Identification and characterization of adhesive proteins in freshwater mussels towards the development of novel bioadhesives

by

Judith Ng

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science
Institute of Biomaterials and Biomedical Engineering
University of Toronto

© Copyright 2018 by Judith Ng
Abstract

Identification and characterization of adhesive proteins in freshwater mussels towards the development of novel bioadhesives

Judith Ng
Master of Applied Science
Institute of Biomaterials and Biomedical Engineering
University of Toronto
2018

The European freshwater mollusk *Dreissena bugensis* (quagga mussel) adheres to a variety of underwater surfaces via the byssus, a proteinaceous anchor, from which threads are secreted. The threads terminate in small, disc-shaped adhesive plaques which have adapted to turbulent conditions with remarkable adhesive strength for their size. Top adhesive candidates that demonstrated evidence for an adhesive role were identified via LC-MS/MS and relative quantification of proteins between the mussel plaque and adhesive interface. For the first time, a top candidate for adhesive activity (Dbfp7) was isolated directly from the quagga mussel. We aim to incorporate relevant protein motifs from these proteins into peptide mimics and to recombinantly express Dbfp7, leading to the development of novel bioadhesives for medical and dental applications. Understanding the molecular mechanisms and proteins responsible for adhesion in freshwater mussels will provide a paradigm for medical adhesives which function reliably on wet tissue surfaces with minimal inflammation.
Acknowledgements

There are several individuals whom I would like to thank. First and foremost, I would like thank my mentor and supervisor, Prof. Eli Sone. It has been an absolute joy to work under his tutelage for the past three years. His ability to empathize with students, to guide and encourage both their strengths and shortcomings, and to be present during every crest and trough of their project has made my personal journey as a graduate student in his group an unparalleled learning experience.

I am very grateful to my committee members, Prof. Kevin Truong and Prof. Jonathan Rocheleau for their valuable feedback and thorough insight during the project. I would also like to thank Prof. Jayachandran Kizhakkedathu who welcomed me into a wonderfully productive and informative learning environment in his laboratory at UBC.

I would like to thank all my colleagues in the Sone Lab who have been instrumental to the success of my project. Thank you to Magdalena Wojtas for her unique expertise in protein purification and for helping me every step of the way in experiments that have made almost an entire chapter of my thesis possible. Thank you to David Rees for his patience, diligence, and excellent research abilities that have made my transition to carrying forth his work well-facilitated and seamless. Endless thanks to Alex Lausch and Bryan Quan for always being willing to advise, troubleshoot, and along with all members of the Sone Lab, making the lab environment an incredible place to be. Thank you to Kenny Kimmins, Kate Li, Denise Yap, Bryan James, Minh-Tam Nguyen and Angelico Obille for their dedication and hours put into maintaining our well-oiled mussel aquarium system and for helping collect the mussels that have made this work possible.

I would also like to acknowledge the labs that allowed access to their equipment and for their expertise, including Prof. Ben Ganss, Prof. Rocheleau, Prof. Radisic, Prof. Houry, and Prof. Master.

I would also like to thank Paul Taylor of the SickKids Proteomics, Analytics, Robotics, & Chemical Biology Centre (SPARC BioCentre) for trypsin digestion, LC-MS/MS analysis and analysis guidance. Thank you to Reynaldo Interior for extensive amino acid analysis assistance and guidance.

Last but not least, I would like to thank my family - my parents and my sister - for their unwavering support of all my goals, no matter how lofty, and for their constant and reassuring belief in my capabilities. The patience, love, and excitement they have always shown for my successes and lessons learned has made every hardship worthwhile.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 Mussel Biology and Adhesion</td>
<td>1</td>
</tr>
<tr>
<td>1.1.2 Marine Mussel Attachment</td>
<td>3</td>
</tr>
<tr>
<td>1.1.3 Recombinant and Peptide Mimic Experiments</td>
<td>5</td>
</tr>
<tr>
<td>1.1.4 Animals as Models of Adhesion</td>
<td>5</td>
</tr>
<tr>
<td>1.1.5 Mussels as Models for Adhesion</td>
<td>6</td>
</tr>
<tr>
<td>1.1.6 Existing Freshwater Mussel Byssal Proteins</td>
<td>7</td>
</tr>
<tr>
<td>1.2 Rationale and Motivation</td>
<td>10</td>
</tr>
<tr>
<td>1.2.1 Limitations of Current Adhesives</td>
<td>10</td>
</tr>
<tr>
<td>1.2.2 Weaknesses of Current Mussel-Inspired Adhesives</td>
<td>11</td>
</tr>
<tr>
<td>1.2.3 Current Progress: Candidate Spatial Verification and Molecular Weight</td>
<td>12</td>
</tr>
<tr>
<td>1.3 Project Objectives and Goals</td>
<td>13</td>
</tr>
<tr>
<td>1.3.1 Objective 1: Identify quagga mussel adhesive protein candidates and select the top candidate</td>
<td>13</td>
</tr>
<tr>
<td>1.3.2 Objective 2: Purify and characterize the top quagga mussel adhesive protein candidate</td>
<td>14</td>
</tr>
<tr>
<td>1.4 Document Overview</td>
<td>15</td>
</tr>
<tr>
<td>2 Quagga Mussel Adhesive Protein Identification</td>
<td>16</td>
</tr>
<tr>
<td>2.1 Abstract</td>
<td>16</td>
</tr>
<tr>
<td>2.2 Introduction</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Methods</td>
<td>19</td>
</tr>
<tr>
<td>2.3.1 Mussel Thread/Plaque and Footprint Collection</td>
<td>19</td>
</tr>
<tr>
<td>2.3.2 SDS-PAGE and Staining</td>
<td>19</td>
</tr>
</tbody>
</table>
### 2.3.3 Liquid Chromatography Tandem Mass Spectroscopy (LC-MS/MS) ... 19
### 2.3.4 Protein Identification, Selection, and Analysis .......................... 20
### 2.4 Results & Discussion ............................................................... 20
#### 2.4.1 Quagga Mussel Protein Distribution ........................................ 21
#### 2.4.2 Determination of Putative Protein Roles ................................... 22
#### 2.4.3 Byssal Proteins: All Sectors ................................................... 24
#### 2.4.4 Relative Protein Quantification of Footprint+Plaque Proteins ........ 26
#### 2.4.5 Relative Protein Quantification: Candidate Selection ................. 28
#### 2.4.6 Caveats of Experimental Design ............................................. 35
### 2.5 Conclusion .............................................................................. 37
### 2.6 Acknowledgements .................................................................. 37
### 2.A All Venn Diagram Sectors: Other Protein Roles ............................ 38
#### 2.A.1 Structural Proteins ................................................................. 38
#### 2.A.2 Mucin-like Proteins ................................................................. 38
#### 2.A.3 Redox Proteins .................................................................. 40
#### 2.A.4 Mussel Proteins ...................................................................... 40
#### 2.A.5 Protease Inhibitors ................................................................. 41
#### 2.A.6 Tyrosinase .................................................................. 41
#### 2.A.7 Cellular .................................................................. 42
### 2.B Relative Protein Quantification: Proteins of Interest ..................... 42
#### 2.B.1 Rhinophore Proteins ............................................................... 42
#### 2.B.2 Structural Proteins ................................................................. 42
#### 2.B.3 Mucus-like proteins ............................................................... 43
#### 2.B.4 Redox proteins .................................................................. 43
#### 2.B.5 Plasminogen-like Proteins ...................................................... 44
### 2.C Justification for Targeted Footprint+Plaque Overlap Search .......... 44
### 3 Dbfp7 Purification and Characterization ........................................ 48
#### 3.1 Abstract .............................................................................. 48
#### 3.2 Introduction .......................................................................... 49
#### 3.3 Methods ............................................................................ 50
##### 3.3.1 Protein Extraction ............................................................... 50
##### 3.3.2 Amino Acid Analysis ............................................................ 51
##### 3.3.3 MALDI-TOF MS Spectrometry ............................................. 51
##### 3.3.4 Liquid Chromatography Tandem Mass Spectroscopy (LC-MS/MS) 51
##### 3.3.5 Protein Identification, Selection, and Analysis ......................... 52
##### 3.3.6 Circular Dichroism ............................................................... 52
##### 3.3.7 Quartz Crystal Microbalance ............................................... 52
##### 3.3.8 Atomic Force Microscopy ..................................................... 53
#### 3.4 Results & Discussion ............................................................... 53
3.4.1 Phenol gland purification of mussel adhesive proteins .......................... 54
3.4.2 Seasonal Purification ........................................................................... 55
3.4.3 Size and secondary structure of native Dbfp’s ..................................... 57
3.4.4 Adsorption studies of Dbfp’s on SiO$_2$, TiO$_2$ and Au ......................... 62
3.5 Conclusion ............................................................................................. 67
3.6 Acknowledgements ................................................................................ 68

