Identification and characterization of a novel secreted β-prism lectin domain from *Vibrio crassostrea*

by

Simone Harrison
Class of 2018

A thesis submitted to the
faculty of Wesleyan University
in partial fulfillment of the requirements for the
Degree of Bachelor of Arts
with Departmental Honors in Molecular Biology & Biochemistry

Middletown, Connecticut
April 17, 2018
**Abstract**

Many species of the genus *Vibrio* are responsible for causing waterborne illness in humans and marine animals. A well-known example is the bacterium *Vibrio cholerae*, responsible for the disease cholera. A related species, *Vibrio crassostrea*, has caused widespread farmed oyster death in recent years. One of its proteins, which we term LPG7-BPC, is partially conserved in many *Vibrio*. It is most highly conserved (87%) in the bacterium responsible for oyster-caused gastroenteritis and death in humans, *Vibrio vulnificus*, as well as two other known pathogenic *Vibrio*: *Vibrio splendidus* (86%) and *Vibrio lentus* (84%). LPG7-BPC contains a β-prism lectin domain conserved with *Vibrio cholerae* cytolysin (VCC), a pore-forming toxin produced by *Vibrio cholerae*, and an amino-terminal secretion signal. Our lab has previously shown that beta prism lectin domains in VCC target complex N-glycans found on the surface of animal cells, suggesting that LPG7-BPC targets animal cells. We show our strategy for purifying this large and complex protein. First, the gene for the protein was split up into 7 plausible domains. These fragments were placed into plasmids, which were placed into E. coli for expression and purification. Here we show the crystallization of the LGP7 BPC beta prism lectin domain, which we term LGP7 BP, and expect to analyze its structure by X-ray crystallography. Then, we show isothermal titration calorimetry studies suggesting that LGP7 BP has different sugar binding properties than known *Vibrio* beta prism lectins. Additionally, we show secondary structure analysis of LGP7 BP by circular dichroism, substantiating its stability and beta sheet-character. Modelling studies suggest it might have a similar sugar binding pattern to the beta prism of VCC. Aside from the β-prism lectin
domain, many parts of the rest of LPG7 do not share clear homology with any other well-studied protein, suggesting that it contains a novel structural fold. Studying this new example of a secreted factor is important to understanding how LPG7-BPC interacts with human, oyster, and animal cells.
Acknowledgments

I would first and foremost like to thank Rich Olson, my research advisor and PI, for all his guidance, encouragement, and hours of explanations. Despite only having helped on a few scattered projects previously, my junior year he suggested that I start a project that I would mainly do independently. This independence caused me to grow and learn tremendously, as my dedication to my project motivated me to spend more and more time in lab, and I will always be grateful to him for giving me this opportunity. Of course, this was also made possible by the members of the Olson Lab, and Katie Kaus in particular. I have the tendency to ask questions about everything there is to ask questions about in lab, and she was always patiently there to answer all of them and guide me through many lab procedures, and offer important insights when things did not work out as expected. This thesis would not have been possible without the support of Meagan MacDonald last semester, as she helped me with many experiments, the guidance of Allison Beister on protein expression, Lorencia Chigweshe’s beginning of this project, and Charlie Visudharomn and Ken Lu’s support and guidance.

Secondly, I would like to thank Ishita Mukerji for her support and guidance as a teacher, through many office hours, and also for being a reader of thesis. I would also like to thank Don Oliver for his patient and wise advice as an academic advisor, and for also being a reader of this thesis.

Finally, I would like to thank my friends and family for their continual support and encouragement, and for always having open ears to hear me talk about my proteins. This thesis would not have been possible without them.
# Table of Contents

Abstract .......................................................................................................................... 2  
Acknowledgments ........................................................................................................ 4  
Table of Contents ......................................................................................................... 5  
List of Figures ................................................................................................................ 8  
List of Tables .................................................................................................................. 9  
Chapter 1 - Introduction ............................................................................................... 10  
  1.1 Lectins ..................................................................................................................... 11  
    1.1.1 Beta prism I fold and plant lectins ................................................................. 11  
    1.1.2 Plant Lectins as response to biotic and abiotic stress .................................... 12  
    1.1.4 Non-plant lectins as defense against pathogens ............................................. 14  
    1.1.5 Lectins in Non-Vibrio Pathogenicity ............................................................... 15  
  1.2 Vibrio cholerae disease and pathogenicity in humans ............................................ 16  
    1.2.1 VCC beta prism domain ................................................................................. 16  
    1.2.2 Beta Prism Lectin Domains in Biofilms ......................................................... 18  
    1.2.3 The specificity of sugar binding of RMBC2 and VCC ..................................... 19  
    1.2.4 Residue participation in sugar binding by VCC and RBMC2 ......................... 20  
  1.3 Beta trefoil domains ............................................................................................... 22  
  1.4 Beta prism and trefoil domains predict vibrio proteins that target eukaryotic cells: .............................................................................................................................. 23  
    1.4.1 Beta prism domain in an Invasin .................................................................... 24  
    1.4.2 LGP7 BPC Discovery ...................................................................................... 26  
  1.5 Vibrio crassostrea pathogenicity ......................................................................... 26  
  1.6 Vibrio Infection of humans ................................................................................... 28  
  1.7 Antibiotic Resistance in Vibrio cholerae and other Vibrio .................................. 29  
  1.8 Research Goals .................................................................................................... 30  
Chapter 2 - Theory and Methods ............................................................................... 31  
  2.1 Methods: Cloning, Expression, and Purification of LGP7 BPC ........................... 31  
    2.1.1 Methods: Cloning ........................................................................................... 31  
    2.1.2 Methods: Protein Expression ........................................................................ 31  
  2.1.3 Methods: Protein Purification ......................................................................... 32  
  2.2 Isothermal Titration Calorimetry ........................................................................ 32  
    2.2.1 Isothermal Titration Calorimetry Theory ...................................................... 32
2.2.2 Methods: Isothermal Titration Calorimetry ........................................ 35
2.3 Circular Dichroism .................................................................................. 35
  2.3.1 Circular Dichroism Theory ................................................................. 35
  2.3.2 Methods: Circular Dichroism ............................................................. 38
  2.3.3 Methods: BeStSel (Beta Structure Selection) ..................................... 38
2.4 X-ray crystallography ............................................................................. 40
  2.4.1 X-ray crystallography theory ............................................................... 40
  2.4.2 Methods: Crystallization (X-ray crystallography) ............................... 43
2.5 Bioinformatics and Modelling ................................................................. 44
  2.5.1 Methods: Psi-BLAST Sequence Analysis .......................................... 44
  2.5.2 Methods: Modelling—Raptor X Structure Prediction ......................... 45
  2.5.3 Methods: Modelling—Phyre2 Structure Prediction ............................. 45
  2.5.4 Methods: Modelling—SWISS-MODEL Structure Prediction ............. 46
  2.5.4 Methods: SignalP 4.0 ........................................................................ 47
Chapter 3 - Results ..................................................................................... 48
  3.1 Approach to protein expression ............................................................. 48
  3.2 Cloning and Expression ....................................................................... 50
    3.2.1 LGP7 BP Cloning ............................................................................. 50
    3.2.2 LGP7 BP + NGFP Expression ............................................................ 51
    3.2.3 LGP7 BP Purification ....................................................................... 52
  3.3 LGP7 BP Crystallization ......................................................................... 54
    3.4.2 Isothermal Titration Calorimetry Suggests that LGP7 BP does not bind to asialofetuin ................................................................. 58
  3.5 Circular Dichroism of RBMC2 and LGP7 BP ....................................... 61
    3.5.1 Calculation of thermodynamic parameters of RBMC2 and LGP7 BP ...... 62
    3.5.2 Secondary structure analysis by BeStSel shows highly similar secondary structures of LGP7 BP and RBMC2, different melting ....................... 64
  3.6 Alignment of LGP7 BP to Vibrio cholerae Beta Prisms: ......................... 72
  3.6 Modelling ............................................................................................... 73
    3.6.1 Modelling of LGP7 BP to VCC (PDB 4GX7) shows sugar binding residues are highly conserved ............................................................ 73
    3.6.2 Modelling of LGP7 BP to RBMC2 (PDB 5V6F) shows some similarities but significant differences in sugar binding residues ............................ 74
    3.6.3 Alignment of models and LGP7 BP modelled with Phyre2 shows difference in models .............................................................................. 75
Chapter 4 - Discussion ............................................................................... 77
4.1 Isothermal Titration Calorimetry Suggests that LGP7 BP may not mammalian glycans, in contrast with the model

4.2 Circular dichroism confirms that LGP7 BP is a well-folded majority antiparallel beta sheet protein, and that it is likely similar to RBMC2

4.3 LGP7 BP Crystallography

4.4 Future Directions

Chapter 5 - References
List of Figures

Figure 1.1.1 the apo form of Jacalin (PDB 1KU8)
Figure 1.1.2 Plant lectin crystal structures
Figure 1.1.3 Cry3Aa3 Toxin from Bacillus thuringiensis tenebrionis
Figure 1.1.4 ZG16p and M-ficolin crystal structures
Figure 1.2.1 VCC Overall
Figure 1.2.2 role of biofilms in Vibrio cholerae infection
Figure 1.2.3: Glycan specificity of VCC and Rbmc2
Figure 1.2.4.1 VCC with methylmannose bound
Figure 1.2.4.2 Crystal structure of RBMC2 bound to trimannose
Figure 1.3 Vibrio vulnificus hemolysin Beta Trefoil domain
Figure 1.4 Beta Prism lectin domains and beta trefoil lectin domains in Vibrio
Figure 1.4.1 Invasin from Yesinia pseudotuberculosis
Figure 1.4.2 Breakdown of LGP7 by amino acids and putative domains
Figure 1.5.1: Vibrio crassostrea strains involved in oyster mortalities
Figure 1.5.2: plasmid pGV1512
Figure 1.8: Signal P prediction of LGP7 BPC secretion signal, residues 1-23
Figure 2.2.1: ITC diagram
Figure 2.3.1: Secondary structure determination by BeStSeL
Figure 2.3.1 characteristic peaks of various secondary structures for CD
Figure 2.4.1 X-ray crystallography set up and unit cells
Figure 3.1.1: PSIPRED secondary structure, binding, and domain boundary prediction of LGP7 BPC with secretion signal (Jones, 1999)
Figure 3.1.2: Psi-Blast boundary prediction based on known matches
Figure 3.1.3: predicted structure of full length LGP7-BPC by Raptor X
Figure 3.2.1 Cloning of LGP7 BP
Figure 3.2.2: Expression of LGP7 BP
Figure 3.2.3: Purification of LGP7 BP
Figure 3.3: small hexagonal LGP7 BP crystals
Figure 3.4.1.1: LGP7 BP + NGFP off the sizing column, before ITC
Figure 3.4.2.2 Origin 7 analysis of the ITC of LGP7 BP + GFP and asialofetuin
Figure 3.5.1 Changing of ellipticity over 20-96ºC for LGP7 BP
Figure 3.5.2 changing ellipticity over 20-96ºC for Rbmc2
Figure 3.5.1.1 Thermodynamic parameters of LGP7 BP at the peak at 233 nm
Figure 3.5.1.2 Thermodynamic parameters of LGP7 BP
Figure 3.5.2.1 Secondary structure of LGP7 BP and RBMC2 at 20ºC
Figure 3.5.2.2 Secondary structure of LGP7 BP and RBMC2 at 50ºC
Figure 3.5.2.3 Secondary structure of LGP7 BP and RBMC2 at 96ºC
Figure 3.5.2.4 Comparison of change in secondary structure of LGP7 with RBMC2
Figure 3.6 alignment (with mega) of known beta prism domains.
Figure 3.6.1 Alignment of SWISS MODEL model of LGP7 BP to VCC
Figure 3.6.2 Alignment of SWISS MODEL of LGP7 BP to RBMC2
Figure 3.6.3 Homology models of LGP7 BP aligned
Figure 4.2: Secondary structure changes of LGP7 BP over 20-96 ºC
List of Tables

**Table 1:** JSCG + screen results of promising conditions of LPG7 BP (those with crystalline like precipitation).

**Table 2:** Crystal screen around the condition 1 M Succinic acid, 1 M HEPES, and 1% PEG 2K MME. 1 = Precipitate 0 = clear

**Table 3:** Thermodynamic parameters determined by CD melting of beta prisms RBMC2 and LGP7 BP

**Table 4:** Total BeStSel secondary structure analysis of LGP7 BP, from 20°-96°C (Micsonai et al., 2015)

**Table 5:** Total BeStSel secondary structure analysis of Rmbc2 from 20°C-96°C (Micsonai et al., 2015)

**Table 6:** direct comparison of changes in secondary structure upon melting, between LGP7 and Rmbc2, analyzed by BeStSel (Micsonai et al., 2015). Data also shown from De et. al 2018
Chapter 1 - Introduction

This introduction seeks to provide a background on all we know about beta prism lectin domains and their roles in infection and pathogenicity. First, it will define lectins generally, then examine the discovery of the beta prism lectin fold I in the jacalin, thus leading to the designation of protein with this fold as “jacalin-like”. Next, it will explore the structures and roles of various lectins as they are related to plant immune response and bacterial pathogenicity. Specifically, the role of two known beta prism lectin domains in *Vibrio cholera*, the bacterium responsible for the disease cholera, will be described in detail. Next, I will focus on the discovery of the beta prism lectin domain which we characterize in this study, LGP7 BPC, and then examine the bacteria it comes from, *Vibrio crassostrea* strain LGP7. Then, a more general background on the pathogenicity of *Vibrio* bacteria in humans will be given. Finally, the research goals of this thesis will be stated.
1.1 Lectins

Lectins are defined as proteins that bind carbohydrates (sugars). All lectins are divided into two groups based on what kind of sugars they bind: lectins that bind oligosaccharides and monosaccharides, and lectins that only bind oligosaccharides. Of the oligosaccharides that lectins can bind, they usually bind either N-glycans or O-glycans (Nilsson, 2007). N-glycans are sugars attached to proteins found on the outside of eukaryotic cells. They share a common core of 2 N-acetylglucosamine, 3 glucose, and 9 mannose (Stanley et. al, 2009). Of the lectins that can bind monosaccharides, they tend to fall into 5 groups: those that bind either glucose/mannose, those that bind galactose/N-acetylgalactosamine, those that bind fucose, those that bind N-acetylglucosamine, or those that bind sialic acid (Nilsson, 2007). Lectins bind these carbohydrates without modifying them. Some lectins tend to be oligomeric or multidomain in order to increase their glycan binding avidity (Nilsson, 2007).

