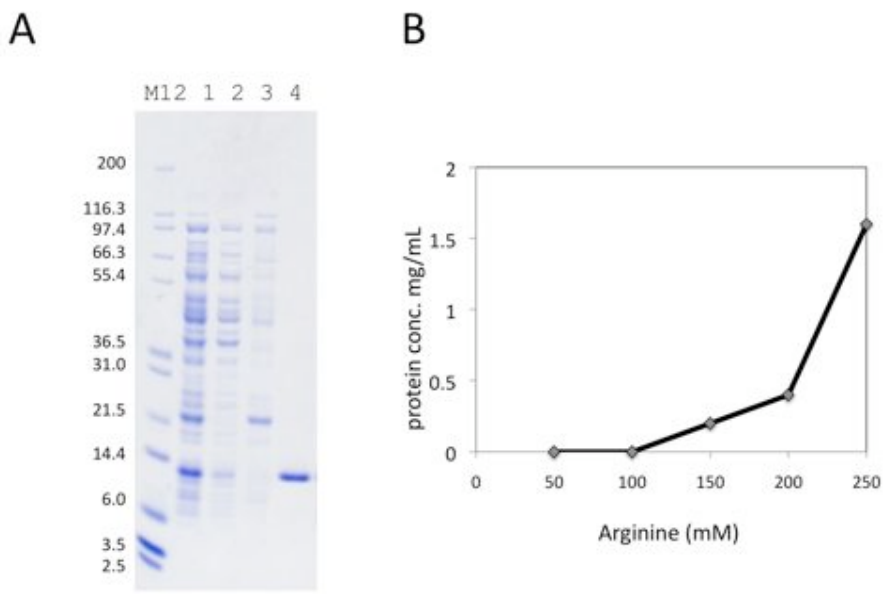


Figure 4-4 Expression and solubility of N-GNBPA2. (A) Purification of N-GNBPA2 from cleared lysate (lane 1) from cells co-expressing GroEL and GroES. Lane 2, flow through; lane 3, wash fraction; lane 4, elution fraction. (B) Solubility of N-GNBPA2 in arginine buffer (pH 8.0).

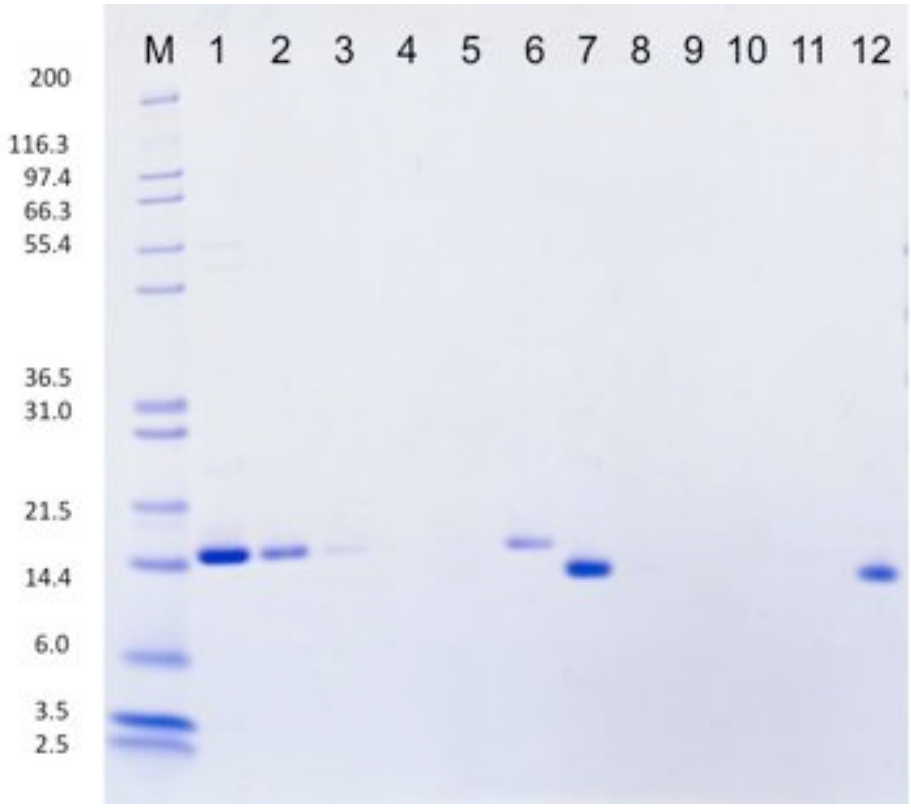


Figure 4-5 Curdlan pull-down assay of N-GNBPA2. N-GNBPA2 and N-βGRP were assayed and loaded on the gel, corresponding to lanes 1-6 and lanes 7-12, respectively. Lanes 1 & 7, loading controls; lanes 2 & 8, unbound fraction; lanes 3-5 & 9-11, three wash fractions; lanes 6 & 12, curdlan-bound fraction.