4 Discussion, Future Work & Conclusion .................................................. 69
  4.1 Future Work & Recommendations ......................................................... 69
    4.1.1 Dbfp7 Structural Characterization Optimization ............................... 69
    4.1.2 Dbfp7 Surface-Sensitive Characterization Optimization ................. 71
    4.1.3 Recombinant Expression of Dbfp7 .................................................. 73
    4.1.4 Expansion of Quagga Mussel Byssal Protein Knowledge .................. 73
  4.2 Significance & Conclusions .................................................................. 73

References ................................................................................................. 74
List of Tables

2.1 Quagga mussel Dbfp proteins found enriched and abundant at the adhesive interface. ........................................... 30
2.2 Large and abundant proteins identified by LC-MS/MS by Rees et al.[81] that did not originally meet acceptance criteria to be listed as a novel quagga mussel byssal protein, but were re-classified as potential byssal candidates due to the presence of Dopa within peptides. Bolded transcripts are both highly enriched and abundant at the footprint-plaque interface. ........................................... 31
2.3 Theoretical molecular weights of full Dbfp7 sequences and N- and C-terminal fragments found in naturally secreted footprint and plaque samples. .............................. 35
A1 Top 30 naturally secreted proteins ranked by decreasing relative enrichment at adhesive interface compared to the bulk plaque. Byssal proteins are bolded. ... 46
A2 Top 30 naturally secreted proteins abundant at the adhesive interface ranked by decreasing abundance by total spectral count. Byssal proteins are bolded. ... 47
3.1 Theoretical molecular weights of full Dbfp7 sequences and N- and C-terminal fragments found in Dbfp7 purified natively from quagga mussel phenol glands. Fold change ratio data in relative quantification experiments of footprint versus plaque samples are also shown. ........................................... 57
3.2 Amino acid analysis performed on Dbfp7 purified natively from quagga mussel phenol glands. ................................. 58
3.3 Dbfp7 fragment and substrate surface net charge at experimental pH conditions (pH 5.0). ........................................... 65
List of Figures

1.1 Lateral view (top) and dorsal-ventral view (bottom) of quagga and zebra mussels. 1

1.2 L-R: A quagga mussel attaching to an aquarium wall via the foot organ; a splayed open quagga mussel with the foot organ indicated; schematic of the byssal components (thread and plaque) extending from the mussel stem in white; SEM micrograph of a detached thread and plaque.) Adapted from Smith[10] with permission. 2

1.3 Interactions of Dopa with mica, polystyrene (PS), TiO$_2$, and rock surfaces via hydrogen bonding, aromatic interactions, and monodentate and bidentate coordination bonds. Adapted from Li et al.[21] with permission. 4

1.4 Artificially induced (KCl injection) secretion of byssal proteins. (a) A splayed open quagga mussel with the internal anatomical structures present. The foot is highlighted, into which KCl is injected to induce secretion of fresh byssal threads and plaques. (b) Magnification of the mussel foot with labels indicating mussel organ structure and protein secretion. (c) Gel electrophoresis of proteins artificially induced to be secreted from the mussel foot. Labels indicate previously identified Dopa-containing Dbfp proteins in addition to newly identified byssal proteins. Reprinted from Rees et al.[56] with permission from Taylor & Francis (www.tandfonline.com). 8

1.5 The plaque-substrate interface in the zebra mussel. (a) TEM micrograph of an ultrathin section of plaque deposited on an epoxy substratum. Arrow indicates the interface while similarly stained granules are visible within the plaque region above the interface, indicated by the asterisk. (b) The interfacial layer remains post-plaque removal. Adapted from Farsad and Sone[55], with permission from Elsevier. 9

1.6 Summary of approach from proteome to candidate. Full pipeline for determining potential quagga mussel adhesive protein candidates from artificially induced and naturally secreted mussel samples. 14
2.1 Two forms of byssal products from the quagga mussel. (a) Naturally secreted proteins collected by allowing mussels to secrete threads and plaques onto a glass slide in a controlled underwater environment. Adapted from Gilbert and Sone[54] with permission. (b) Artificially induced proteins triggered for release via KCl injection into the mussel foot. Uncrosslinked whole thread and plaque proteins rise to the surface of the mussel foot. Adapted from Rees et al.[81] with permission.

2.2 Area-proportional Venn diagram in the center, representing 385 unique proteins sourced via artificial induction and natural secretion. Numbers shown in each sector represent the number of proteins localized uniquely to the respective sector. The induced family of proteins is tinted yellow while the footprint and plaque proteins, which together compose the naturally secreted family of proteins, are shown in blue. Breakdown of the ten most ubiquitous protein categories is indicated in separate pie charts. Pie charts referring to proteins represented in more than one sector are highlighted for emphasis. Colour scheme follows identical labelling as Figure 2.3.

2.3 385 proteins sourced via artificial induction and natural secretion were sorted into a number of categories, sequentially narrowed by predicted or known roles based on confidence to existing proteins via BLAST results.

2.4 Dbfp proteins separated by representation across footprint, plaque, and induced protein samples. Values in parentheses represent total peptide counts for each Dbfp within each respective sector. Footprint and plaque Dbfp breakdowns shown outside of the venn diagram indicate total Dbfp counts within entire footprint and entire plaque sectors. Repeated Dbfp’s appearing within different sectors represent transcripts of Dbfp homologs unique to each sector (e.g. Dbfpα, Dbfpβ, etc.)