1.1.1 Beta prism I fold and plant lectins

Beta prism lectin domains were first discovered in the protein jacalin, which was purified from the seeds of jackfruit (Artocarpus integrifolis), and is a galactose and mannose specific lectin, made up of two separate chains (alpha and beta). This specificity is unusual because, as stated above, monosaccharide binding lectins are usually glucose/mannose or galactose/N-acetylgalactosamine specific. This flexibility of binding is attributed to the unusually large sugar binding site of jacalin, allowing it to interact with these monosaccharides despite their different hydroxyl conformations (Bourne et al., 2002). The main role of jacalin is to bind T-antigen disaccharide
Galβ1,3GalNAc, which is thought to be an immune response to viral infection.

Interestingly, the T-antigen is an antigen often present on human cancerous cells (Esch & Schaffrath, 2017). Jacalin has been shown to also be a stimulator of T-cells in humans (Bunn-Moreno & Campos-Neto, 1981). The family in which the hypothetical protein LGP7 BP is found is called the “jacalin-like superfamily”. This family is characterized by the beta-prism I fold, which contains three Greek key beta sheets (four stranded), with the strands parallel to 3-fold axis. (Sankaranarayanan et al., 1996).

![Figure 1.1.1 A) the apo form of Jacalin, the first beta prism lectin domain purified, as an example of a beta prism I fold. (PDB 1KU8) (Bourne et al., 2002)](image)

### 1.1.2 Plant Lectins as response to biotic and abiotic stress

The role of plant lectins is not always known, but they often participate in the recognition of glycoproteins, either intracellularly as a sorting mechanism of carbohydrates, which are usually in the form of N-linked glycoproteins, or extracellularly as an immune response to the carbohydrates on the outside of cells (Nilsson, 2007). Many organisms have specific glycoproteins on the outside of their cells, rendering them susceptible to lectin binding and immune identification of this
type. The *Parkia platycephala* seed lectin, which binds mannose, has been shown to cross link carbohydrates in its hexamer form in order to inactivate pathogens (Gallego del Sol et al., 2005). Beta prism I fold lectin domains are also often proteins expressed in plants during salt stress, such as the SALT protein from *Oryza sativa* (Jiang et al., 2006). This protein binds mannose and forms a dimer (Jiang et al., 2006).

![Figure 1.1.2](image)

**Figure 1.1.2** A) the hexamer form of Parkia platycephala seed lectin (PDB ID 1ZGS) (Gallego del Sol et al., 2005) B) the dimer form of the SALT protein from *Oryza sativa*, bound to methyl-mannose (PDB ID 5GVY) (Jiang et al., 2006)

### 1.1.3 Crystal toxin uses beta prism lectin domain for insecticidal activity

The well-known crystal (Cry) toxin of *Bacillus thuringiensis* is a pore forming toxin with insecticidal activity (Xu et al., 2014). This toxin is so specific for insect cells that introduction of this toxin into the genome of plants has been suggested as a strategy creating insect resistant transgenic food crops, as the toxin will not bind to human cells (Xu et al., 2014). This toxin contains three domains. Domain I is a seven helix bundle that is thought to form the transmembrane lytic pore. Domain II contains
the beta prism domain, which is thought to be important for this highly specific recognition (and consequent lysis) of insect cells. The short helix between strands beta 1 and 2 is also found in ZG16p protein, discussed below. The exact sugar binding mechanism is not known. Domain III might also be involved in insect cell recognition (Xu et al., 2014).

1.1.4 Non-plant lectins as defense against pathogens

Many other non-plant eukaryotes use lectins as defense against pathogens. Horseshoe crabs secrete proteins called tachylectins that recognize the saccharides on the surface of pathogens: for example, LPS (lipopolysaccharide) or N-acetylglucosamine and N-acetylgalactosamine (Nilsson, 2007) (Beisel, Kawabata, Iwanaga, Huber, & Bode, 1999). In mammals, there are lectins called collectins and ficolins that can bind to oligosaccharides on the surface of microorganisms and create an immune response (Nilsson, 2007). They recognize these carbohydrates through their C-terminal domains (Garlatti et al., 2007). Other mammalian lectins include the

Figure 1.1.3 Cry3Aa3 Toxin from Bacillus thuringiensis tenebrionis (Li et al., 1991). Domain I is purple, domain II is blue, and domain III is green.

Non-plant lectins as defense against pathogens
jacalin-related human zymogen granule protein 16 (ZG16p), which is the first beta prism lectin to be found in humans (Kanagawa et al., 2014). It binds short ligands that end in mannose preferentially. This protein’s function is unknown, but it might mediate the binding of aggregated zymogens to the granule membrane in the pancreas (Kanagawa et al., 2014). Protein WGA16 (see Results) is a homologue of this protein (Garéniaux et al., 2015).

1.1.5 Lectins in Non-Vibrio Pathogenicity

Pathogens may also produce lectins which mediate pathogen cellular adhesion and invasion to host cells through binding to host glycoproteins, and therefore are important when considering medicine against these pathogens. *Helicobacter pylori*, which infects humans and causes ulcers, has two known adhesins (Kusters et. al, 2006). The hemagglutinin protein is a lectin found on the surface of the flu virus and binds host cell sialic acids (Nilsson, 2007).
1.2 Vibrio cholerae disease and pathogenicity in humans

The toxins that have been primarily implicated in pathogenicity of the epidemic and virulent Vibrio cholera strain O1, El Tor biotype, are the toxin-co-regulated pili (TCP) and the cholera toxin (CT) (Zhang et al., 2014). CT, an enterotoxin, is thought to cause the diarrheal symptoms of the disease (Cinar et al., 2010). This toxin is activated through quorum sensing, and subsequently activates cyclic adenosine monophosphate. This increases Cl⁻ secretion and decreases NaCl⁻ coupled absorption (Morris & Acheson, 2003). This efflux of Cl⁻ leads to an osmotic release of water from cells, leading to the watery diarrhea characteristic of cholera, and consequent severe dehydration that is not remediable by water and salt (though water supplemented with glucose does work, as glucose channels remain unaffected.) (Morris & Acheson, 2003). This diarrhea can lead to death rapidly without proper rehydration. An estimated 3-5 million cases occur worldwide annually with around 100,000 deaths, according to the World Health Organization (WHO). However, TCP and CT deficient strains are still able to cause acute diarrheal disease in humans and lethality in C. elegans, highlighting the importance of other virulence factors (Cinar et al., 2010). One of these factors is thought to be Vibrio cholerae cytolysin VCC (Cinar et al., 2010). VCC is even required for lethality in C. elegans.

1.2.1 VCC beta prism domain

VCC has a well characterized beta-prism lectin domain. It is vital to cellular adhesion and consequent host cell lysis through pore formation of the VCC pore forming toxin. Pore formation is thought to follow the following steps: a VCC monomer binds to the cell membrane through N-glycan interaction with the beta
prism domain and direct membrane interaction with the rim domain of VCC (De et al., 2015). Monomers then diffuse across the cell membrane and form oligomeric pre-pore intermediates before entering the pore state, which lyses the cell (De et al., 2015). Mutation of residues in the beta prism domain leads to up to 50 fold loss of hemolysis, demonstrating the importance of this domain in virulence (De, Kaus, Sinclair, Case, & Olson, 2018). However, other domains, such as the stem and cytolytic core, are essential to host cell attachment, pore formation, and consequent lysis. The Olson lab has characterized this protein’s affinity for different mammalian N-glycans, and found it binds certain N-glycans highly specifically, suggesting it might target neutrophils (De et al., 2018). Lysis of neutrophils allows *Vibrio cholerae* to evade the host immune system.

Figure 1.2.1 A) VCC in its oligomeric pore and monomer form, and the location of the beta prism lectin domain (De et al., 2018) and (Levan, De, & Olson, 2013) B) VCC bound to methyl-mannose (PDB 4GX7) (Levan et al., 2013)
1.2.2 Beta Prism Lectin Domains in Biofilms

Beta prism lectin domain specificity for mammalian N-glycans has also been found in the biofilm matrix proteins Rbmc1 and Rbmc2 (Rugosity and Biofilm structure Modulator C), as well as Bap1 (Biofilm associated protein 1), in the Vibrio cholerae genome (De et al., 2018). A biofilm is matrix that surrounds a group of bacteria, using containing an exopolysaccharide (for Vibrio cholerae, VPS), several matrix proteins, and other small molecules (Yildiz & Visick, 2009). Biofilms help bacteria attach to surfaces and tissues and help protect the bacteria against things like stomach acid and unfavorable conditions. Biofilms have been proven to be important in the transmission of cholera, through increasing the bacterium’s resistance to stressful conditions in the host organism. Vibrio cholera is thought to form biofilms during infection (Silva & Benitez, 2016). Consequently, a lower amount of the bacteria is needed to produce symptoms of infection when the bacteria is in a biofilm (Silva & Benitez, 2016). Finally, biofilms may increase the propagation of the disease through the excretion and ingestion (Silva & Benitez, 2016).

![Diagram of biofilm formation](image)

Figure 1.2.2 A) role of biofilms in Vibrio cholerae infection, (Silva & Benitez, 2016)
1.2.3 The specificity of sugar binding of RMBC2 and VCC

The binding affinities of RBMC2 and VCC to the sugars listed in the Table 1 suggest that they have the strongest affinities for the NGA2 heptasaccharide core, and the pentasaccharide core, respectively, with binding affinities of 2 µM and 1.5 nM, respectively (De et al., 2018). Strong binding to asialofetuin, a fetal serum glycoprotein, is a good indicator of binding mammalian glycoproteins (De et al., 2018). NGA2 is a human glycoprotein that is often found on Immunoglobin G, as well as other cells, suggesting that VCC might target neutrophils as a way to evade the human immune systems (Rohrer, Basumallick, & Hurum, 2016). RMBC2’s decreased specificity for this glycan, and strong binding overall, suggest that its sugar binding could serve a different function.

<table>
<thead>
<tr>
<th>Glycan/Sugar</th>
<th>VCC β-prism affinity</th>
<th>RbmC2 β-prism affinity</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-acetylglucosamine</td>
<td>2.7 mM (2.6, 2.8)</td>
<td>178.7 µM (89.6, 268.0)</td>
</tr>
<tr>
<td>LacNAc</td>
<td>Did not bind</td>
<td>Did not bind</td>
</tr>
<tr>
<td>GlcNAc-Man</td>
<td>--</td>
<td>158.8 µM (138.9, 162.1)</td>
</tr>
<tr>
<td>Mannotriose</td>
<td>228.6 µM (224.5, 240.6)</td>
<td>4.1 µM (4.0, 5.6)</td>
</tr>
<tr>
<td>Pentasaccharide</td>
<td>--</td>
<td>1.1 ± 0.6 nM</td>
</tr>
<tr>
<td>NGA2</td>
<td>0.2 ± 0.1 µM</td>
<td>1.5 ± 0.5 nM</td>
</tr>
<tr>
<td>NA2</td>
<td>1.0 ± 0.1 µM</td>
<td>--</td>
</tr>
<tr>
<td>A2</td>
<td>1.1 ± 0.4 µM</td>
<td>--</td>
</tr>
<tr>
<td>Asialofetuin</td>
<td>1.5 µM (1.5, 3.6)</td>
<td>18.4 nM (9.9, 29.3)</td>
</tr>
</tbody>
</table>

Figure 1.2.3: Glycan specificity of VCC and Rbmc2, from either ITC (blue) or intrinsic tryptophan fluorescence spectroscopy (red) (De et. al 2018)
1.2.4 Residue participation in sugar binding by VCC and RBMC2

There is an essential aspartic acid in both RMBC2 (D853) and VCC (D617) that mediates hydrogen bonding the sugar. This residue is found to be conserved across many beta prism domains (De et al., 2018). The crystal structure of VCC bound to methyl-mannose suggests that the other VCC residues that participate in direct interactions with the sugar are A614, G632, A615, and Y654 (Levan et al., 2013). These interactions are between the carboxyl group of D617 and the sugar, and the backbone amide nitrogens of Gly632, A614, and A615, and the sugar. The Y654 participates in a Van der Waals interaction with the methyl group of this sugar (Levan et al., 2013).

![Figure 1.2.4.1](image)

Figure 1.2.4.1 A) crystal structure of VCC beta prism with Me-alpha-mannose bound in its heptamer form (Levan et al., 2013). B) A close up of the sugar binding site with hydrogen bonds shown. Os represent the oxygens in the sugar. (Levan et al., 2013)

From the crystal structure of RMBC2 to trimannose, the residues that are thought to form direct hydrogen bonding with the sugar are W896, W948, N871, F859, and Y894, due to their distance in the crystal structure being less than 3.35 Å (De et al., 2018). The PVQGT loop containing N871 loop is mostly unstructured in the apo crystal structure of RBMC2, and ordered in the ligand bound to GlcNAc-
Man, suggesting that binding of the sugar leads to ordering of this group, potentially mediated by the water molecule coordinated by N871 and W948 (De et al., 2018).

This model of sugar binding provides us with an explanation for why RMBC2 has a greater affinity for sugars than VCC. F850 and W896 participate in Van der Waals and ring stacking interactions with the sugars, leading to increased affinity. Instead of these residues, VCC has A614 and Y654, respectively (De et al., 2018). Though tyrosine can still participate in aromatic ring stacking, tryptophan’s increased surface area leads to a stronger interaction. Additionally, Y894 forms a hydrogen bond with the central sugar in trimannose bound to RBMC2. VCC has a phenylalanine instead of this residue, which cannot participate in hydrogen bonding.