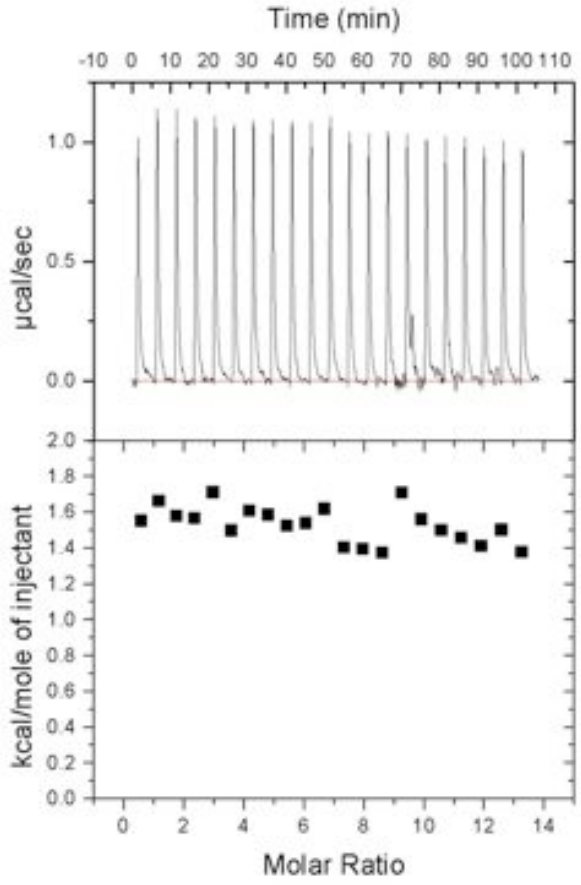


Figure 4-6 Titration of laminarin with N-GNBPA2 as monitored by ITC. The protein concentration was 38 μM and the ligand (injectant) concentration was 2.9 mM.

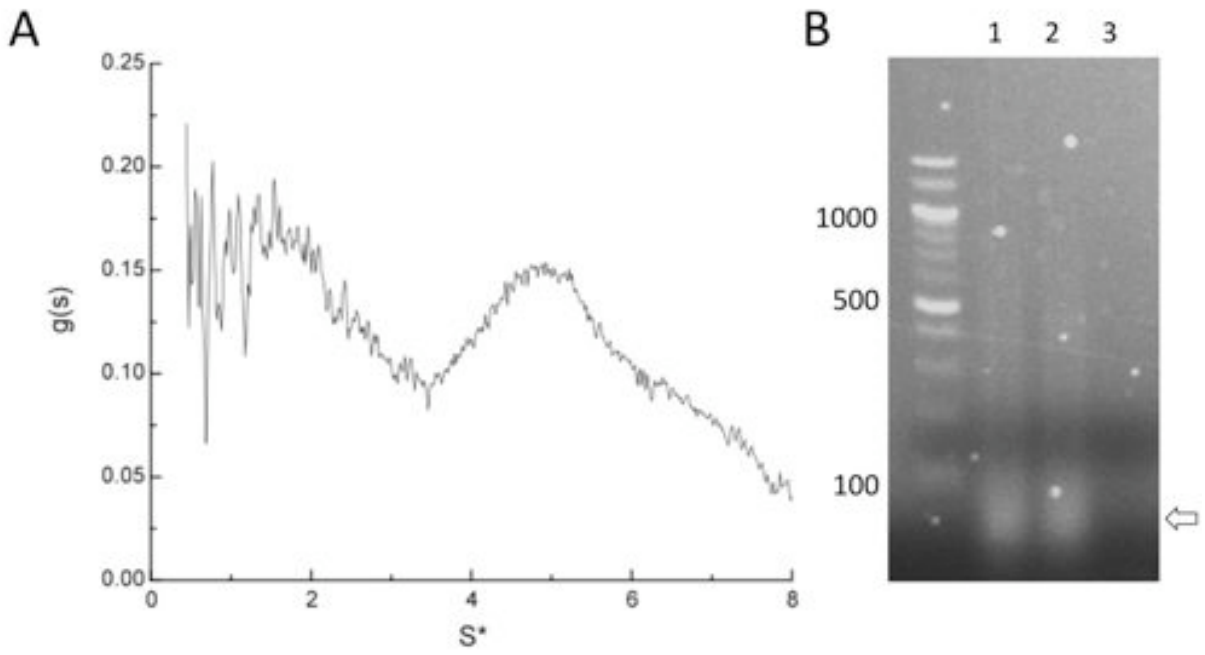


Figure 4-7 Interaction of N-GNBPA2 with DNA. (A) Sedimentation velocity analysis. The peak at ~5S* corresponds to a N-GNBPA2:DNA complex; (B) Agarose gel under UV light. Lane 1, purified N-GNBPA2; lane 2, after RNase treatment; lane 3, after DNase treatment.