2.5 (a) 155 byssal proteins ranked in order of decreasing relative abundance at the adhesive interface compared to the bulk plaque. Values are displayed as fold change ratio. Black arrows indicate Dbfp proteins; labels indicate the specific Dbfp. (b) Top 30 proteins abundant at the adhesive interface relative to the bulk plaque, arranged in decreasing relative abundance at the adhesive interface. Proteins are grouped by families based on similar characteristics as determined by BLAST. Proteins with an asterisk (*) were also found among the list of 233 induced proteins in a separate LC-MS/MS experiment. Colour scheme follows identical labelling as Figures 2.2 and 2.3.
2.6 Top 30 proteins abundant at the adhesive interface ranked in decreasing order of abundance (Table A2) plotted by fold change ratio and total spectrum count. Dbfp7 is emphasized in red. comp53519_c2, with an unusually high spectrum count, has a top BLAST match to a mucus component. Dbfp proteins are shown in black and labelled above the relevant point. 30

2.7 LC-MS/MS peptide analysis shows variable localization of Dbfp7 peptide enrichment. Six transcripts of Dbfp7 present in the relative quantification experiment are represented by their sequence numbers (seq). All peptides appearing before the red line represent peptides enriched in the footprint; those appearing after represent peptides enriched in the plaque. 34

A1 CGLCGN motif conserved in FCγBP and mucins were present and align in both the query transcripts, their top BLAST hits, and the native protein associated with the top BLAST hit. 39

3.1 PAGE gels of separated D. bugensis proteins. (a) Sequentially isolated fractions (Supernatants S1, S2 of the original homogenized aliquot of D. bugensis phenol glands in extraction buffer (5% acetic acid with protease inhibitors). Gel was run on 12% Bis-Tris gel and visualized with Coomassie blue. The band of interest is highlighted in red. (b) Induced quagga mussel byssal visualized with silver stain, performed by a colleague[56]. Protein bands marked as novel indicate quagga mussel proteins not previously identified by researchers who had isolated crosslinked, Dopa-containing byssal proteins. 54

3.2 RP-HPLC chromatogram of partially purified quagga mussel phenol gland extract Supernatant-2 (S2). (a) RP-HPLC chromatograms were detected at 220 nm (blue line) and 280 nm (black line). A gradient of acetonitrile was employed. Highlighted regions in yellow represent fractions that were further analyzed by gel electrophoresis. Fractions pooled from 38-48 min correspond to Dbfp7 elution. (b) SDS-PAGE of representative column fractions in (a) coinciding with HPLC peaks. Ladder with stained protein standards is located in the left lane with MW standards whose apparent molecular masses are given in kDa. Numbers above the lanes correspond to elution time and fraction number in (a). Sections marked as A and B represent the same sections highlighted as A and B in (a). 55

3.3 Dbfp7 purification efficiency is maximal within 24 hours of mussel retrieval from native conditions. (a) Chromatographic peak area integration versus time between mussel obtainment and mussel detachment for Dbfp7 extracted directly from the phenol gland and separated by HPLC; (b) for Dbfp7 extracted directly from the phenol gland, pre-purification with a 50 kDa syringe filter, and separated by HPLC. 56
3.4 Select (6/11) amino acid sequences of phenol gland-derived Dbfp’s shown with corresponding pI and content of Tyr (Y in red). Sequences represented are those found with a footprint:plaque fold change ratio present in relative quantification experiments. ................................................................. 58
3.5 Representative MALDI-TOF MS spectra of purified native Dbfp7 demonstrating the absence of consistent Dbfp7 polymorph separation in separate fractions by RP-HPLC. (a) MALDI-TOF spectra of all fractions pooled together, purified from RP-HPLC from 4000 to 10,000 kDa. The main molecular weight families represented are within the 6-8 kDa region. (b) Representative MALDI-TOF spectra of one fraction purified from RP-HPLC from 4000 to 20,000 kDa. ........ 59
3.6 (a) Relative percent intensities of four most consistently present families of protein sizes in kilodaltons (kDa). (b) Representative HPLC spectra indicating fractions isolated from RP-HPLC for MALDI-TOF analysis. Each division indicates a fraction that was analyzed by MALDI-TOF, numbered for the time at which it was eluted, in minutes (44-53). ................................................................. 60
3.7 CD spectra of 500 µg ml⁻¹ Dbfp7 in 20 mM acetate buffer showing mean residue ellipticity [θ] as a function of wavelength. ................................................................. 61
3.8 (a, b) Representative AFM images of 150µg ml⁻¹ Dbfp7 in 0.1 M acetate buffer (pH 5.0) adsorbed on a SiO₂-coated silicon wafer. (c) AFM image of untreated silicon wafer. ................................................................. 62
3.9 (a) ∆f and ∆D for QCM-D measurements performed with Dbfp7 adsorption onto SiO₂, TiO₂, and Au sensors. (b) Evolution of the change in the effective thickness, obtained from the best fit between the viscoelastic model (obtained using Q-Soft), and the experimental data. ................................................................. 63
3.10 Dbfp7 N-terminal fragment charge distribution in solution at pH 5.0. .......... 66
Chapter 1

Introduction

1.1 Background

1.1.1 Mussel Biology and Adhesion

The zebra mussel (*Dreissena polymorpha*) and closely related quagga mussel (*Dreissena bugensis*) (Figure 1.1) are members of a small family of bivalve molluscs which, until the 19th century, were restricted to the Black, Caspian, and Azov Seas[1]. Their exemplary ability for mussel attachment has enabled their spread and colonization of the Great Lakes. These invasive species are known for their rapid spread in the waterways of eastern North America due to their ability to adhere to a variety of surface chemistries underwater and their reputation as a nuisance species[2, 3].

![zebra mussel and quagga mussel](image)

**Figure 1.1** – Lateral view (top) and dorsal-ventral view (bottom) of quagga and zebra mussels.

Despite considerable separation in their evolution, marine and freshwater mussels share macrostruc-
tural similarities. These include features such as a byssal retractor muscle, a root which connects the threads to retractor mussels within the shells, a stem extending from the muscular root of the byssus, and individual byssal threads which are attached to the root and extend forth, terminating at the distal end of the thread in adhesive pads called plaques, which attach to the substrate[4, 5] (Figure 1.2). The byssus functions as the proteinaceous anchor and is secreted by the foot of the mussel, which contains a ventral groove where threads and plaques are first formed. Threads and plaques are covered by a highly cross-linked, protective coating[6, 7]. Mussel adhesives permit mussels to attach strongly to various surfaces in turbulent, wet environments of high ionic strength[8]. Secreted in liquid form and undergoing solidification via crosslinking to form the byssal thread and plaque, the mussel is able to anchor to a foreign surface via the byssus, a structure comprised of exocrine byssal glands embedded in the mussel’s foot. Mussel binding activity must be expeditiously opportunistic due to factors including water turbulence, persistent surf and tides, immersion, and competition for space and predators[9].