Figure 1.2.4.2 A) crystal structure of RBMC2 bound to trimannose B) hydrogen bonding interactions of RMBC2 and trimannose C) crystal structure of RMBC2 bound to GlcNAc-Man D) hydrogen bonding interactions of RBMC2 and GlcNAc-Man. (De et al., 2018)
1.3 Beta trefoil domains

A potential evolutionary precursor to beta prism lectin domains are beta trefoil domains (Kaus, Lary, Cole, & Olson, 2014). The beta trefoil fold consists of six two stranded hairpins, three of which are in a beta barrel formation. The other three hairpins form a triangular cap to the barrel (Murzin, Lesk, & Chothia, 1992). VCC contains a known inactive beta trefoil domain in addition to its beta prism lectin domain. *Vibrio vulnificus* is a bacterium responsible for many cases of seawater caused food poisoning and septicemia in humans (J. G. Morris & Acheson, 2003). Examination of the beta trefoil lectin domain of the *Vibrio vulnificus* hemolysin (VVH) shows that this beta trefoil domain also targets N-glycans, but with lower affinity (micromolar vs. nanomolar affinity) and specificity than the beta prism lectin domains (Kaus et al., 2014). VVH shows similar affinity for many N-glycans on a glycan chip screen in contrast with VCC, which shows strong affinity for only a select few (Kaus et al., 2014). Accordingly, VVH shows less preference for intestinal colonization than VCC, as demonstrated by its propensity for causing septicemia. VCC contains a nonfunctional beta prism trefoil domain, suggesting that the beta prism domain replaced this domain as a way of targeting the toxins to the host organism’s cells (Kaus et al., 2014).
1.4 Beta prism and trefoil domains predict vibrio proteins that target eukaryotic cells:

Figure 1.3 *Vibrio vulnificus* hemolysin Beta Trefoil domain, bound to N-Acetyl-D-Lactosamine. Forms a heptamer in the crystal structure. PDB ID 4OWL (Kaus et. al 2014)

Figure 1.4 Beta Prism lectin domains and beta trefoil lectin domains are found in many *Vibrio* cytolysins and hemolysins. (Kaus et. al 2014)
As evidenced by this phylogenetic tree, the incidence of beta trefoil domains and beta prism lectin domains can predict where cytotoxic proteins exist in the genomes of many *Vibrio*. A PSI-BLAST search with a beta prism domain previously characterized in this lab, RBMC2, found two novel proteins containing beta prism domains.

1.4.1 Beta prism domain in an Invasin

The first protein we found was an invasin in *Yesina pekkaneni*. Invasins have been shown to be essential to M cellular uptake of *Yersinia (pseudotuberculosis* and *entercolitica)* in humans (Hamburger, Brown, Isberg, & Bjorkman, 1999). This transports bacteria across the intestinal epithelium, where they then colonize the liver and spleen and cause disease. Invasins bind β₁ integrins, which is then thought to cause the cytoskeleton of the host cell to rearrange to bring the bacteria into the cytoplasm. Invasins outcompete fibronectin, the native binding partner of integrins. Integrins themselves are integral membrane proteins that are essential to cellular communication by binding various components of the cytoskeleton to extracellular proteins like extracellular matrix proteins or cell surface proteins. The invasin in *Yersinia pseudotuberculosis* contains 5 domains, D1-D5 (Hamburger et al., 1999). D4 and D5 are implicated in integrin binding. In the novel protein we found in *Yersinia pekkaneni* through PSI-BLAST, D4 and D5 are replaced with a putative beta prism lectin domain. This suggests that the invasin could be using a novel method of host cell bind cell binding, and that beta prism lectin domains have a broader role in pathogenicity than previously identified.
Figure 1.4.1 Invasin from *Yersinia pseudotuberculosis*, domains shown (Hamburger et al., 1999). In *Yersinia pseudotuberculosis*, D4 and D5 are replaced with a putative beta prism lectin domain.
1.4.2 LGP7 BPC Discovery

The second protein that we found through PSI-BLAST is the protein that is the subject of this thesis, which we term LGP7 BPC, for the bacterial strain in which it was found, *Vibrio crassostrea* LGP7 and the fact that it is beta-prism containing. Except for the beta prism domain, which we term LGP7 BP, and a putative secretion signal, the rest of the protein’s structure cannot be predicted by traditional homology models and is dissimilar to any other protein previously studied, suggesting it has a novel structure and function. Discovering the structure and the carbohydrate specificity of this beta prism could elucidate the function of this novel hypothetical protein.

![Figure 1.4.2 Breakdown of LGP7 by amino acids and putative domains](image)

1.5 Vibrio crassostrea pathogenicity

*Vibrio crassostrea* has also been of interest lately due to several recent mass farmed oyster mortalities. A recent study has found that populations of different species and subspecies of *Vibrio* are responsible for the death of these oysters, and that virulence is localized around a certain subset of genes (Lemire et al., 2015). These genes are necessary, but not sufficient, for pathogenicity, suggesting that there are other genes spread throughout the genome that are important for survival of the bacteria in the hemolymph of the oysters, evading its immune system (Lemire et al., 2015). The particular strains involved in virulence are thought to be J5-4, J5-19, J5-5,
J2-9, J2-13, J5-15, LGP8, LGP7, LGP15, J5-23, J5-24, J5-28, J5-20, J5-6, LGP107, and LGP108 (Lemire et al., 2015), shown in Figure 1.5.1. The protein described in this document is from strain LGP7, a strain which has been isolated from more than one farmed oyster mortality event in the past (Lemire et al., 2015).

A later study hypothesized that, due to the fact that *Vibrio crassostrea* was the only bacteria that inhabited oysters at a significantly higher percentage than the seawater when the ocean temperatures reached above 16°C (which is when the oyster mortalities were observed to happen), and the fact that it was observed to cause oyster mortalities in vitro, *Vibrio crassostrea* is the main bacteria involved in oyster pathogenesis (Bruto et al., 2017). It also theorized that the virulence of *Vibrio crassostrea* strains is correlated to the presence of a large mobilizable plasmid, pGV1512, containing hypothetical virulence factors such as a type six secretion system (T6SS) domain, a type four secretion system (T4SS) domain, a heavy metal resistance domain (Cus/Cop), among other genes (Bruto et al., 2017). The deletion of
those 3 mentioned domains did not result in virulence attenuation while deletion of various regions in between those three domains did. However, no other known virulence factors were found in the plasmid, suggesting that the virulence factor is novel (Bruto et al., 2017). The LGP7 BPC gene is not in this plasmid: however, this plasmid is not sufficient for virulence, suggesting that LGP7 BPC could play a role in host organism infection.

Figure 1.5.2: plasmid pGV1512 with its hypothetical domains, and showing its presence causes a significant increase in mortality of Crassostrea gigas oysters (Bruto et al., 2017)

1.6 Vibrio Infection of humans

With rising ocean temperatures, there has also been interest in the increased incidence of Vibrio infection of humans (Morris & Acheson, 2003). Usually, this infection is caused by Vibrio vulnificus or Vibrio parahaemolyticus, but many other Vibrio have also been found to infect humans. Symptoms can include gastroenteritis as well as fever, and can even lead to septicemia if the infection spreads to the
bloodstream (Morris & Acheson, 2003). The increased incidence of septicemia in *Vibrio vulnificus* infection is thought to be caused by the fact that its hemolysin, as described above, has a beta trefoil domain with less specificity for certain sugars, allowing it to lyse cells anywhere in the body (Kaus et al., 2014). *Vibrio vulnificus* caused 124 cases of infection in the United States in 2014, the last year on record, 97 hospitalizations and 21 deaths, according for the Center for Disease Control and Prevention (CDC). These infections were most prevalent in summertime and with consumption of raw shellfish, particularly oysters. *Vibrio vulnificus* can also cause infections through wounds exposed to sea water (Morris & Acheson, 2003).

Pathogenic *Vibrio* infection of marine animals is also common. *Vibrio lentus* is known to cause lesions and disease in octopuses, often with bacterial colonization of the internal organs of the octopus and consequent mortality (Farto et al., 2003). *Vibrio splendidus* has recently been linked to lesions and mass mortality of aquacultured carpet shell clam (*Ruditapes decussatus*) (Gómez-León, Villamil, Lemos, Novoa, & Figueras, 2005). *Vibrio vulnificus* and *Vibrio splendidus* have been found to have the LGP7 BPC hypothetical protein with 87% and 86% identity, and *Vibrio lentus* contains the protein with 84% identity.

**1.7 Antibiotic Resistance in Vibrio cholerae and other Vibrio**

The virulence of *Vibrio* is greatly increased due to the ease in which these bacteria share virulence factors and antibiotic resistance factors. In *Vibrio cholerae*, antibiotic resistance usually arises through SXT elements. SXT elements are mobile DNA elements in the class of integrative conjugating elements for a conjugative transposons (Kitaoka, Miyata, Unterweger, & Pukatzki, 2011). These SXT elements
can spread through horizontal gene transfer, or even through conjugative plasmids or genomic islands *in trans* (Kitaoka et al., 2011). Outside of SXT elements, mobile integrons can spread antibiotic resistance (Kitaoka et al., 2011).

### 1.8 Research Goals

In this thesis, the beta prism lectin domain, which we term LGP7 BP, was targeted for structural and function studies to examine its role and determine if it could bind human or animal cells. The presence of the beta prism lectin domain in combination with a secretion signal (shown) suggest that it is likely secreted by the bacteria to effect a host organism, and we seek to characterize this interaction.

![Signal P prediction of LGP7 BPC secretion signal, residues 1-23](image)

Figure 1.8: Signal P prediction of LGP7 BPC secretion signal, residues 1-23
Chapter 2 - Theory and Methods

2.1 Methods: Cloning, Expression, and Purification of LGP7 BPC

2.1.1 Methods: Cloning

Full length LGP7-BPC and its fragments were cloned out of the purified genomic DNA of the LGP7 strain, provided by Frédérique Le Roux, with Phusion High Fidelity DNA Polymerase from New England Biolabs. These PCR products were purified using the Wizard SV PCR Clean-up system, and were digested for 2 hours at 37°C with enzymes XhoI and BamHI. This digestion was gel purified using a Wizard SV Gel Cleanup System. The fragments were then ligated into the pNGFP-BC plasmids using T4 ligase overnight at 16°C, with a 3 to 1 insert to plasmid ratio. The plasmid was also digested and gel purified. pNGFP-BC attaches a cleavage N-terminal His-tag and GFP to the protein. It also contains a number of digestion sites as well as a T7 promoter, T7 terminator, and stop codon, as well as a region encoding ampicillin resistance (Kawate & Gouaux, 2006).

2.1.2 Methods: Protein Expression

For every liter of expression culture, 20 mLs of Luria Broth culture with 100µg/mL of carbenicillin was inoculated from glycerol stocks of the T7 express cells containing the pNGFP-BC LGP7 BP plasmids, and grown overnight at 37°C with shaking. This overnight culture was added to larger liters of Luria broth at 20 mL overnight/liter, with 1 mL of 50 mg/mL carbenicillin per liter of Luria broth. These liters were grown at 37°C with shaking for 3.5 hours to an optical density of ~0.6, then induced for expression with 1mL 1M IPTG per liter, then grown for four
hours at 30°C with shaking. Cells were pelleted using Thermo Scientific Sorvall Lynx 6000 centrifuge at 9,000 RPM for 10 minutes. The cells were frozen at -80°C.

2.1.3 Methods: Protein Purification

Cells were then thawed, and resuspended in 9 mL TBS per liter, with protease inhibitor cocktail and 10 mM imidazole. Samples were then lysed by homogenization, and centrifuged at 18,000 RPM in Thermo Scientific Sorvall Lynx 6000 centrifuge for 25 minutes to separate cell matter (pelleted) from soluble protein in the supernatant. The supernatant is then passed over a 5mL Ni NTA-chelating column in TBS pH 7.5 using a Biorad Profina protein purification system. The protein attached to GFP sticks to this column to due to its His-tag. The column is washed with 10 column volumes of 40 mM imidazole in TBS buffer, then eluted with 3 column volumes of 250 mM imidazole in 20 mM Tris, 150 mM NaCl, pH 7.5 (referred to TBS from here on, excluding the imidazole) buffer, and passed through a 50mL desalt column so the final elution is in TBS pH 7.5 buffer. The protein is then cut with thrombin at 1:250 w/w thrombin:protein, and then run over a GE Healthcare Size Exclusion machine on a Superose 6 10/300 GL column, in TBS buffer. The fractions ~B7-B11 are collected and concentrated to get a pure sample of the beta prism domain (after around 1.1 column volumes) as confirmed using SDS-PAGE gels stained with commassie brilliant blue G-250.

2.2 Isothermal Titration Calorimetry

2.2.1 Isothermal Titration Calorimetry Theory

Isothermal titration calorimetry (ITC) is a highly sensitive way of measuring binding affinities. It does this through measuring the amount of heat released when
incremental amounts of ligand are added to the protein of interest. The protein of interest is placed in the sample cell, while a syringe is loaded with the ligand and placed above it. The protein of interest and the ligand must be in the same buffer in order to negate the energy released when buffers are diluted or mixed (Sheehan, David, 2009). The amount of heat released is measured due to the fact that the machine is constantly comparing a sample cell to the reference cell, which is internal to the machine and at a constant temperature (Figure 2.2.1). The machine cools or heats the sample cell in order for it to be the same temperature as the reference cell. This measures how much heat the cell absorbs or releases as the binding event occurs (Sheehan, David, 2009).

The binding curve is then extrapolated from the heat gain or loss which is needed to keep the sample chamber the same temperature as the reference chamber, which is shown as $\mu$cal/sec against time (Sheehan, David, 2009). This heating or cooling gets smaller over time due to the binding saturating. When the concentration

Figure 2.2.1: The ITC machine and how a graph of the experimentally determined data can give us the binding data. (“Isothermal Titration Calorimetry to study biomolecular interactions,” n.d.)
of the solutions is also input into the program, a sigmoid can be fit to the kcal/mol vs. molar ratio of the ligand and the solution (Figure 2.2.1). The change in enthalpy can be determined from this graph of kcal/mol vs molar ratio: the overall kcal released per mol is the change in enthalpy. The $K_a$ or $K_d$ can be determined by estimating the slope of the steepest part of the curve (Figure 2.2.1). This is due to the fact that this steep part of the curve occurs when the proteins binding sites are filling up and the ligand is both associating with and dissociating from the protein. The $n$, or stoichiometry of binding, can be determined from the molar ratio value at the steepest part of the curve. This is due to the fact that this drastic reduction in the release of heat means that the binding sites are filled up. The equations $\Delta G = -RT \ln(K_a)$ and $\Delta G = \Delta H - T \Delta S$ can be used to relate the parameters measured from experiment to the free energy of binding (Sheehan, David, 2009), where $\Delta G$ is change in free energy, $K_a$ is the association constant, $\Delta H$ is the change in enthalpy, $T$ is temperature, and $\Delta S$ is the change in entropy.

The free energy of binding ($\Delta G$) can tell us about the favorability of the binding event. A negative $\Delta G$ means that the binding is favorable. A positive $\Delta G$ means that the binding event is unfavorable. The equations above tell us that the enthalpy, entropy and temperature of binding determine this free energy of binding. Enthalpy describes the energy of the binding event in terms of the bonds formed or broken. Bond breaking requires energy, whereas bond formation releases energy. Therefore, one might expect a binding event to have a negative enthalpy (meaning energy is released), adding to the favorability of the reaction. Entropy ($\Delta S$) also determines the favorability of the reaction. Entropy is a way of describing the amount
of disorder in a system, or the number of microstates it can have. The more disorder that is present (or the greater quantity of microstates) a system can have, the more favorable it is. Therefore, an increase in entropy means an increase in free energy. One might expect a binding event to decrease the entropy, due to the restricted flexibility of two things bound together, and the decreased number of microstates they can form. The free energy of binding would therefore be negative and favorable only if the favorability of the enthalpy of binding overcomes the loss of entropy. N tells us the number of ligand binding sites a protein is predicted to have.

2.2.2 Methods: Isothermal Titration Calorimetry

Sugar binding was measured with isothermal titration calorimetry on a Microcal VP-ITC Isothermal Titration Calorimeter. 1.5 mL of 85 micromolar of LGP7 BP + GFP was loaded into the sample chamber, and ~300 µL of 88 micromolar asialofetuin was loaded into the syringe, and 2 µL was injected in for 4 seconds at 210 second intervals, with constant stirring, at 25ºC. A total of 29 injections were performed per experiment. The data was analyzed through Origin7.

2.3 Circular Dichroism

2.3.1 Circular Dichroism Theory

Circular dichroism probes the secondary structure of proteins. It uses a beam of circularly polarized light. This is created by adding together two beams of singularly polarized light that are out of phase with each other by $\frac{1}{4}$ wavelength. There are two types circularly polarized light: left and right. Different peptide bond angles will absorb these two types circularly polarized light differentially (Venyaminov, Sergie Yu, 1996). This will confer an ellipticity to the circularly
polarized light (Venyaminov, Sergie Yu, 1996). Measuring this ellipticity has allowed people to figure out that the common secondary structures of proteins create a characteristic peaks at certain wavelengths in when absorbing circularly polarized light, as pictured below in Figure 2.3.1.

![Figure 2.3.1 characteristic peaks of various secondary structures for circular dichroism (“4.2.1 Circular dichroism spectroscopy,” n.d.)](image)

Analysis of the ellipticity of these peaks can determine the relative amount of secondary structure in a protein. This would be predicted to decrease as the protein is denatured by rising temperature, therefore leading to a decrease in peak intensity. The proteins are predicted to go from being mainly beta sheet, to a random coil (denatured) as they melt.
A graph of the peak ellipticity values vs. the temperature can provide us with a melting curve of the protein’s secondary structure. This melting curve allows us to extrapolate the ellipticity values at which the protein is fully folded (we assume the value at 20ºC), and the value at which it is fully unfolded (we assume the value at 96ºC). The fraction folded and unfolded can then be calculated from the relative ellipticity values. These fractions can then be used to calculate the Keq at every temperature. A plot of the lnKeq vs the 1/T gives us the Van’t Hoff plot, which we can use to calculate various thermodynamic parameters (Venyaminov, Sergie Yu, 1996). The slope of this graph gives us \(-\Delta H/R\), from which we can calculate the enthalpy, by multiplying by \(-R\), while the intercept gives us \(\Delta S/R\), from which we can calculate entropy by multiplying by \(R\) (Venyaminov, Sergie Yu, 1996). With these parameters, we can calculate \(\Delta G\) at various temperatures by the equation \(\Delta G = \Delta H - T\Delta S\).

These thermodynamic parameters will elucidate the stability of the tertiary and secondary structures of these different forms of the proteins: a larger \(\Delta \Delta G\) corresponding to a larger free energy of unfolding, a larger \(\Delta H\) corresponding to a greater loss of energy due to more bond breaking, and a larger \(\Delta S\) corresponding to a greater gain of entropy (due to the structure becoming significantly more flexible and dynamic).

The \(T_m\) can be estimated from the x-intercept of the graph (when Keq = 1, or when lnKeq = 0) because the \(T_m\) is the point at which the fraction folded is equal to the fraction unfolded. A larger \(T_m\) signifies that a greater thermal energy is needed to
unfold the protein, and that it therefore has a more stable secondary and tertiary structure.

2.3.2 Methods: Circular Dichroism

Circular Dichroism was performed with 15 micromolar RBMC2 and LGP7 BP in 10 mM pH 7.4 phosphate buffer on a Jasco J-810 spectropolarimeter machine. Samples were equilibrated to the temperature, and then spectra were measured every two degrees from 20ºC to 96ºC. Using a single peak wavelength in the spectra (233 nm for LGP7 BP and 222 nm for RBMC2), a Van’t Hoff plot was created in the following manner: First, the fraction unfolded and folded was determined with the assumption that at 20ºC, the protein was fully folded, and at 96ºC the protein was fully unfolded. \( \frac{\Theta - \Theta_{folded}}{\Theta_{unfolded} - \Theta_{folded}} = \text{fraction unfolded, and } 1 - \text{fraction unfolded} = \text{Fraction folded.} \) From that, we can calculate \( K_{eq} \): 

\[
K_{eq} = \frac{[\text{unfolded}]}{[\text{folded}]}.
\]

From this, we can create the Van’t Hoff plot. The slope of the Van’t Hoff plots gives us: \(-\Delta H/R\) from which we can calculate the enthalpy, by multiplying by \(-R\) (-8.3145 J/molK) (Provost, n.d.). With these parameters, we can calculate \( \Delta G \) at various temperatures by the equation \( \Delta G = \Delta H - T \Delta S \). Next, we calculated the \( \Delta \Delta G \) by subtracting the \( \Delta G \) at 95ºC from the \( \Delta G \) at 10ºC. The \( T_m \) can be estimated from the x-intercept of the graph (when \( K_{eq} = 1 \), or when \( \ln(K_{eq}) = 0 \)) because the \( T_m \) is the point at which the fraction folded is equal to the fraction unfolded.

2.3.3 Methods: BeStSel (Beta Structure Selection)

This algorithm operates under the assumption that the CD spectra of a given protein is a linear combination of various base spectra (Micsonai et al., 2015). These
base spectra, shown in Figure 2.3.3, were optimized through a set of protein spectra with known secondary structures (Micsonai et al., 2015). Then, to parse out the secondary structures, a set of basis spectra were calculated from the reference database by linear least-square approximation (Micsonai et al., 2015). To determine the secondary structure of a novel spectrum, the optimized basis sets are linearly fit to the new spectra for each of the seven given secondary structure elements (Figure 2.3.3), and the secondary structure contents therefore parsed out (Micsonai et al., 2015). The same pre-calculated basis spectra are used for analysis of all new CD spectra (Micsonai et al., 2015).

**Figure 2.3.3**: Secondary structure determination by BeStSel

A) Secondary structures used in the algorithm, derived from secondary structure basis components by DSSP (Define Secondary Structure of Proteins) B) Basis spectra for the beta sheet secondary structures. Wavelength regions that were left out of the algorithm are shown by thin lines. C) Basis spectra for the non beta sheet secondary structures. Wavelength regions that were left out of the algorithm are shown by thin lines. D) A flowchart of the algorithm. (Micsonai et al., 2015)
2.4 X-ray crystallography

2.4.1 X-ray crystallography theory

X-ray crystallography is a process by which high resolution structures of biological molecules can be obtained. This often involves a rigorous protein purification process to obtain a homogeneous protein solution. Then, proteins are put through various screens to see what conditions are favorable for crystallization (Rhodes, 2014). Finally, diffraction data is obtained from the crystal from one of various sources of x-rays, leading to an electron density map (after solving the phase problem through various methods) (Rhodes, 2014). The protein sequence can then be built into the electron density map, and after rounds of refinement, lead to a final crystal structure (Rhodes, 2014).

Proteins can be purified from *E. coli* bacteria. Creation of crystals suitable for x-ray crystallography requires a very pure protein sample, to create a homogenous crystal lattice. Obtaining a crystal involves finding a properly concentrated solution, with the correct buffer molecules to keep the protein from aggregating and precipitating in a non crystalline form. In this thesis, we used the hanging drop method to conduct crystallization trial. This method involves placing a drop of protein in solution mixed with equal parts buffer on a coverslip. This is place on top of a reservoir solution, containing the same buffer, but not diluted with the protein solution. The lower concentration of the protein drop causes the water to preferentially evaporate and then recondense into the buffer solution below, slowly diluting that solution while concentrating the protein solution, until the two are at
roughly equal concentrations. This slow increase in concentration promotes the creation of crystals. Various conditions, like pH, ionic strength, and various organic molecules and cofactors, can promote the crystallization of proteins: screens are set up to find what exact conditions are suitable for the particular protein, and these screens are continually refined. Ions in particular can cause the proteins to “salt out” and form crystallized. A proper crystal is not overnucleated, and therefore only contains one crystal lattice and not two or more overlapping ones, or a poorly formed crystal lattice. Forming pure and stable protein crystals is usually an iterative process with much refinement involved. Protein crystals are saturated with solvent, and are not stable outside the solvent (Rhodes, 2014).

Once a proper crystal is obtained, diffraction data can be gathered. X-rays need to be used to create a high resolution picture, due to their near atomic wavelength of roughly $10^{-10}$ m (~1.5 Angstroms) (Rhodes, 2014). A narrow beam of X-ray is produced, usually from a rotating anode or synchrotron (significantly stronger) X-ray source. This beam is directed at the crystal, which is usually cryoprotected from X-ray damage by being frozen in liquid nitrogen. Most of the beam passes through the sample, which is stopped from being collected as data by a beam stop (Rhodes, 2014). The rest of the diffracted X-rays are collected on X-ray film, or another form of digital detection. Each spot gives us 2 pieces of information: intensity, and position (in h,k,l). From this, we can determine the unit cell of the crystal, due to the fact that the diffraction data essentially gives us a picture of the reciprocal lattice (the reciprocal of the unit cell) (Rhodes, 2014). From this, you want to find the asymmetric unit, or the smallest unit of the crystal that can reproduce the
whole crystal through rotation and translation (Rhodes, 2014). Usually, the sample is rotated to obtain multiple perspectives of the crystal, and therefore a 3-D structure.

From this data, we need 3 pieces of information to calculate electron density from the x-ray diffraction. The first two, the wavelength and the amplitude of the waves, are known. However, the third piece of information, phase has to be determined experimentally. Once all 3 are calculated, one can use a Fourier transformation to get an electron density map from diffraction data (Rhodes, 2014).

Once the preliminary electron density map has been determined, various rounds of refinement need to occur before the final structure is obtained. $R$ values and $R_{	ext{free}}$ describe the accuracy of the protein structure. The R factor measures how well the predicted diffraction pattern (based on the model you build) matches the experimentally determined diffraction pattern. For structural determination and refinement, only around 95% of the data are used while ~5% are removed. The $R_{	ext{free}}$ value measures how well our model predicts the remaining ~5% of data that was
removed from the data set. A completely random group of atoms would give an R value of around 0.6. R values below 0.2 are regarded as good, whereas R values below .1 are likely on obtained in small-molecule crystallography. An R_free value that is more than 7% above the R value suggests that you are overinterpreting the data and creating an erroneous model (Rhodes, 2014). If a model of a protein already exists which is someone somewhat homologous to the protein of interest, molecular replacement can be used to solve the phase problem, and once the electron densities are obtained, a difference map can be created, where the known map is subtracted from the calculated data (Rhodes, 2014). This can also highlight incongruencies between the model and the new crystal structure. In refinement, one must often manually build in predicted residues, based on favorable peptide bond angles, electron density, and the known sequence of the protein. Once this occurs, refinement can happen, which takes the calculated data and the newly predicted data and constructs a new model. This usually leads to reduction in the R-values, demonstrating it is a more accurate model. If the R-value goes up, it suggests your model is incorrect.

Ramachandran plots can also be used to test the validity of a model by showing if the backbone angle are mainly in the favorable, or if they are at least allowed (Wlodawer, Minor, Dauter, & Jaskolski, 2008).

2.4.2 Methods: Crystallization (X-ray crystallography)

A hanging drop crystal screens were set up with a JSCG+ screen, from Molecular Dimensions, in a 24 well format. See results for specific conditions. Half a milliliter of the condition was added to the well. 0.7 microliters of 11.5 mg/mL LGP7
BP in pH 7.5 TBS was mixed with 0.7 microliters of TBS for the drop. This drop was placed over the well and sealed to be airtight, allowing slow evaporation of the drop.

2.5 Bioinformatics and Modelling

2.5.1 Methods: Psi-BLAST Sequence Analysis

BLAST is a simple alignment search engine. Biologically relevant proteins that have lower sequence identity with the protein of interest sometimes do not show up with a simple alignment search. PSI-BLAST uses a new algorithm to find more distantly related but biologically relevant sequences (Altschul et al. 1997).

To understand how PSI BLAST works, it’s important to understand how BLAST works. BLAST compares two sequences directly with each other, with a score $s_{ij}$ for aligned amino acids $i$ and $j$. In other words, it looks for equal length segments that have similar amino acids. High quality local alignments are called “high scoring segment pairs” or HSPs (Altschul et al., 1997). BLAST searches the database for words, or the segments described above, and any word alignment is a hit (Altschul et al., 1997). The PSI BLAST algorithm takes into account how far two words are from each other, or the window length $A$, and therefore requires two non-overlapping hits instead of one. Consequently, the threshold parameter for what qualifies as a hit must be lowered for comparable accuracy. This if similarities within a window are found, “gapped alignment” proceeds (Altschul et al., 1997).

Making this alignment position-specific greatly increases the chance of finding more distantly related proteins, as does the iteration (Altschul et al., 1997). Each iteration, a choice must be made about which proteins to base the next iteration
on, therefore refining the search with more sequences to align to. This is how PSI-BLAST, or Position Specific Iterated BLAST, functions. This can identify large proteins with domains that contain a similar protein to the protein being used to search, as happened here.

2.5.2 Methods: Modelling—Raptor X Structure Prediction

Raptor X is a mostly template-based modelling program, meaning it uses structural templates (from the PDB) that have similar protein sequences to the protein sequence of interest to model the protein, but also has a number of strategies for protein prediction when no clear homologs are present (Källberg, Margaryan, Wang, Ma, & Xu, 2014). Raptor X uses conditional neural fields (CNF) to combine various protein biological motifs. To determine these, it uses a nonlinear probabilistic scoring function (Källberg et al., 2014). Secondly, RaptorX uses a multiple-template threading (MTT) procedure, which enables the use of multiple templates to model one protein sequence (Källberg et al., 2014).