**Figure 1.2** – L-R: A quagga mussel attaching to an aquarium wall via the foot organ; a splayed open quagga mussel with the foot organ indicated; schematic of the byssal components (thread and plaque) extending from the mussel stem in white; SEM micrograph of a detached thread and plaque.) Adapted from Smith[10] with permission.

The underwater adhesion capabilities of mussels and sessile marine invertebrates are an increasingly attractive paradigm for designing water-resistant adhesives. This is largely due to their remarkable ability to adhere underwater, which is typically hindered by the high dielectric constant of water ($\epsilon = 80$), thus greatly diminishing the magnitude of interaction energies possible in water[11]. Adhesive performance underwater is hampered by four pathways: water presenting itself as a weak boundary layer at interfaces, the wicking of water into interfaces, the erosion of the adhesive, and the swelling of adhesive by water absorption[8]. The presence of this weak boundary layer that forms between water and a substrate prevents direct contact between the adhesive and the adherend, leaving cohesive strength with the boundary layer itself as the strongest link in bond performance, rather than the adhesive capability itself[12]. To combat these impeding pathways, mussel adhesion involves the removal of weak boundary layers and thus ensures the spread of the adhesive, the development of numerous and strong interfacial contacts, and curing of the adhesive[8].
Chapter 1. Introduction

Precursors of thread and plaque proteins of marine mussels are stockpiled in different locations of the foot, over ten of which are produced in the phenol gland, though likely not in the same cells[13]. Mussel adhesive protein synthesis occurs via a regulated protein secretion pathway[13]. Protein synthesis occurs in the endoplasmic reticulum (ER), inside which many enzymes which covalently-modified amino acid residues reside. The nascent protein is sequestered into ER regions which bud off into membrane-bound vesicles. Byssal adhesives, as proteins destined for regulated secretion pathways, migrate towards the Golgi apparatus and bud off the trans portion of the Golgi, fusing with each other to form large vacuoles which serve as reservoirs for accumulating protein[13]. The proteins are stockpiled within the foot organ and then secreted and/or released into the byssal groove, providing a template for thread and plaque formation[4].

1.1.2 Marine Mussel Attachment

Marine mussels have received much attention in studies of proteins responsible for adhesion, the putative adhesive mechanism, and in the development of synthetic adhesive materials which mimic the structure of, or contain motifs of, mussel proteins known to demonstrate adhesive behaviour. Mussel adhesive proteins are known as a particularly powerful source of adhesion due to their long-lasting strength in wet environments[12] on virtually all types of organic and inorganic materials while maintaining flexibility and elasticity[4, 14]. Marine mussels are typically 4-5 cm long and attach to a substrate with 50-100 protein threads, each approximately 2-4 cm long and 0.1 mm in diameter[13, 15]. Plaques are 0.15 mm in diameter at the thread-plaque interface and 2-3 mm at the plaque-substrate interface[4]. Marine mussels, including Geukensia demissa, Mytilus edulis, and Bathymodiolus childressi have been observed to retain the byssus in their adult stage, thus permitting them to live entirely exposed attached to hard surfaces. The blue mussel (Mytilus edulis) has been shown to contain M. edulis foot protein 5 (Mefp-5), which bonds to glass, plastic, wood, concrete, and Teflon[4]. The strength of mussel adhesive proteins is attributed to the cross-linking that occurs between polymer chains of individual adhesive proteins, with forces of 250-300 N/mussel required to dislodge the California mussel Mytilus californianus, while individual threads have detachment forces of 5-6 N/thread[16, 17]. Marine mussels have also been demonstrated to stick with strengths of almost 300 kilopascal (kPa)[18]. Considering the mussels’ size and ability to adhere underwater, these are impressive values even in comparison to those of polyvinyl acetate (PVA, “white glue”) at 900 kPa and cyanoacrylate (superglue) at 8000 kPa[19].

The curing and hardening process of marine mussel adhesives is largely due to the presence of a catecholic precursor such as 3,4-dihydroxyphenylalanine (Dopa), formed by post-translational hydroxylation of tyrosine[4, 20]. It is able to interact through hydrogen bonding, metal-catechol coordination, electrostatic interactions, cation-π and π-π interactions, and aromatic interac-
tions, all of which contribute to marine mussel ability to adhere in wet adhesion conditions[21], the interactions of which are summarized in Figure 1.3. This breadth of bonding phenomena explains the mussel’s promiscuity in adherence patterns. Its ability to adhere to rocks and glass is likely due to the ability for Dopa to complex with metal ions, oxides, and semimetals[22], while cation-π and π-π stacking allows for adhesion to more hydrophobic surfaces[21].

![Figure 1.3](image.png)

**Figure 1.3** – Interactions of Dopa with mica, polystyrene (PS), TiO$_2$, and rock surfaces via hydrogen bonding, aromatic interactions, and monodentate and bidentate coordination bonds. Adapted from Li et al.[21] with permission.

Although the mussel adhesion mechanism is not fully understood, it appears to be dependent on Dopa content[12], redox chemistry [23, 24], and metal interactions[25, 26, 27]. Adhesive proteins from marine mussel byssal thread and plaque commonly contain post-translational modifications. The most modified byssal proteins include Mefp-3 and Mefp-5, in which 42% and 37% of amino acids are modified, respectively[13]. Modifications include O-phosphoserine and 4-hydroxyarginine[12]; however the primary modification among byssal proteins is tyrosine hydroxylation to 3,4-dihydroxyphenylalanine (Dopa)[13]. The significance of these modifications is logical as Dopa and phosphoserine both bolster adhesive interactions with mineral and metal surfaces that exceed noncovalent possibilities in water[13].
1.1.3 Recombinant and Peptide Mimic Experiments

Producing peptide mimics of protein sequences with tandem repeats of consensus sequences of interest is a common strategy to better understand marine mussel byssal proteins in their role for applications in biological adhesives. This has been done previously to elucidate the structure and interactions of these repeats [28]. This has immense potential for commercial applications downstream, as this has been done for *Mytilus galloprovincialis* (Mediterranean mussel) wherein a fusion peptide (fp-151) consisting of an fp-5 adhesive protein sequence from the mussel was flanked with copies of the decapptide repeat of the mussels cuticle protein [29]. This mimic displayed macro-scale adhesion and biocompatibility [29]. Mussel-mimetic tissue adhesives have extensive biomedical applications such as use in fetal membrane repair [30, 31]. Expansions on this research can inspire fouling-resistant technologies to prevent mussels from attaching to underwater structures and contribute to coatings that decrease drag, fuel consumption, and maintenance [32].

The currently existing approaches, such as Messersmith’s strategy via chemical synthesis and conjugation - polymers of poly(ethylene glycol) capped with Dopa - are practical, easy to scale, and are several times stronger than fibrin-based glues [33]. However, toxicity and biodegradability of the materials remains to be confirmed. In the context of a surgical adhesive to be used in live humans, this represents a significant limitation. Conversely, the approach of Hwang et al. has been to functionally express recombinant mussel foot proteins in *E. coli* [29]. Both methods exploit the chemical prowess of Dopa interaction with a wide variety of substrates and circumvent the limitations of purification directly from the mussel; however, complete biocompatibility and biodegradability, hallmarks of a truly functional surgical adhesive, is an ongoing challenge.

1.1.4 Animals as Models of Adhesion

Naturally produced adhesives are ubiquitous among biological systems and are known for their superior strength in comparison to man-made materials. These include adhesive research based in the biological systems beyond those of mussels, including spiders, marine tubeworms, barnacles, and mussels. Adhesives for wound closure and healing are increasingly inspired by adhesives from nature, from proteins in the feet of gecko lizards to patches fabricated from quill-inspired plastic needles, which provide optimal force balance on tissues [34]. Surgical glues have also been modeled after secretions of the sandcastle worm, which have been advanced to create biomimetic adhesives that can hold together bone fragments in rats with craniofacial defects and mend ruptures in *in vitro* human tissue [35, 36]. *Pomphorhynchus laevis*, a fish parasite known as the spiny-headed worm, possesses a proboscis which swells with water to secure the worm to the intestinal wall of its host, which has inspired the design of polystyrene microneedles that when arranged into a 100-needle array, are able to stick to pig skin with four
times the strength of a metal staple[34]. Other specialized biological systems used in adhesives research include bacteria, fungi, spiders, sea cucumbers, worms, algae, fish, holothurians, arthropods, and barnacles[4, 37].

1.1.5 Mussels as Models for Adhesion

Marine mussels secrete natural glues to fasten themselves to rocks and ship hulls; thus the adhesives produced by mussels must maintain fluidity while being secreted from the secretory organ to the extracellular space, and then must undergo self-assembly and irreversible curing to form a durable adhesive joint[38]. Dopa, a hydroxylated derivative of tyrosine, crosslinks matrix proteins within mussel secretions to effectively cure the adhesive[34]. Particular attention has been paid to the *Mytilus* genus, where *Mytilus* foot proteins (mfp) are post-translationally modified with Dopa. Despite the independent evolution of freshwater and marine mussels, the byssi of freshwater mussels (zebra and quagga mussels in particular) are composed of proteins containing the rare amino acid Dopa, indicating that Dopa plays an integral role in the mussels adhesive properties[39]. However, the lower levels of Dopa in freshwater mussels indicate that a novel mechanism for mussel adhesion in zebra and quagga mussels may exist.

Mussel adhesive protein attachment to living substances such as porcine skin[40] and mammalian cells have been investigated[14, 29, 41, 42]. Water-insoluble mussel adhesives are biodegradable[12] and not immunogenic, thus making them an attractive option for use within the body. Functionalizing polymers with catechol-containing analogues has also been studied[43, 44]. Under oxidative conditions, the conversion of catechol groups produces highly reactive quinones that enable the polymer to adhere strongly to wet polymers and form gels in saline environments[35, 45]. A polyethylene glycol polymer mimic of marine mussel glue that was catechol-functionalized has been successfully used in pancreatic islet transplantation in mice with *in vivo* stability, good tissue integration, and no inflammatory response[46]. Incorporation of Dopa and Dopa-mimetic catechols into hydrogel models have seen success in *ex vivo* adhesive sealing as a means of treating punctured human fetal membranes[46]. An *in vivo* murine model of extrahepatic islet transplantation which utilized an adhesive polymer consisting of a poly(ethylene glycol) (PEG) core whose endgroups were functionalized with catechol was demonstrated under oxidizing conditions to immobilize transplanted islets with minimal acute or chronic inflammatory responses[31].

Current studies on Dopa-containing adhesives were performed on TiO$_2$, from promising studies demonstrating a 1 nN reversible interaction between a single tethered Dopa moiety and a wet TiO$_2$ surface[45] to Dopa-functionalized synthetic acrylamide polymers, demonstrating less than 70 pN of adhesion per catechol on TiO$_2$. Experimentation on TiO$_2$ is commonplace in the study of developing biomimetic adhesives and understanding mussel adhesion mechanisms. This is largely due to titanium’s ubiquity in medical implant devices and the associated
2-20 nm thick TiO$_2$ layer on the titanium surface formed under physiological conditions, which yields a hydroxyl-terminated surface that promotes biocompatibility\cite{47, 48}. Some demonstrations of Dopa-functionalized polyethylene glycol show adhesion to biological surfaces only after periodate oxidation\cite{49}, which is another limitation to its potential use in clinical scenarios.

### 1.1.6 Existing Freshwater Mussel Byssal Proteins

Zebra and quagga mussel proteins have been isolated and identified by previous researchers, spanning $D. \textit{polymorpha}$ foot proteins 1, 2 and 3 (Dpfp1-3)\cite{39, 50} and $D. \textit{bugensis}$ foot proteins 0, 1, 2 and 3 (Dbfp0-3), the latter three of which appear to be homologs of the analogous Dpfp proteins\cite{51, 52}. These quagga mussel (Dbfp) proteins were only able to be isolated post-secretion, and thus detected only after crosslinking via a Dopa-specific stain\cite{51}. This only permitted identification of Dopa-containing byssal proteins. To circumvent this caveat, a KCl injection protocol was developed to secrete byssal material on demand from the mussel foot, from which eight novel zebra mussel byssus proteins (Dpfp5-Dpfp12) were revealed\cite{53} (Figure 1.4). Furthering the ingenuity of the freshwater mussel bonding mechanism and an indication of adhesion beyond the role of Dopa chemistry was the absence of tyrosine hydroxylations in Dpfp proteins 6-12. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) revealed zebra mussel proteins in the 5.8-7 kDa range at the adhesive interface between the mussel bulk plaque and substrate surface that were not present within the plaque itself, indicating their putative role in adhesion\cite{54}. The presence of a 10-20 nm electron-dense adhesive layer (Figure 1.5) at the adhesive interface of freshwater mussel plaque and substrate has been found\cite{55}, but the molecular mechanism of adhesion remains to be elucidated.
Figure 1.4 – Artificially induced (KCl injection) secretion of byssal proteins. (a) A splayed open quagga mussel with the internal anatomical structures present. The foot is highlighted, into which KCl is injected to induce secretion of fresh byssal threads and plaques. (b) Magnification of the mussel foot with labels indicating mussel organ structure and protein secretion. (c) Gel electrophoresis of proteins artificially induced to be secreted from the mussel foot. Labels indicate previously identified Dopa-containing Dbfp proteins in addition to newly identified byssal proteins. Reprinted from Rees et al.[56] with permission from Taylor & Francis (www.tandfonline.com).
Figure 1.5 – The plaque-substrate interface in the zebra mussel. (a) TEM micrograph of an ultrathin section of plaque deposited on an epoxy substratum. Arrow indicates the interface while similarly stained granules are visible within the plaque region above the interface, indicated by the asterisk. (b) The interfacial layer remains post-plaque removal. Adapted from Farsad and Sone[55], with permission from Elsevier.