2.5.3 Methods: Modelling—Phyre2 Structure Prediction

Phyre2, in normal mode instead of intensive mode, is a template-based protein structure prediction server (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015). The rough pipeline it uses to determine the structure of novel protein sequences is as follows: first, it uses HHblits to create a collection of diverse but biologically relevant homologs to the given protein sequence (Kelley et al., 2015). HHblits uses profile-profile matching, and has a 50-100% increase in sensitivity over PSI-BLAST (Kelley et al., 2015). The secondary structure is simultaneously predicted with PSI-PRED. PSI-PRED uses neural networks trained on known protein sequence profiles to
predict secondary structure with an accuracy of 75-80% (Jones, 1999). All this information is converted to a hidden Markov model (HMM). This HMM is scanned against the HMMs of proteins with known sequences, which is known as the fold library (Kelley et al., 2015). To do this, Phyre2 uses the algorithm HHsearch (Kelley et al., 2015). This creates a library of templates to which the novel protein can be aligned to, creating a model for its backbone (Kelley et al., 2015). Next, gaps that are not accounted for by these templates are modeled in through a sort of “loop library”, in which templates are searched for with similar endpoint distances and loop distances (Kelley et al., 2015). Sometimes, the server is unable to model anything in for these loops, which suggests an error in the original alignment (Kelley et al., 2015). Then, the side chains are modelled into the backbone in their most probable rotamer while avoiding steric clashes (Kelley et al., 2015).

2.5.4 Methods: Modelling—SWISS-MODEL Structure Prediction

SWISS-MODEL uses HHblits and BLAST in parallel to search its HMM library of protein structure templates, from the PDB (Biasini et al., 2014). Sequence profiles, secondary structure predicted by PSIPRED, DSSP and SSpro, solvent accessibility predicted by ACCpro, and per-residue solvent accessibility(NACCESS) are all stored together with the structure in the template library (Biasini et al., 2014). Additionally, ligands of these templates are identified and separated from buffer molecules and solvents (Biasini et al., 2014). The alignment properties are then modelled as a probability density function (PDF), which takes into account that some properties are more important than others at predicting the quality of the template (Biasini et al., 2014). The template most likely to have structural similarity has a
value at which the joint distribution is maximized of the combined PDF estimates of each property. This is the global quality estimation score (GMQE) (Biasini et al., 2014).

2.5.4 Methods: SignalP 4.0

SignalP 4.0 is a neural network-based method used to detect secretion signal peptides, and to distinguish them from N-terminal transmembrane regions (as both will be hydrophobic) (Petersen, Brunak, Heijne, & Nielsen, 2011). Earlier versions could not reliably distinguish between these two. This network was trained on known secretion signals of eukaryotes, and Gram-positive and -negative bacteria (Petersen et al., 2011). Transmembrane sequences were used to train SignalP-TM to predict transmembrane sequences (Petersen et al., 2011). If four or more positions are predicted as being transmembrane, SignalP-TM is used for the prediction of the protein (Petersen et al., 2011). Otherwise, SignalP is used.
Chapter 3 - Results

3.1 Approach to protein expression

Expression of the full length LGP7 BPC resulted only in insoluble protein material, as demonstrated by Western Blot analysis of the cell lysate. We therefore used RaptorX (Källberg et al., 2014) structure and binding prediction, as well as PSIPRED secondary structure and domain prediction (Jones, 1999), to divide the protein up into seven plausible domains. The first of these domains is the beta prism lectin domain excluding the secretion signal, residues 46-204.

Figure 3.1.1: PSIPRED secondary structure, binding, and domain boundary prediction of LGP7 BPC with secretion signal (Jones, 1999)
Figure 3.1.2: Psi-Blast boundary prediction based on known matches, from PSIPRED (Jones, 1999)

The domains were therefore created as follows, with care being taken to not interrupt any of the predicted secondary structures, and with the putative domains in mind but often broken down into smaller fragments. The fragments are shown in Figure 3.1.3 with on a Raptor X ligand binding prediction model of the full LGP7 BPC protein (Källberg et al., 2014). These domains were also roughly predicted by RaptorX structure prediction (Källberg et al., 2014).

Fragment 1 (beta prism, in red): residue 46-204
Fragment 2 (in yellow): residue 189 - 393
Fragment 3 (in green): residue 372 - 506
Fragment 4a (in blue): residue 478 - 626
Fragment 4b (purple): residue 594-930
Fragment 4c (in pink): residue 909 - 1184
Fragment 5 (in grey): residue 1131 - 1437
Much of the middle of the protein appears unstructured simply due to it being unable to be predicted by traditional homology models, as Raptor X bases its modelling on templates of previously solved structures. The putative beta prism lectin domain, LGP7 BP, is the only domain predicted with a high degree of certainty (shown in red, Figure 3.1.3).

3.2 Cloning and Expression

3.2.1 LGP7 BP Cloning

This first domain, LGP7 BP, was successfully cloned out of the Vibrio crassostrea LGP7 genome and cloned into the plasmid pNGFP-BC. It is apparent as the band at ~0.45 kilobases on the gel, (it is predicted to be 0.477 kilobases), lanes 2 and 3. The plasmid is 6.2 kb and also apparent on the gel, lanes 2 and 3.
3.2.2 LGP7 BP + NGFP Expression

This plasmid attaches a cleavage GFP+His-tag to the protein of interest. The attached GFP often aids in solubility and visibility of proteins during purification. The pCGFP-BC attaches the GFP to the C-terminal, as opposed to the N-terminal, but resulted in more insoluble protein material here. As is evident in Figure 3.2.2, LGP7 + NGFP + His-tag construct resulted in significant soluble protein expression following Ni/NTA affinity chromatography purification of the soluble cell lysate on a Biorad Profina Protein Purification system. The full-length protein is apparent in the 2nd lane, at 45 kDa (LGP7 is 18 kDa, GFP<sub>UV</sub> is 27 kDa), with a few impurities.
Figure 3.2.2: 12.5 % SDS-PAGE gels stained with G-250 coomassie brilliant blue of Nickel column fractions. Washes were with 40 mM imidazole. Eluate was in 250 mM imidazole. Eluate contained \( \sim 30 \) mgs protein. Band at \(~45\) kDa corresponds to LGP7 BP + GFP (18 kDa + 27 kDa)

### 3.2.3 LGP7 BP Purification

The nickel column eluate was concentrated to around \(~10\) mg/mL of protein. This protein solution was then digested for 1 hour with 1:250 w/w protein/thrombin, with 10 mM \( \text{CaCl}_2 \), as this had proved optimal for complete GFP cleavage through the thrombin site without being so concentrated that it participated in nonspecific cleavage that cleaved the protein itself. Comparison of lane 2 of Figure 3.2.2 and lane 2 of Figure 3.2.3 A shows the protein pre and post digestion, and shows complete cleavage, due to the upper band at 45 kDa disappearing almost completely post digestion. The digestion was stopped with 20 mM EDTA, and the sample was then centrifuged to remove any precipitation or aggregates, and then run over a Superose 6 10/300 GL sizing exclusion column in pH 7.5 TBS.
Figure 3.2.3: A) 12.5 % SDS-PAGE gels stained with G-250 Coomassie brilliant blue. Each lane (besides ladder and post digestion) contains a sizing column fraction, as shown on the graph in B. The major bands shown at ~27 and ~18 kDa correspond to GFP and pure LGP7 BP respectively. As shown in this gel, after digestion, LGP7 BP presents a heterogeneous population, perhaps due to a loop being nicked or a disordered section being cut off.

B) Superose 6 10/300 GL column chromatogram trace. Measured by UV absorbance and fluorescence. LGP7 BP shows a monodisperse peak, as confirmed by the gel in A.
Around 4 mgs of pure protein were recovered from the four liters of bacteria. The thrombin digestion appeared to create two populations of pure protein, as shown in Figure 3.2.3 A, fractions B5-B10 or at 26 mL, slightly more than 1 column volume (24 mLs), perhaps due to a disordered loop being cut off or nicked by the enzyme. Despite the heterogeneous population, both appeared to have the same slight affinity for the sugars in the sizing column (agarose), as they came out later than even the salts in the column (which come out at around fraction B3), and it did not impede crystallization. This same affinity is seen in RBMC2, which comes out in fractions ~B4-B9 or one column volume. Buffer molecules are very small and therefore move through the column very slowly—larger proteins like GFP should move through the column relatively quickly. GFP does do this, as it comes out at 16 mL or two thirds of a column volume. If LGP7 BP did not have an affinity for the sugar to the column, it would be impossible to separate this protein, at 18kDa, from GFP, at 25 kDa, through size exclusion chromatography alone.

3.3 LGP7 BP Crystallization

The first small crystals were found through a molecular dimensions JSCG+ screen, in the condition 2-23: 1.0 M succinic acid, 0.1 M HEPES and 1% 2K MME PEG, pH 7. 0.7 µL of 11.5 mg/mL protein in TBS pH 7.5 was mixed with 0.7 µL of each of the conditions below, and placed on a coverslip over a well 0.5 mL of each of the conditions below, with an airtight seal.
A few other conditions showed promising crystalline like precipitation, but there were no outstanding patterns from these conditions. (Table 2) The rest of the conditions produced precipitation or stayed clear.

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1-7</td>
<td>None</td>
<td>0.1 M</td>
<td>CHES</td>
<td>9.5</td>
<td>20</td>
<td>PEG 8000</td>
</tr>
<tr>
<td>1-8</td>
<td>0.2 M</td>
<td>Ammonium formate</td>
<td>None</td>
<td>-</td>
<td>20</td>
<td>PEG 3350</td>
</tr>
<tr>
<td>1-33</td>
<td>None</td>
<td>0.1 M</td>
<td>Sodium/potassium phosphate</td>
<td>6.2</td>
<td>25</td>
<td>1,2-Propanediol</td>
</tr>
<tr>
<td>1-37</td>
<td>None</td>
<td>None</td>
<td>-</td>
<td>24</td>
<td>PEG 1500</td>
<td></td>
</tr>
<tr>
<td>2-7</td>
<td>0.2 M</td>
<td>Zinc acetate dihydrate</td>
<td>0.1 M</td>
<td>Sodium cacodylate</td>
<td>6.5</td>
<td>10</td>
</tr>
<tr>
<td>2-17</td>
<td>0.2 M</td>
<td>Magnesium chloride hexahydrate</td>
<td>0.1 M</td>
<td>Tris</td>
<td>8.5</td>
<td>50</td>
</tr>
<tr>
<td>2-22</td>
<td>1.1 M</td>
<td>Sodium malonate dibasic monohydrate</td>
<td>0.1 M</td>
<td>HEPES</td>
<td>7</td>
<td>0.5</td>
</tr>
<tr>
<td>2-31</td>
<td>0.1 M</td>
<td>Succinic acid</td>
<td>None</td>
<td>-</td>
<td>15</td>
<td>PEG 3350</td>
</tr>
<tr>
<td>2-32</td>
<td>0.15 M</td>
<td>DL-Malic acid</td>
<td>None</td>
<td>-</td>
<td>20</td>
<td>PEG 3350</td>
</tr>
</tbody>
</table>

Table 1: JSCG + screen results of promising conditions of LPG7 BP (those with crystalline like precipitation).

Figure 3.3: small hexagonal proteins in a drop with 1.0 M succinic acid, .1 M HEPES, and 1% PEG 2K MME, condition 2-3 of the Molecular Dimensions JCSG+ screen condition 2-23. A drop of 0.7 µL 11.5 mg/mL was mixed with 0.7 µL of this condition, leading to these crystals.
The screen in Table 2 was created with the idea to screen around the pHs, and concentrations of the various reagents, in order to optimize crystals, of the condition 2-23. However, repetition of this crystallization condition, but with sodium succinate dibasic hexahydrate instead of succinic acid, did not yield any crystals, and neither did any of the other screened conditions.

<table>
<thead>
<tr>
<th></th>
<th>Conc. Sodium succinate dibasic hexahydrate (M)</th>
<th>Conc. HEPES (M)</th>
<th>% PEG 2000 MME</th>
<th>pH</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>1</td>
<td>0.1</td>
<td>1%</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>A2</td>
<td>0.8</td>
<td>0.08</td>
<td>1%</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>A3</td>
<td>0.6</td>
<td>0.06</td>
<td>1%</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>A4</td>
<td>1</td>
<td>0.1</td>
<td>0%</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>A5</td>
<td>0.8</td>
<td>0.08</td>
<td>0%</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>A6</td>
<td>0.6</td>
<td>0.06</td>
<td>0%</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>B1</td>
<td>1</td>
<td>0.1</td>
<td>1%</td>
<td>6.5</td>
<td>1</td>
</tr>
<tr>
<td>B2</td>
<td>0.8</td>
<td>0.08</td>
<td>1%</td>
<td>6.5</td>
<td>1</td>
</tr>
<tr>
<td>B3</td>
<td>0.6</td>
<td>0.06</td>
<td>1%</td>
<td>6.5</td>
<td>1</td>
</tr>
<tr>
<td>B4</td>
<td>1</td>
<td>0.1</td>
<td>0%</td>
<td>6.5</td>
<td>1</td>
</tr>
<tr>
<td>B5</td>
<td>0.8</td>
<td>0.08</td>
<td>0%</td>
<td>6.5</td>
<td>1</td>
</tr>
<tr>
<td>B6</td>
<td>0.6</td>
<td>0.06</td>
<td>0%</td>
<td>6.5</td>
<td>1</td>
</tr>
<tr>
<td>C1</td>
<td>1</td>
<td>0.1</td>
<td>1%</td>
<td>7.5</td>
<td>1</td>
</tr>
<tr>
<td>C2</td>
<td>0.8</td>
<td>0.08</td>
<td>1%</td>
<td>7.5</td>
<td>0</td>
</tr>
<tr>
<td>C3</td>
<td>0.6</td>
<td>0.06</td>
<td>1%</td>
<td>7.5</td>
<td>0</td>
</tr>
<tr>
<td>C4</td>
<td>1</td>
<td>0.1</td>
<td>0%</td>
<td>7.5</td>
<td>0</td>
</tr>
<tr>
<td>C5</td>
<td>0.8</td>
<td>0.08</td>
<td>0%</td>
<td>7.5</td>
<td>1</td>
</tr>
<tr>
<td>C6</td>
<td>0.6</td>
<td>0.06</td>
<td>0%</td>
<td>7.5</td>
<td>0</td>
</tr>
<tr>
<td>D1</td>
<td>1</td>
<td>0.1</td>
<td>5%</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>D2</td>
<td>0.8</td>
<td>0.08</td>
<td>5%</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>D3</td>
<td>0.6</td>
<td>0.06</td>
<td>5%</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>D4</td>
<td>1</td>
<td>0.1</td>
<td>5%</td>
<td>6.5</td>
<td>1</td>
</tr>
<tr>
<td>D5</td>
<td>0.8</td>
<td>0.08</td>
<td>5%</td>
<td>6.5</td>
<td>1</td>
</tr>
<tr>
<td>D6</td>
<td>0.6</td>
<td>0.06</td>
<td>5%</td>
<td>6.5</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2: Crystal screen around the condition 1M Succinic acid, 1 M HEPES, and 1% PEG 2K MME. 1 = Precipitate 0 = clear
3.4 Isothermal Titration Calorimetry of LGP7 BP and Asialofetuin

3.4.1 ITC Preparation and Experiment

LGP7 BP + GFP was concentrated to 85 µM (with some free GFP and free LGP7 BP contamination) in TBS and dialyzed overnight in pH 7.5 TBS in a membrane with a 3-kDa cutoff. A 88 micromolar asialofetuin solution was made from the dialysate. This was then loaded into the syringe, and 2 µL was injected in for 4 seconds at 210 second intervals in the sample chamber with 1.5 mL of the LGP7 + GFP, with constant stirring, at 25ºC. A total of 29 injections were performed per experiment.