MALDI-TOF MS was performed on quagga mussel proteins to identify those preferentially located at the adhesive interface, as these proteins are presumed to play a significant role in adhesion. However, although MALDI-TOF data provides accurate information on molecular weight, the technique does not provide sequence information. Thus, quagga mussel byssal material was isolated by means of the aforementioned KCl injection protocol. Novel quagga mussel proteins were identified by gel electrophoresis[56]. Gel bands were clipped and MALDI-TOF spectra were produced for the 6/7 kDa bands, which showed significant peaks in the 6-10 kDa range[56]. Separate MALDI-TOF experiments performed on freshly induced plaques and threads were analyzed individually. The cluster of peaks in the 6.1-6.7 kDa range were significantly more intense in the induced plaque spectra, although peaks in this range were found in both thread and plaque spectra[56], indicating a potential structural or cohesive role of the 6-7 kDa proteins in the thread and plaque. A 6.5 kDa peak also featured prominently in the MALDI-TOF spectra of fresh quagga mussel footprints, indicative of a role at the adhesive interface between the plaque and substrate. In an effort to expand the breadth of quagga mussel proteomic data, a quagga mussel foot transcriptome library was created with next-generation RNA sequencing and liquid chromatography tandem-mass spectrometry (LC-MS/MS)[56]. Putative adhesive roles and functionality of proteins as obtained from these experiments remain to be assessed, in order to determine the specific roles of byssal proteins of interest.
1.2 Rationale and Motivation

Marine and freshwater mussels attach to surfaces underwater by the means of the byssus, a proteinaceous anchor containing threads that extend to adhesive plaques at the tips. The distribution and characterization of proteins in marine mussels is considerably more vast and complete than current work in freshwater mussels[43]. Although the byssus of freshwater and marine mussels share macroscale similarities, freshwater byssal activity functions in the absence of high amounts of Dopa, a rare amino acid that has been shown extensively to play a significant role in mytilid adhesion[12]. The phylogenetic distinction between freshwater and marine mussels further indicates a novel mechanism for adhesion in freshwater mussels that has yet to be determined. However, recent work has reported the identification and sequence analysis of eight novel byssal proteins from induced, freshly-secreted byssal threads with minimal cross-linking in the zebra mussel, and seventeen novel byssal proteins in the quagga mussel[53, 57].

Zebra and quagga mussels have Dopa-containing proteins in both the thread and plaque; thus further analysis and characterization of these proteins will reveal the full potential of adhesive properties in freshwater mussels. However, understanding mussel adhesion has expanded beyond the role of Dopa chemistry[53, 55]. Elucidating the mechanism of wet-resistant adhesion in freshwater mussels, either from sequence features or physical characterization, provides preliminary data regarding mechanisms and conditions in which mussel adhesive proteins work optimally. Furthermore, characterization of byssal protein functions is imperative to understanding their role in adhesion. Studying the interactions of putative adhesive proteins with surface-sensitive and optical techniques is particularly useful because of the complicated nature of protein interaction with material surfaces[58], which aligns with the downstream goal of investigating adhesive protein activity in potential applications of mussel-inspired adhesives. Our goal is to determine adhesive protein candidates in both zebra and freshwater mussels, and to perform biophysical characterization of top candidates to evaluate their activity and structure on different surfaces and adsorption conditions.

1.2.1 Limitations of Current Adhesives

All adhesive events in living organisms occur in the presence of water; however, some adhesions using chemical-based adhesives are easily broken by water[28]. Biological adhesives permit the attachment of tissues with different mechanical properties and make it possible to connect tissues with interface prosthetics. Surgical glues are essential to the success and efficiency of minimally invasive surgeries, which rely on sealing or attaching devices to tissues without tissue penetration or compression[59]. The body is able to heal damage incurred by cuts or breakages if the tissue on either side of the cut is held close enough together to permit regrowth; however, this is not always possible. The typical procedure for fastening soft tissue
involves mechanical fastening methods such as suturing and stapling, or pinning of bone\cite{60}. However, these methods may be inadequate because of (1) turbulent and moist conditions such as cardiovascular, tracheobronchial, or pulmonary surgeries, (2) tissue sensitivity to mechanical fasteners, such as kidneys and other organs with high perfusion rates, (3) minimally invasive surgeries in which access to apply staples and sutures is restricted, AND (4) scarring that may impede the outcomes of cosmetic surgery\cite{60}.

Less invasive procedures are increasingly prevalent due to their reduction of postoperative complications, recovery times, and patient discomfort; however, their effectiveness is partially reliant on clinically approved surgical glues that must meet three basic criteria: be nontoxic, bind strongly to tissue, and work well within wet and highly dynamic environments within the body\cite{59}. In addition, an optimal surgical adhesive will be able to hold two sides of tissue together until there lies enough mechanical strength to support wound healing\cite{14}, be biodegradable, be gradually metabolized without inciting an immunogenic response\cite{60}, be inexpensive, and be easy to use. Dental restoration techniques including synthetic adhesives, coatings, and sealants tend to experience failure, deterioration, and detachment due to weak bonding between dental resin and tooth surfaces\cite{61, 62, 63}, leading to the irreversible loss of tooth tissues\cite{64}. This has underscored the demand for broadened understanding of the biomimetic wet adhesion approach, for which marine and freshwater mussel adhesion mechanisms have become leading models.

Currently existing adhesives are generally classified into two categories: synthetic reactive glues such as cyanoacrylates and bioderived fibrin- and collagen-based glues\cite{14, 60}. Although fibrin glues are non-toxic, they have weak tensile, cohesive, and adhesive strengths\cite{14, 65, 66} and their derivation from autologous tissues presents financial and labour-intensive limitations\cite{14}. The main components of fibrin glues include thrombin and fibrinogen, proteins obtained from the human blood, leading to potential viral infections if improperly screened\cite{67}. Commercial synthetic adhesives based on cyanoacrylate function by polymerizing upon contact with anionic substances to create a film that holds wound edges together\cite{33, 68}. However, neither cyanoacrylate-based adhesives or fibrin sealants are bioabsorbable\cite{69}, and their applications for external use are not recommended because cyanoacrylates incite an intense inflammatory response upon contact with noncutaneous surfaces\cite{70}.

### 1.2.2 Weaknesses of Current Mussel-Inspired Adhesives

Although Dopa-based mussel-inspired adhesives are well-studied, caveats with this model are threefold: (1) isolating mussel adhesive proteins natively from mussels is laborious, inefficient, and impure, with 10,000 individual mussels required to obtain one gram of marine mussel adhesive proteins\cite{71}; (2) conjugations of Dopa onto polymers may present limitations in biocompatibility and variability of use, and (3) even if able to be recombinantly produced and
genetically engineered via *E. coli* to produce mussel proteins en masse purely and efficiently, Dopa has its own disadvantages in breadth and stability of use as an adhesive due to its potential cytotoxicity. The oxidative crosslinking of catechol occurs at high pH levels (pH ≥ 8), which limits regions in the body in which a Dopa-based adhesive may be used, including mildly acidic tissues such as cancer cells (pH ≤ 7)[72] and subcutaneous tissues (pH = 6.7-7.1)[73, 74]. Moreover, catechol oxidation generates reactive oxygen species (ROS) including super oxide anion (O2^-) and hydrogen peroxide (H2O2)[75, 76]. Although ROS presence can have both beneficial and detrimental effects depending on their concentration among tissues, the regulation of ROS generation is an additional consideration in an already sensitive system, the failure of which could lead to chronic inflammation or tumor initiation responses[74, 77].

Dopa readily undergoes a two-electron oxidation to dopaquinone in the presence of alkaline pH or with trace oxidants such as O2. This greatly affects its versatility in clinical conditions as alkaline and oxidative environments are certainly present, and an 80% decrease in the amount of Dopa binding with TiO2 surfaces upon oxidation to dopaquinone has been demonstrated in AFM single molecule studies[48]. At low pH (pH < 5.5), Dopa favors binding via molecular adsorption to TiO2 - i.e. through hydrogen bonding[48]. At high pH (pH ≥ 8) and in the presence of metal ions, Dopa will undergo bidentate coordination bonding with surface-bound, available Ti sites[48]. The binding strength of a Dopa-TiO2 coordination bond far outweighs that of a Dopa-Ti hydrogen bond (44 kT vs. 4 kT), resulting in an increase of binding strength of a single Dopa group to the TiO2 surface[48, 78]. However, high pH also imparts a competing effect on Dopa activity by oxidation of the catechol group, which decreases the number of Dopa residues available for binding. Although marine mussels do regulate pH in the distal depression[79, 80] and secrete mfp-6, which has been shown to demonstrate antioxidant properties, it is abundantly clear that interactions involving Dopa are highly environmentally-sensitive, and although condition parameters can be tuned precisely to fit the requirements for optimal Dopa adhesive activity, this may not always be possible in a surgical or clinical setting.

1.2.3 Current Progress: Candidate Spatial Verification and Molecular Weight

MALDI-TOF performed on induced plaque and threads indicates that a family of proteins at approximately 6.6 kDa are intensified in the induced plaque, while a family of 8.1 kDa and 10-14 kDa proteins appear to be plaque-specific, implying that there is a selection of proteins in these molecular weight ranges which play a role of elevated importance in the plaque compared to the thread. The presence of a selection of plaque-specific and plaque-intensified proteins thus gives direction regarding target molecular weights of proteins with an enhanced role within the plaque as compared to the thread. Putative roles include a structural component exclusive to the plaque or an adhesive component with a partial or preliminary role in preparing the plaque for attachment. Another consideration borne as a consequence of the experimental design of
isolating induced threads from plaques (which are inherently connected together at the time of production) is that some thread components will be incorporated into the plaque samples - and these will also appear in the induced thread samples - while thread components that do not also appear in the induced plaque samples are truly unique to the thread samples. Thus, we can believe that peaks only appearing in the induced plaque MALDI-TOF spectra are truly unique to the plaque. Conversely, for peaks which appear in both induced plaque and induced thread spectra, we look to the relative ionization counts to infer the abundance of that molecular weight family within that sample, and face these peaks with more skepticism as they could be carried over from the alternate sample.

MALDI-TOF was compared among samples of induced plaque and three forms of naturally secreted proteins directly at the adhesive interface: the upturned plaque, and both fresh and aged plaque footprints left behind on substrates. Most significant of these results was that a 8.1 kDa protein that was found to be unique to the induced plaque (when compared to the induced thread), was also found to not only be consistently present in both the upturned plaque and plaque footprints, but also had a relative increase in ionization count compared to their presence in induced plaque. This implies that a protein at approximately 8.1 kDa has a characteristic that enhances the activity of naturally secreted plaques or is enhanced by its localization, and is localized particularly to the adhesive interface formed at the base of the plaque, providing further evidence of protein activity in this 6-14 kDa molecular weight family.

1.3 Project Objectives and Goals

1.3.1 Objective 1: Identify quagga mussel adhesive protein candidates and select the top candidate

The proteins that are responsible for the adhesive strength at the quagga mussel plaque-substrate interface will be identified and spatially verified. The aim is to identify proteins that are isolated to the adhesive interface of zebra and quagga mussels, and to deduce purported functions and primary sequences, which may be reliable indicators of adhesive behaviour. Proteins isolated to the adhesive interface may contain motifs of interest that are indicative of adhesive characteristics when examined for protein homology and conserved domains. Determining the relative abundance of proteins at the adhesive interface versus the bulk plaque is another useful proxy to determine the potential adhesive role of a novel byssal protein.
1.3.2 Objective 2: Purify and characterize the top quagga mussel adhesive protein candidate

Purification of a top quagga mussel adhesive candidate will be performed. The top protein candidate will be purified and characterized with adsorption, surface-sensitive, optical analysis techniques. From an adhesive perspective, preliminary data from adsorption studies can reveal if the top protein candidate is able to form interactions with various substrates.

Figure 1.6 – Summary of approach from proteome to candidate. Full pipeline for determining potential quagga mussel adhesive protein candidates from artificially induced and naturally secreted mussel samples.
1.4 Document Overview

This thesis consists of four chapters. Chapter 1 introduces the overall background on zebra and quagga mussels, our research motivation and goals, and objectives of this project. Chapter 2 and Chapter 3 represent manuscripts for scientific journal submission. Objective 1, as described in section 1.3.1, is addressed in Chapter 2: “Quagga Mussel Adhesive Protein Identification”. Objective 2 as described in section 1.3.2 is addressed in Chapter 3: “Dbfp7 Purification and Characterization”. Chapter 4 summarizes the results and discussion from the previous chapters and addresses future work that may be undertaken to further extend our understanding of quagga mussel adhesion.
Chapter 2

Quagga Mussel Adhesive Protein Identification

Judith Ng, David Rees, Kate Li, Eli Sone

Chapter 2 is in preparation for a manuscript to be submitted for peer-reviewed journal publication. All work was performed by Judith Ng with the exception of David Rees, whose contribution was the collection and purification of all protein samples presented in this chapter. Kate Li’s contribution was assistance in analyzing relative quantification proteomics results and determining Dbfp7 peptide enrichment.

2.1 Abstract

The European freshwater mollusk *Dreissena bugensis* (quagga mussel) adheres to a variety of underwater surfaces via the byssus, a proteinaceous anchor, from which threads are secreted. The threads terminate in small, disc-shaped adhesive plaques which have adapted to turbulent conditions with remarkable adhesive strength for their size. We have identified the most likely proteins to be responsible for adhesion in *D. bugensis* by developing a novel protocol to collect naturally secreted (plaque and footprint) and artificially induced (whole thread and plaque) proteins from *D. bugensis*. Protein samples were analyzed for bottom-up sequencing with LC-MS/MS. Relative quantification of naturally secreted proteins between the plaque and footprint samples was performed with tandem mass tag (TMT) mass spectrometry. Dbfp7 was selected on the basis of being the most highly footprint-enriched and abundant protein which manifests as two separate protein fragments matching the target 6-8 kDa molecular weight range predicted by previous MALDI-TOF and SDS-PAGE experiments, and was carried forward as the top quagga mussel adhesive protein candidate.
Chapter 2. Quagga Mussel Adhesive Protein Identification

2.2 Introduction

The quagga mussel (*Dreissena bugensis*) and the closely related zebra mussel (*Dreissena polymorpha*) are invasive freshwater mussel species known for their rapid spread in the waterways of eastern North America due to their ability to adhere to a variety of surface chemistries underwater[2, 3, 55]. This adhesion occurs by the means of the byssus, a proteinaceous anchor composed of threads that terminate in adhesive plaques. Mussel-produced adhesives must maintain fluidity while being secreted from the secretory organ to the extracellular space, and then must undergo self-assembly and irreversible curing to form a durable adhesive joint[38]. 3,4-dihydroxyphenylalanine (Dopa), a rare amino acid and hydroxylated derivative of tyrosine found to be abundant and instrumental to adhesion within marine mussel adhesive proteins, crosslinks matrix proteins within mussel secretions to effectively cure the adhesive[34]. The byssi of both zebra and quagga mussels are composed of proteins containing small amounts of the otherwise rarely-appearing amino acid Dopa, indicating that Dopa plays an integral role in the mussels adhesive properties[51].