Figure 3.4.1.1: LGP7 BP + NGFP off the sizing column, before ITC. Fractions in the red box were combined for the experiment
This raw ITC data shows evidence that asialofetuin does not bind LGP7 BP. Though there is a release of heat upon titration of the asialofetuin, this could simply be due to the dilution of the sugar. The greatly shifting baseline and the fact that this release in heat is not changing suggests it probably is not binding.

3.4.2 Isothermal Titration Calorimetry Suggests that LGP7 BP does not bind to asialofetuin

In comparable ITC experiments, both RBMC2 and VCC show a large release of heat upon titration of the asialofetuin, which decreases as the binding sites are filled up (molar ratio .4 for RBMC2, and ~.2 VCC). The concentrations of RBMC2 (48 µM), VCC (23 µM) and asialofetuin (~115 µM) in the experiments was very close to the concentration of sugar and protein used here, so it should not have been a limiting factor.
Origin 7 analysis of the ITC of LGP7 BP + GFP and asialofetuin gives messy data with a large error, suggesting that the $\mu$cal/sec are merely due to the increase in entropy when the sugar is diluted creating heat, or due to machine error, and that it does not bind LGP7 BP. This lack of binding could however be explained by three other factors. Firstly, the protein attached to GFP comes off the column significantly earlier than the pure protein (fractions A14-B2), suggesting that GFP blocks sugar binding to some extent. In addition to this, the contamination bands on the gel are significant, and would make the concentration of protein appear greater than it is. The second factor is that asialofetuin solutions have been shown to be more dilute than expected by nanodrop spectrometers, potentially due to the static-y, flyaway nature of the sugar. The third factor is that expression in *E. coli* has lead to LGP7 BP being unstable not well-folded, and therefore unable to participate in native interactions.
Figure 3.4.2.2 Origin 7 analysis of the ITC of LGP7 BP + GFP and asialofetuin
3.5 Circular Dichroism of RBMC2 and LGP7 BP

To indicate whether this protein is well folded, we performed thermal melts monitored by circular dichroism. This protein was concentrated to 15 µM (from fractions B6-B10 off the sizing column, shown in Figure 3.2.3 A), and buffer exchanged into pH 7.4 sodium phosphate buffer. 0.6 mL of this solution was analyzed through circular dichroism on a Jasco J-810 spectropolarimeter machine over a range of 20 ºC to 96 ºC. Characteristic beta sheet (minimum at ~215 nm) and random coil ellipticities (minimum at ~200 nm) were observed. RMBC2 does not go through the characteristic breakdown into random coil, as LGP7 BP does, and stays rather beta sheet in nature.

Figure 3.5.1 Changing of ellipticity over 20-96ºC for LGP7 BP
3.5.1 Calculation of thermodynamic parameters of RBMC2 and LGP7 BP

Analysis of the CD led to determination of certain thermodynamic parameters of LGP7 BP, as compared with that of RBMC2 as a control. Analysis of the changing of the peak at 233 nm for LGP7 BP, and at 222 nm for RBMC2, lead to a determination of the following thermodynamic parameters, for the range 20ºC to 96ºC, after the creation of a Van’t Hoff plot. A Van’t Hoff plot was created in the following manner: first, the fraction unfolded and folded was determined with the assumption that at 20ºC, the protein was fully folded, and at 96ºC the protein was fully unfolded. 

\[
\frac{(\Theta - \Theta_{\text{folded}})}{(\Theta_{\text{unfolded}} - \Theta_{\text{folded}})} = \text{fraction unfolded, and } 1 - \text{fraction}
\]
unfolded = Fraction folded. From that, we can calculate Keq: 

$$\text{Keq} = \frac{[\text{unfolded}]}{[\text{folded}]}.$$  

From this, we can create the Van’t Hoff plot, shown below.

**A)**

![LGP7 BP CD Melt at 233 nm](image1)

**B)**

![LGP7 BP Van't Hoff](image2)

Figure 3.5.1.1 A) Ellipticity of LGP7 BP at the peak at 233 nm from 20-96°C. B) Van’t hoff plot of the melting of this protein

**A)**

![Rbmc2 CD Melt at 222 nm](image3)

**B)**

![Rbmc2 Van't Hoff](image4)

Figure 3.5.1.2 A) Ellipticity of RBMC2 at the peak at 222 nm from 20-96°C. B) Van’t hoff plot of the melting of this protein

The slope of the Van’t Hoff plots gives us: $-\Delta H/R$ from which we can calculate the enthalpy, by multiplying by $-R (-8.3145 \text{ J/molK})$ (Provost, n.d.). With these parameters, we can calculate $\Delta G$ at various temperatures by the equation $\Delta G = \Delta H - T\Delta S$. Next, we calculated the $\Delta G$ by subtracting the $\Delta G$ at 96°C from the $\Delta G$ at 20°C. The $T_m$ can be estimated from the x-intercept of the graph (when $\text{Keq} = 1$, or
when lnKeq= 0) because the T\textsubscript{m} is the point at which the fraction folded is equal to the fraction unfolded.

<table>
<thead>
<tr>
<th></th>
<th>$\Delta H$ (kJ/mol)</th>
<th>$\Delta S$ (J/mol)</th>
<th>Tm (°C)</th>
<th>$\Delta G$ 20°C (kJ/mol)</th>
<th>$\Delta G$ 96°C (kJ/mol)</th>
<th>$\Delta \Delta G$ (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rbmc2</td>
<td>110</td>
<td>337</td>
<td>53.5</td>
<td>11.3</td>
<td>-14.3</td>
<td>25.6</td>
</tr>
<tr>
<td>LGP7 BP</td>
<td>103</td>
<td>319</td>
<td>51.8</td>
<td>10.2</td>
<td>-14.1</td>
<td>24.3</td>
</tr>
</tbody>
</table>

Table 3: Thermodynamic parameters determined by CD melting of beta prisms RBMC2 and LGP7 BP

The RBMC2 T\textsubscript{m} is almost identical to the one previously determined by the Olson lab (55°C). The melting point of LGP7 BP is only slightly lower, at about 52°C. The changes in enthalpy and entropy ($\Delta H$ and $\Delta S$) are marginally smaller for LGP7 BP, suggesting that it has marginally less strong bonds and marginally more disorder than RBMC2, though this discrepancy may be due to machine error. The $\Delta \Delta G$s over the melts are only marginally different, suggesting that LGP7 BP is well folded and maybe marginally more unstable.

3.5.2 Secondary structure analysis by BeStSel shows highly similar secondary structures of LGP7 BP and RBMC2, different melting

We used the program BeStSel (for beta sheet selection) to more precisely determine the changes in secondary structure of LGP7 BP and Rbmc2. Previously, CD was not considered a very reliable way to determine the secondary structure of beta sheet proteins and mixed alpha beta proteins, due to the spectral diversity of beta sheet proteins. This study posits that this is due to different types of beta sheets having different characteristic spectra. Their algorithm therefore parses out parallel beta sheets, antiparallel relaxed beta sheets, antiparallel left twisted beta sheets, and
antiparallel right twisted beta sheets, as well as alpha helices and distorted helices, and “turn” and “other” secondary structure. On a test set of beta rich proteins whose secondary structure was known, BeStSel predicted the secondary structure with RMSDs all less than .04 (Micsonai et al., 2015). We therefore considered this program a reliable to characterize and compare general changes in secondary structures of LGP7 BP and RMBC2. Some sample analysis of spectra being fit to a basis spectra at certain temperatures are shown below in Figure 3.5.2.1-3

These spectra show the folded proteins’ secondary structure content. It is apparent that this is highly similar for both RBMC2 and LGP7 BP, even though the CD signals were quite different. Both proteins are majority antiparallel beta sheet
(LGP7 BP - 49.2% vs. RBMC2 56.0%), characteristic of beta prism lectin domains, and of this mostly right twisted beta sheet. They are predicted to have no parallel beta sheet character, are predicted to have slight distorted helical content (LGP7 BP – 3.9% vs. RBMC2 5.2%), though we know this is untrue for RMBC2 due to its crystal structure, and can be attributed to noise or error in BeStSel. It is therefore likely that LGP7 BP also has no helical content. The rest of the secondary structure is predicted to be “turn” or “other”, which is accordance with the RMBC2 crystal structure.

Figure 3.5.2.2 Secondary structure of LGP7 BP and RBMC2 at 50ºC as determined by BeStSel. Concentration of protein (15 µM), sequence length of the protein (LGP7 = 159 amino acids and RMBC2 = 138 amino acids), and pathlength (0.1 cm) were input (Micsonai et al., 2015)

These proteins have very similar secondary structures at 50ºC to the proteins at 20ºC, demonstrating that at this temperature the proteins’ secondary structure has not really started to melt.
The discrepancies in the melting of these two proteins is apparent at 96°C. While LGP7 BP has a large “other” character (50.7 %) signifying it is most likely mostly random coil, RBMC2 has very similar secondary structure to its 50°C and 20°C states, except that its right twisted beta sheet character has increased. This suggests the formation of amyloids. It does indeed melt, as the peak shifting demonstrates, but it melts into a different structure, instead of melting into random coil. These graphs show how the BeStSel algorithm fit a curve to the data, and the detailed breakdown of secondary structures.

Figure 3.5.2.3 Secondary structure of LGP7 BP and RBMC2 at 96°C as determined by BeStSel. Concentration of protein (15 µM), sequence length of the protein (LGP7 = 159 amino acids and RMBC2 = 138 amino acids), and pathlength (0.1 cm) were input (Micsonai et al., 2015)
<table>
<thead>
<tr>
<th>Temperature</th>
<th>Helix1</th>
<th>Helix2</th>
<th>Anti1</th>
<th>Anti2</th>
<th>Anti3</th>
<th>Para</th>
<th>Turn</th>
<th>Others</th>
<th>NRMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10.58</td>
<td>5.21</td>
<td>4.29</td>
<td>19.6</td>
<td>22.28</td>
<td>0</td>
<td>0.53</td>
<td>37.52</td>
<td>0.0835</td>
</tr>
<tr>
<td>22</td>
<td>0.69</td>
<td>4.41</td>
<td>4.14</td>
<td>19.67</td>
<td>38.57</td>
<td>0</td>
<td>2.56</td>
<td>29.97</td>
<td>0.0741</td>
</tr>
<tr>
<td>24</td>
<td>12.52</td>
<td>4.11</td>
<td>5</td>
<td>22.58</td>
<td>20.54</td>
<td>0</td>
<td>0</td>
<td>35.26</td>
<td>0.0694</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>4.71</td>
<td>3.89</td>
<td>17.07</td>
<td>37.97</td>
<td>0</td>
<td>3.77</td>
<td>32.59</td>
<td>0.0718</td>
</tr>
<tr>
<td>28</td>
<td>9.38</td>
<td>4.86</td>
<td>4.93</td>
<td>19.46</td>
<td>25.21</td>
<td>0</td>
<td>1.63</td>
<td>34.54</td>
<td>0.0783</td>
</tr>
<tr>
<td>30</td>
<td>5.89</td>
<td>5.16</td>
<td>4.51</td>
<td>17.43</td>
<td>31.73</td>
<td>0</td>
<td>2.47</td>
<td>32.81</td>
<td>0.0747</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>5</td>
<td>3.52</td>
<td>14.05</td>
<td>39.52</td>
<td>0</td>
<td>0.53</td>
<td>37.37</td>
<td>0.0724</td>
</tr>
<tr>
<td>34</td>
<td>9.48</td>
<td>4.31</td>
<td>4.58</td>
<td>18.6</td>
<td>22.66</td>
<td>0</td>
<td>3.07</td>
<td>37.31</td>
<td>0.0777</td>
</tr>
<tr>
<td>36</td>
<td>0</td>
<td>4.67</td>
<td>3.72</td>
<td>14.3</td>
<td>39.38</td>
<td>0</td>
<td>0</td>
<td>37.94</td>
<td>0.0787</td>
</tr>
<tr>
<td>38</td>
<td>11.45</td>
<td>4.92</td>
<td>4.52</td>
<td>19.16</td>
<td>16.53</td>
<td>0</td>
<td>1.7</td>
<td>41.72</td>
<td>0.0822</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
<td>5.81</td>
<td>2.61</td>
<td>10.48</td>
<td>40.19</td>
<td>0</td>
<td>4.6</td>
<td>36.3</td>
<td>0.0653</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
<td>5.35</td>
<td>3.21</td>
<td>13.66</td>
<td>45.43</td>
<td>0</td>
<td>2.35</td>
<td>29.99</td>
<td>0.0578</td>
</tr>
<tr>
<td>44</td>
<td>8.79</td>
<td>6.14</td>
<td>5.43</td>
<td>16.45</td>
<td>23.45</td>
<td>0</td>
<td>4.65</td>
<td>35.1</td>
<td>0.0708</td>
</tr>
<tr>
<td>46</td>
<td>5.84</td>
<td>5.19</td>
<td>3.87</td>
<td>11.16</td>
<td>27.09</td>
<td>0</td>
<td>5.36</td>
<td>41.48</td>
<td>0.0727</td>
</tr>
<tr>
<td>48</td>
<td>9.28</td>
<td>5.92</td>
<td>5.97</td>
<td>14.87</td>
<td>18.76</td>
<td>0</td>
<td>6.95</td>
<td>42.27</td>
<td>0.0721</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>5.34</td>
<td>0</td>
<td>0</td>
<td>52.03</td>
<td>0</td>
<td>9.21</td>
<td>33.43</td>
<td>0.0421</td>
</tr>
<tr>
<td>52</td>
<td>0</td>
<td>9.17</td>
<td>0</td>
<td>0</td>
<td>26.24</td>
<td>0</td>
<td>16.33</td>
<td>48.26</td>
<td>0.0494</td>
</tr>
<tr>
<td>54</td>
<td>0</td>
<td>7.9</td>
<td>0</td>
<td>0</td>
<td>26.59</td>
<td>0</td>
<td>13.3</td>
<td>52.22</td>
<td>0.0357</td>
</tr>
<tr>
<td>56</td>
<td>10.69</td>
<td>9.85</td>
<td>1.34</td>
<td>0</td>
<td>9.55</td>
<td>0</td>
<td>11.77</td>
<td>56.8</td>
<td>0.0516</td>
</tr>
<tr>
<td>58</td>
<td>12.98</td>
<td>9.7</td>
<td>0</td>
<td>0</td>
<td>2.17</td>
<td>0</td>
<td>12.88</td>
<td>62.28</td>
<td>0.0506</td>
</tr>
<tr>
<td>60</td>
<td>9.68</td>
<td>11.07</td>
<td>0</td>
<td>0</td>
<td>13.26</td>
<td>0</td>
<td>13.66</td>
<td>52.33</td>
<td>0.0330</td>
</tr>
<tr>
<td>62</td>
<td>0.39</td>
<td>8.68</td>
<td>0</td>
<td>0</td>
<td>17.47</td>
<td>0</td>
<td>14.74</td>
<td>58.72</td>
<td>0.0237</td>
</tr>
<tr>
<td>64</td>
<td>2.29</td>
<td>6.11</td>
<td>0</td>
<td>0</td>
<td>15.33</td>
<td>0</td>
<td>13.41</td>
<td>62.88</td>
<td>0.0292</td>
</tr>
<tr>
<td>66</td>
<td>11.59</td>
<td>11.12</td>
<td>0</td>
<td>0</td>
<td>13.37</td>
<td>0</td>
<td>18.89</td>
<td>45.02</td>
<td>0.0407</td>
</tr>
<tr>
<td>68</td>
<td>14.11</td>
<td>7.8</td>
<td>0</td>
<td>0</td>
<td>19.37</td>
<td>0</td>
<td>16.7</td>
<td>42.01</td>
<td>0.0450</td>
</tr>
<tr>
<td>70</td>
<td>8.31</td>
<td>7.34</td>
<td>0</td>
<td>0</td>
<td>20.38</td>
<td>0</td>
<td>15.41</td>
<td>48.56</td>
<td>0.0277</td>
</tr>
<tr>
<td>72</td>
<td>8</td>
<td>9.23</td>
<td>0</td>
<td>0</td>
<td>9.12</td>
<td>0</td>
<td>14.78</td>
<td>58.87</td>
<td>0.0261</td>
</tr>
<tr>
<td>74</td>
<td>4.92</td>
<td>6.58</td>
<td>0</td>
<td>0</td>
<td>18.3</td>
<td>1.1</td>
<td>15.43</td>
<td>53.67</td>
<td>0.0235</td>
</tr>
<tr>
<td>76</td>
<td>19.44</td>
<td>10.75</td>
<td>0</td>
<td>0</td>
<td>5.45</td>
<td>0</td>
<td>16.87</td>
<td>47.49</td>
<td>0.0290</td>
</tr>
<tr>
<td>78</td>
<td>16.53</td>
<td>5.48</td>
<td>0</td>
<td>0</td>
<td>28.83</td>
<td>0</td>
<td>16.62</td>
<td>32.54</td>
<td>0.0453</td>
</tr>
<tr>
<td>80</td>
<td>11.56</td>
<td>9.79</td>
<td>0</td>
<td>0</td>
<td>6.45</td>
<td>0</td>
<td>15.7</td>
<td>56.5</td>
<td>0.0201</td>
</tr>
<tr>
<td>82</td>
<td>17.96</td>
<td>10.94</td>
<td>0</td>
<td>0</td>
<td>16.63</td>
<td>0</td>
<td>17.34</td>
<td>37.12</td>
<td>0.0451</td>
</tr>
<tr>
<td>84</td>
<td>12.57</td>
<td>10.58</td>
<td>0</td>
<td>0</td>
<td>14.2</td>
<td>1.81</td>
<td>16.86</td>
<td>43.98</td>
<td>0.0317</td>
</tr>
<tr>
<td>86</td>
<td>8.46</td>
<td>8.17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14.13</td>
<td>69.24</td>
<td>0.0267</td>
</tr>
<tr>
<td>88</td>
<td>20.08</td>
<td>11.6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15.76</td>
<td>52.56</td>
<td>0.0335</td>
</tr>
<tr>
<td>90</td>
<td>7.74</td>
<td>8.48</td>
<td>0</td>
<td>0</td>
<td>7.45</td>
<td>0</td>
<td>13.24</td>
<td>63.09</td>
<td>0.0217</td>
</tr>
<tr>
<td>92</td>
<td>3.58</td>
<td>4.56</td>
<td>0</td>
<td>0</td>
<td>24.59</td>
<td>4.26</td>
<td>14.05</td>
<td>48.95</td>
<td>0.0198</td>
</tr>
<tr>
<td>94</td>
<td>12.64</td>
<td>5.25</td>
<td>0</td>
<td>0</td>
<td>25.61</td>
<td>0.56</td>
<td>15.18</td>
<td>40.76</td>
<td>0.0204</td>
</tr>
<tr>
<td>96</td>
<td>8.78</td>
<td>8.14</td>
<td>0</td>
<td>0</td>
<td>2.24</td>
<td>0</td>
<td>13.44</td>
<td>67.4</td>
<td>0.0254</td>
</tr>
</tbody>
</table>

Table 4: Total BeStSel secondary structure analysis of LGP7 BP, from 20°-96°C (Micsonai et al., 2015)
<table>
<thead>
<tr>
<th>Temperature °C</th>
<th>Helix 1</th>
<th>Helix2</th>
<th>Anti1</th>
<th>Anti2</th>
<th>Anti3</th>
<th>Para</th>
<th>Turn</th>
<th>Others</th>
<th>NRMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0</td>
<td>5.24</td>
<td>4.02</td>
<td>20.36</td>
<td>31.61</td>
<td>0</td>
<td>14</td>
<td>24.77</td>
<td>0.076</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>4.81</td>
<td>7.63</td>
<td>19.92</td>
<td>44.98</td>
<td>0</td>
<td>14.96</td>
<td>7.7</td>
<td>0.0569</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
<td>8.24</td>
<td>5.36</td>
<td>27.34</td>
<td>28.36</td>
<td>0</td>
<td>13.98</td>
<td>16.73</td>
<td>0.0366</td>
</tr>
<tr>
<td>26</td>
<td>0</td>
<td>6.38</td>
<td>5.46</td>
<td>22.52</td>
<td>31.78</td>
<td>0</td>
<td>13.76</td>
<td>20.11</td>
<td>0.0524</td>
</tr>
<tr>
<td>28</td>
<td>0</td>
<td>6.4</td>
<td>4.6</td>
<td>22.7</td>
<td>29.28</td>
<td>0</td>
<td>12.55</td>
<td>24.46</td>
<td>0.0585</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>3.33</td>
<td>2.96</td>
<td>15.72</td>
<td>36.94</td>
<td>0</td>
<td>15.43</td>
<td>25.62</td>
<td>0.1081</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
<td>5.14</td>
<td>2.81</td>
<td>19.84</td>
<td>30.4</td>
<td>0</td>
<td>15.47</td>
<td>26.35</td>
<td>0.0791</td>
</tr>
<tr>
<td>34</td>
<td>0</td>
<td>7.35</td>
<td>5.34</td>
<td>24.36</td>
<td>18.36</td>
<td>0</td>
<td>12.34</td>
<td>32.25</td>
<td>0.0458</td>
</tr>
<tr>
<td>36</td>
<td>0</td>
<td>4.41</td>
<td>4.92</td>
<td>18.43</td>
<td>39.8</td>
<td>0</td>
<td>15.01</td>
<td>17.43</td>
<td>0.0674</td>
</tr>
<tr>
<td>38</td>
<td>0</td>
<td>3.54</td>
<td>4.79</td>
<td>15.69</td>
<td>37.69</td>
<td>0</td>
<td>11.78</td>
<td>26.5</td>
<td>0.0895</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
<td>6.3</td>
<td>2.82</td>
<td>21.86</td>
<td>24.94</td>
<td>0</td>
<td>15.28</td>
<td>28.8</td>
<td>0.0604</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
<td>6.61</td>
<td>4.88</td>
<td>22.45</td>
<td>31</td>
<td>0</td>
<td>14.99</td>
<td>20.07</td>
<td>0.0398</td>
</tr>
<tr>
<td>44</td>
<td>0</td>
<td>8.5</td>
<td>6.05</td>
<td>26.72</td>
<td>24.04</td>
<td>0</td>
<td>15.43</td>
<td>19.26</td>
<td>0.0508</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>7.59</td>
<td>5.73</td>
<td>23.41</td>
<td>27.11</td>
<td>0</td>
<td>12.72</td>
<td>23.44</td>
<td>0.0372</td>
</tr>
<tr>
<td>48</td>
<td>0</td>
<td>7.57</td>
<td>4.29</td>
<td>22.99</td>
<td>27.75</td>
<td>0</td>
<td>14.64</td>
<td>22.76</td>
<td>0.0358</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>4.65</td>
<td>6.34</td>
<td>15.52</td>
<td>41.96</td>
<td>0</td>
<td>13.4</td>
<td>18.13</td>
<td>0.0446</td>
</tr>
<tr>
<td>52</td>
<td>0</td>
<td>5.72</td>
<td>6.39</td>
<td>19.9</td>
<td>27.14</td>
<td>0</td>
<td>10.56</td>
<td>30.3</td>
<td>0.042</td>
</tr>
<tr>
<td>54</td>
<td>0</td>
<td>6.72</td>
<td>6.37</td>
<td>24.29</td>
<td>24.36</td>
<td>0</td>
<td>13.99</td>
<td>24.27</td>
<td>0.0273</td>
</tr>
<tr>
<td>56</td>
<td>0</td>
<td>4.64</td>
<td>13.05</td>
<td>3.03</td>
<td>20.39</td>
<td>0</td>
<td>16.61</td>
<td>42.26</td>
<td>0.0247</td>
</tr>
<tr>
<td>58</td>
<td>0</td>
<td>0</td>
<td>9.87</td>
<td>0</td>
<td>28.37</td>
<td>0</td>
<td>14.74</td>
<td>47.03</td>
<td>0.0703</td>
</tr>
<tr>
<td>60</td>
<td>0</td>
<td>2.71</td>
<td>11.7</td>
<td>0</td>
<td>29.88</td>
<td>0</td>
<td>14.41</td>
<td>41.3</td>
<td>0.0399</td>
</tr>
<tr>
<td>62</td>
<td>0</td>
<td>3.43</td>
<td>11.69</td>
<td>0</td>
<td>24.15</td>
<td>0</td>
<td>12.63</td>
<td>48.1</td>
<td>0.03</td>
</tr>
<tr>
<td>64</td>
<td>0</td>
<td>1.26</td>
<td>9.98</td>
<td>0</td>
<td>33.79</td>
<td>0</td>
<td>14.84</td>
<td>40.14</td>
<td>0.0507</td>
</tr>
<tr>
<td>66</td>
<td>0</td>
<td>2.43</td>
<td>10.21</td>
<td>0</td>
<td>30.21</td>
<td>0</td>
<td>14.57</td>
<td>42.58</td>
<td>0.0473</td>
</tr>
<tr>
<td>68</td>
<td>0</td>
<td>3.06</td>
<td>12.28</td>
<td>0</td>
<td>24.53</td>
<td>0</td>
<td>14.9</td>
<td>45.23</td>
<td>0.026</td>
</tr>
<tr>
<td>70</td>
<td>0</td>
<td>2.69</td>
<td>11.05</td>
<td>0</td>
<td>26.87</td>
<td>0</td>
<td>14.14</td>
<td>45.25</td>
<td>0.0213</td>
</tr>
<tr>
<td>72</td>
<td>0</td>
<td>1.63</td>
<td>10.64</td>
<td>5.31</td>
<td>25.85</td>
<td>0</td>
<td>12.97</td>
<td>43.6</td>
<td>0.0369</td>
</tr>
<tr>
<td>74</td>
<td>0</td>
<td>4.65</td>
<td>12.52</td>
<td>9.25</td>
<td>11.15</td>
<td>0</td>
<td>11.48</td>
<td>50.95</td>
<td>0.0207</td>
</tr>
<tr>
<td>76</td>
<td>0</td>
<td>1.99</td>
<td>12.64</td>
<td>0</td>
<td>28.15</td>
<td>0</td>
<td>13.56</td>
<td>43.66</td>
<td>0.0279</td>
</tr>
<tr>
<td>78</td>
<td>0</td>
<td>1.24</td>
<td>10</td>
<td>0</td>
<td>29.94</td>
<td>0</td>
<td>15.21</td>
<td>43.62</td>
<td>0.0494</td>
</tr>
<tr>
<td>80</td>
<td>0</td>
<td>3.28</td>
<td>12.88</td>
<td>0</td>
<td>26.28</td>
<td>0</td>
<td>12.73</td>
<td>44.83</td>
<td>0.0308</td>
</tr>
<tr>
<td>82</td>
<td>0</td>
<td>1.54</td>
<td>13.53</td>
<td>0</td>
<td>23.81</td>
<td>0</td>
<td>18.48</td>
<td>42.64</td>
<td>0.0397</td>
</tr>
<tr>
<td>84</td>
<td>0</td>
<td>0.34</td>
<td>10.69</td>
<td>0</td>
<td>29.06</td>
<td>0</td>
<td>15.69</td>
<td>44.22</td>
<td>0.049</td>
</tr>
<tr>
<td>86</td>
<td>0</td>
<td>3.01</td>
<td>11.57</td>
<td>0</td>
<td>20.87</td>
<td>0</td>
<td>14.55</td>
<td>50</td>
<td>0.0362</td>
</tr>
<tr>
<td>88</td>
<td>0</td>
<td>2.61</td>
<td>11.77</td>
<td>0</td>
<td>25.13</td>
<td>0</td>
<td>14.76</td>
<td>45.72</td>
<td>0.0365</td>
</tr>
<tr>
<td>90</td>
<td>0</td>
<td>2.83</td>
<td>9.32</td>
<td>0</td>
<td>26.85</td>
<td>0</td>
<td>14.24</td>
<td>46.76</td>
<td>0.0405</td>
</tr>
<tr>
<td>92</td>
<td>0</td>
<td>4.21</td>
<td>14.11</td>
<td>0</td>
<td>18.39</td>
<td>0</td>
<td>16.6</td>
<td>46.69</td>
<td>0.0242</td>
</tr>
<tr>
<td>94</td>
<td>0</td>
<td>6.73</td>
<td>11.76</td>
<td>0</td>
<td>15.46</td>
<td>0</td>
<td>20.22</td>
<td>45.84</td>
<td>0.0193</td>
</tr>
<tr>
<td>96</td>
<td>0</td>
<td>0.96</td>
<td>8.61</td>
<td>0</td>
<td>38.9</td>
<td>0</td>
<td>16.91</td>
<td>34.61</td>
<td>0.0613</td>
</tr>
</tbody>
</table>

Table 5: Total BeStSel secondary structure analysis of Rmbe2 from 20°C-96°C (Micsonai et al., 2015)
Figure 3.5.2.4 Comparison of change in secondary structure of LGP7 BP and RMBC2 (lower graph), with data from CD analyzed through BeStSel (Micsonai et al., 2015).
Though the RMBC2 data is noisier, it is apparently that upon melting, LGP7 becomes unstructured, in a classical model of melting, whereas RBMC2 maintains a high beta sheet character (Anti 1 and 3, or left twisted and right twisted) though it does also become more unstructured. Interestingly, Anti 2 or relaxed beta sheet character dropped drastically right at the melting temperature of RBMC2, and the right and left twisted beta sheet character rose simultaneously. This observation was also noted in the paper from the Olson lab (De et al., 2018). They suggest that this indicates the formation of amyloids.
3.6 Alignment of LGP7 BP to Vibrio cholerae Beta Prisms:

Alignment of known beta prism domains to LGP7 shows it conserves all but one of the important sugar binding residues of VCC, but only about half of those involved in sugar binding of RMBC2. This is an indication that LGP7 BP might bind with a similar strength and specificity of VCC, as opposed to the ~ 80 fold stronger and less specific binding of RBMC2 (to asialofetuin: $1.5 \mu M$ vs $18.4 \text{nM}$), contrary to the ITC data.

Figure 3.6  alignment (with mega) of known beta prism domains. Residues important to binding highlighted by the arrows. Red arrow points out the residue in VCC (D617), which, when mutated, leads to a greater than 1000 fold loss of hemolysis. Residues specifically important to RMBC2 binding are highlighted in green. Residues important to VCC binding or both RMBC2 and VCC binding are shown in red.
3.6 Modelling

3.6.1 Modelling of LGP7 BP to VCC (PDB 4GX7) shows sugar binding residues are highly conserved

VCC has a 28.00 % identity to LGP7 BP in this alignment. In this model, modeled by SWISS PROT, residues G613, D617, F652, W706, L707 are aligned. In VCC, loss of these residues leads to a 10 to 100 fold loss of hemolytic activity. The one residue for sugar binding that is not conserved is Y654, which is replaced by another aromatic amino acid, tryptophan. The mutation of this tyrosine to alanine in VCC leads to a ~10 fold loss in hemolytic activity (De et al., 2018). However, this exact mutation occurs in RMBC2, and is suggested to lead to stronger binding due to greater ring stacking with oligosaccarides (De et al., 2018), so based on that logic, I this mutation would lead to slightly increased affinity as it does for RBMC2. The RMS between the predicted structure and that of VCC is .082, showing that they are predicted to be highly similar by SWISS MODEL.
3.6.2 Modelling of LGP7 BP to RBMC2 (PDB 5V6F) shows some similarities but significant differences in sugar binding residues

In the alignment of LGP7 BP to RMBC2, less of the residues important for binding are conserved. The residues that are conserved are D853 (as shown in the sequence alignment), V874 and W896, both of which VCC does not preserve and participates in direct binding with the sugar, and W948 (De et al., 2018). The residues that were not conserved were: F850, which is an alanine, just like VCC, the PVQGT loop, which is thought to make binding stronger, including the T870 residue, N871 right outside this loop, which is glutamic acid for LGP7 BP and could therefore participate in similar binding but carries a stronger charge, R876, and Y894, which is phenylalanine LGP7 BP, as well as for VCC, suggesting that LGP7 BP will also not participate in hydrogen bonding to the sugar with this residue, just like VCC (De et al., 2018). As there is no current model to measure the activity of RMBC2, the effect of the mutation of these residues is impossible to measure. However, one might still

Figure 3.6.2 SWISS MODEL prediction of LGP7 BP structure, modelled on the crystal structure of RMBC2 PDB 5V6F, aligned with a crystal structure of RMBC2 (PDB 5V6F) bound to mannotriose (Biasini et al., 2014)
deduce that LGP7 BP might bind sugars in a more similar way to VCC than that of RMBC2, but the fact that it maintains some of the residues important to sugar binding that VCC does maintain could mean that it has a different specificity than VCC, or that it binds oligosaccharides with a slightly greater affinity than VCC. The RMS of this predicted structure to RBMC2 is .088, showing they are predicted to be slightly less similar than the VCC predicted structure, but still very similar.

3.6.3 Alignment of models and LGP7 BP modelled with Phyre2 shows difference in models

I additionally created a homology model based off all known similar structures with Phyre2 (Kelley et al., 2015). This program reports an 100% confidence level in this structure, even though it only covers 81% of the LGP7 BP sequence. The RMS between the VCC homology structure from SWISS MODEL and the Phyre2 structure was .861, showing they are pretty similar. The RMS between the RMBC2 homology model and the Phyre2 structure was .1 showing they are more similar. This suggests that model was mostly based on the RMBC2 crystal structure, which we can also see through looking at the top templates of the Phyre2 prediction. The next most confident templates were VCC, and then other plant (“Parkia platycephala seed lectin in complex with 5-bromo-4-2 chloro-3-indolyl-a-d-mannose”, “crystal structure of salt protein from Oryza sativa”) and even human lectins (“crystal structure of a prostate-specific wga16 glycoprotein lectin,2 form ii”, “crystal structure of human pancreatic secretory protein zg16p”). This algorithm specifically tries to use a diverse set of biologically relevant templates. These templates however might not be as biologically relevant as they are in eukaryotes.
This is reflected in the model, where certain loops could not be modelled in and show up as dashed lines. Phyre2 says that this means there is a potential error in the original templates. While drawing a parallel with human and boar proteins is interesting, it is not likely biologically relevant here. The RMS of the SWISS PROT structures aligned to each other was .796, showing that all of these structures are pretty similar. However, this alignment highlights the discrepancies of predicted structures based on their template(s), confirming the need for a crystal structure of this domain for true structural and function analysis.

Figure 3.6.3 Homology models of LGP7 BP aligned, with Phyre2 model in green (Kelley et al., 2015), SWISS PROT VCC model in pink, and SWISS PROT RMBC2 model in blue (Biasini et al., 2014)
Chapter 4 - Discussion

4.1 Isothermal Titration Calorimetry Suggests that LGP7 BP may not mammalian glycans, in contrast with the model

The most unexpected result of this study is the fact that LGP7 BP does not bind asialofetuin (assuming the experiment did not fail for technical reasons). Of the Vibrio beta prism domains that have been studied in the Olson lab, both RBMC2 and VCC, bind asialofetuin fairly strongly, with micromolar to nanomolar affinity. The modelling affirms that LGP7 BP conserves all of the predicted important sugar binding residues of VCC, besides the Y654W mutation, which is also found in RBMC2. This could suggest that for a VCC-like beta prism lectin domain, this tyrosine cannot be replaced by the larger, bulkier tryptophan without disrupting the sugar binding pocket. Alternatively, Y654 could be participating in important hydrogen bonding with the sugar instead of the supposed nonpolar interaction, which would make its replacement with a tryptophan inactivating of mammalian sugar binding.

Outside of the preservation of the important sugar binding residues of VCC, LGP7 BP only shares maximum 28% identity with VCC, and is found in an oyster pathogen Vibrio crassostrea, as opposed to a human pathogen, Vibrio cholerae. Just as RBMC2 binds sugars completely differently than VCC, it is likely that this novel beta prism lectin domain has a completely novel sugar specificity and function. This is supported by the fact that it is found attached to a protein with a likely novel structural fold.
As previously outlined, lectins bind many different sugars outside of mammalian N-glycans. Further analysis of LGP7 BP with ITC will elucidate what kinds of sugars it could bind, and therefore its function. Carbohydrates are found on the outside of almost all cells, and it is plausible that this lectin binds one of them. Instead of mammalian sugars, it could bind oyster sugars and therefore inactivate its immune system, or algae sugars to form a sort of biofilm in the ocean.

To ensure that ITC data are accurate, more care will be taken to have a pure sample of LGP7 BP, and properly concentrated asialofetuin.

4.2 Circular dichroism confirms that LGP7 BP is a well-folded majority antiparallel beta sheet protein, and that it is likely similar to RBMC2

Another reason that LGP7 BP does not bind asialofetuin could have been that it was not well folded, as *E. coli* does not have all of the same folding and secretion systems of *Vibrio crassostrea*. However, its high *T*<sub>m</sub> of 52.8 °C, close to that of RBMC’s *T*<sub>m</sub> of 53.5 °C, confirms it is well folded. Additionally, the calculated secondary structures over temperatures show that it is likely similarly folded to RBMC2. They are nearly identical in secondary structure content at 20°C, shown in Figure 4.2 and Table 7. However, it is apparent that LGP7 BP’s melting does not follow the atypical, possible amyloid formation of RBMC2, as RBMC2 stays retains high beta character while LGP7 BP melts classically to random coil (“other” and “turn”). Also shown is data from De et. al, 2018 on the melting of RBMC2. This data appears to be slightly more accurate, with a much lower RMSD at 96°C of .18 as opposed to .34, so this model of RBMC2 melting is slightly more accurate. This means that RBMC2 likely has no parallel beta sheet character at 96°C. However, both
analyses agree on the high right and left twisted beta sheet character of RBMC2, with no relaxed beta sheet character.

Figure 4.2: Secondary structure changes of LGP7 BP over 20-96 °C, compared to that of RMBC2
Table 6: direct comparison of changes in secondary structure upon melting, between LGP7 and Rbmc2, analyzed by BeStSel (Micsonai et al., 2015). Data also shown from De et al 2018

<table>
<thead>
<tr>
<th></th>
<th>alpha helix</th>
<th>antiparallel beta sheet (left twisted)</th>
<th>antiparallel beta sheet (relaxed)</th>
<th>antiparallel beta sheet (right twisted)</th>
<th>parallel beta sheet</th>
<th>turn</th>
<th>other</th>
<th>RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RBMC2 20°C</strong></td>
<td>5.20%</td>
<td>4.00%</td>
<td>20.40%</td>
<td>31.60%</td>
<td>0%</td>
<td>14.00%</td>
<td>24.80%</td>
<td>.2106</td>
</tr>
<tr>
<td><strong>LGP7 BP 20°C</strong></td>
<td>3.90%</td>
<td>5.20%</td>
<td>17.20%</td>
<td>26.80%</td>
<td>0%</td>
<td>8.90%</td>
<td>38.10%</td>
<td>.1591</td>
</tr>
<tr>
<td><strong>RBMC2 96°C</strong></td>
<td>1%</td>
<td>8.60%</td>
<td>0.00%</td>
<td>38.90%</td>
<td>0%</td>
<td>16.90%</td>
<td>34.60%</td>
<td>.1844</td>
</tr>
<tr>
<td><strong>LGP7 BP 96°C</strong></td>
<td>5.10%</td>
<td>2.20%</td>
<td>2.90%</td>
<td>12.40%</td>
<td>0%</td>
<td>17.70%</td>
<td>59.70%</td>
<td>.0761</td>
</tr>
<tr>
<td><strong>RBMC2 20°C</strong></td>
<td>2.50%</td>
<td>3.60%</td>
<td>12.80%</td>
<td>27.20%</td>
<td>0%</td>
<td>14.70%</td>
<td>39.30%</td>
<td>.1702</td>
</tr>
<tr>
<td>(previously determined)</td>
<td>0%</td>
<td>20.1%</td>
<td>0%</td>
<td>20.7%</td>
<td>5.6%</td>
<td>16.7%</td>
<td>36.9%</td>
<td>.3746</td>
</tr>
</tbody>
</table>

4.3 LGP7 BP Crystallography

The results from the first crystal refinement screen demonstrates the need to repeat the condition 1.0 M succinic acid, 0.1 M HEPES, and 1% PEG 2K MME more exactly to repeat crystallization. In future screens, succinic acid will be used instead of sodium succinate hexahydrate dibasic. Additionally, according to the new screen, the protein is very sensitive to low pHs, as all conditions with pH below 7 showed precipitation. A new screen will be set up with more pH ranges from 7-7.5.

Additionally, a screen could be created with concentrations of succinic acid and HEPES that did not always decrease or increase concurrently.

The other JCSG+ conditions that showed promising crystalline precipitation include one with a lower concentration of succinic acid (0.1 M) and a heavier weight
PEG, PEG 3350, at 15%. The next screen will sample different percentages of PEG and heavier weight PEGs. There are no other apparent trends among the promising JSCG+ conditions (see table), but they demonstrate potential reagents to try if crystallization continues to need more optimization.

If these crystals can be repeated but remain small, seeding is a viable option to create one large crystal suitable for X-ray crystallography.

If the ITC study is correct so far, we expect for the X-ray crystallography structure to have a binding pocket that looks quite different.
4.4 Future Directions

We expect to proceed with crystallization optimization of LGP7 BP, and then solve its crystal structure. This will be done first by attempting to recreate the exact condition in which crystals were found. Next, a screen around this condition, within a pH range of 7 to 7.5 will be created, potentially using a different PEG. If this fails to create crystals, the other conditions mentioned will be screened with varying concentrations and pHs to try to create crystals. If small crystals are created, seeding will be performed to create bigger crystals. Once diffraction data is gathered, the structure will be solved, and potential binding residues and partners examined.

We also expect do to isothermal titration calorimetry binding studies of this protein to various sugars to elucidate its sugar binding, beginning with asialofetuin, and then methyl-mannose, and then other potential binding partners, like alginic acid, galactose, and other types of N-glycans. If LGP7 BP does show binding to N-glycans, a glycan chip screen may be performed.

Finally, we expect to try to elucidate the structure and function of the rest of LGP7 BPC, perhaps through expression in Vibrio crassostrea itself. This can be achieved through expression in the plasmid “pBAD/Myc-His”, from Thermo Fisher. This is an arabinose induced expression plasmid. If this still does not reliably produce the full length protein, expression of the domains in E. coli, likely in GFP vectors, will continue.
Chapter 5 - References


Technology. Retrieved from


https://doi.org/10.1371/journal.pntd.0004330