The distribution and characterization of proteins in marine mussels is considerably more vast and complete than current work in freshwater mussels[43]. Although the byssus of freshwater and marine mussels share macroscale similarities, freshwater byssal activity displays functionality at a fraction of the high amounts of Dopa shown extensively to play a role in mytilid adhesion[51]. The phylogenetic distinction between freshwater and marine mussels further indicates a novel mechanism for adhesion in freshwater mussels that has yet to be determined.

Recent work has reported the identification and sequence analysis of eight novel byssal proteins from induced, freshly-secreted byssal threads with minimal cross-linking (Figure 2.1b) in the zebra mussel, and seventeen novel byssal proteins in the quagga mussel[53, 57]. Armed with novel information regarding quagga mussel byssal protein identities and sequences, we aimed to determine the proteins primarily responsible for the adhesive characteristics at the quagga mussel plaque-substrate interface. This is accomplished by determining proteins that are isolated to the adhesive interface of naturally set quagga mussels (Figure 2.1b) and to deduce their purported functions and sequences, which may be reliable indicators of adhesive behaviour. Proteins isolated to the adhesive interface may contain motifs of interest that are indicative of adhesive characteristics when examined for protein homology and conserved domains. Determining the relative abundance of proteins at the adhesive interface versus the bulk plaque is a final useful proxy to determine the potential adhesive role of a novel byssal protein.
In the determination of a protein candidate with a potential adhesive role, attention was focused on proteins localized to the adhesive interface. It is expected that a top adhesive protein candidate with a role in protein-substrate activity would be located at the junction between the mussel footprint and substrate, which includes two components - the plaque and footprint. MALDI-TOF, LC-MS/MS, and SDS-PAGE were conducted on a selection of artificially induced and naturally secreted quagga mussel proteins localized to regions at the adhesive interface between the mussel footprint and substrate. Together, this information provided guidelines for further purification procedures that would optimize candidate protein recovery, including clues to an organ from which the potential adhesive candidate would be purified from (i.e. the phenol gland or the entire foot) and the approximate molecular weight of the protein of interest, particularly when compared to miscellaneous and unrelated byssal proteins. This informed relevant downstream protein purification procedures such as protein separation by molecular weight, isoelectric point, charge, or hydrophobicity, among other parameters.

MALDI-TOF was performed on three separate experimental setups to investigate the molecular weight of protein(s) with the highest relative intensity among the thread, plaque, and footprint components of the quagga mussel. Experiments revealed a consistent peak within the 6-8 kDa range of proteins among the induced plaque, upturned naturally secreted plaque, and footprint samples, indicating an abundance of proteins between 6-8 kDa concentrated at the adhesive interface[56]. MALDI-TOF performed on excised gel bands representing a doublet of 6-8 kDa proteins isolated from artificially induced threads and plaques run together on SDS-PAGE revealed high relative intensity of proteins in this same 6-8 kDa range.

With knowledge of the target protein molecular weight, proteins were isolated directly from the adhesive interface. To provide a comparison with proteins found in the plaque, both samples from the adhesive interface (further indicated as “footprint” samples) and plaque components were collected. This is the
first time proteins from the quagga mussel adhesive interface and bulk plaque were isolated, sequenced, and quantified. Based on the distribution of proteins in the naturally secreted plaque and footprint, identification of proteins responsible for adhesion based on localization, molecular weight, novelty as a byssal protein, and predicted role based on homology to currently characterized proteins was performed.

2.3 Methods

2.3.1 Mussel Thread/Plaque and Footprint Collection

Quagga mussels were collected from the origin of the St. Lawrence River at Kingston, Ontario. Mussels were kept in a freshwater aquarium with closed circulation (12°C) and were fed powdered green algae. Mussel cages were collected by attaching Teflon strips to border glass slides. Mussels were placed into these cages and allowed to sit for 12-24 hours. Slides were removed from the aquarium and investigated for mussel adhesion. Mussels were gently detached from the slides, and mussel footprints (residue left behind after mussel plaque attachment) were collected and submerged in extraction buffer. Mussel plaques extending from the mussel foot were detached and submerged in extraction buffer (0.2M sodium borate, 4M urea, 1mM KCN, 1mM EDTA, 10mM ascorbic acid)[51]. Samples were homogenized, centrifuged, frozen, and stored for future analysis. The plaque and footprint samples were homogenized for two minutes in a 1 mL ground glass hand-held tissue grinder containing extraction buffer kept on ice. Plaques from 25 mussels were collected, sonicated on ice in two-second intervals, homogenized, centrifuged at 17,000xg for eight minutes at 4°C, flash-frozen in liquid nitrogen, and stored at -20°C for later use. 400 footprints from mussels were homogenized, desalted, concentrated, flash-frozen and stored at -20°C for later use.

2.3.2 SDS-PAGE and Staining

Samples were separated using Life Technologies Bolt 12% Bis-Tris precast 10-well gels, using Bolt MES-SDS running buffer, with lanes loaded with 20 µl of Bolt LDS sample buffer and 20 µl of protein solution. Proteins were separated using a Bolt Mini Gel Tank, with a constant 165V voltage setting, for 35 - 40 min. The gels were silver stained using the Life Technologies SilverQuest Staining Kit. Briefly, gels were fixed for 20 min using ethanol/acidic acid/water (40:10:50), washed for 10 min in ethanol/water (30:70), and then sensitized for 10 min in ethanol/sensitizer/water (30:10:60). After multiple washes, the gel was stained for 15 min with the Kit Stainer/water (1:99), washed, and developed for 8-10 min. Gels were photographed immediately following development termination. The concentrated thread and plaque solutions were separated with SDS-PAGE, and silver-stained.

2.3.3 Liquid Chromatography Tandem Mass Spectroscopy (LC-MS/MS)

LC-MS/MS of extracted peptides was performed at the SickKids Proteomics, Analytics, Robotics & Chemical Biology Centre (SPARC BioCentre) at the Peter Gilgan Centre for Research Learning at the Hospital for Sick Children. The digested peptides were loaded onto a 100 µm ID pre-column (Dionex) at 4µl/min and separated over a 50 µm ID analytical column (C18 2 µm, Dionex). The peptides were eluted
using a 0 to 35% acetonitrile gradient with an EASY-nLC 1000 nano-chromatography pump (Thermo Fisher, Odense, Denmark). Data was acquired at 70,000 FWHM resolution in the MS mode and 17,500 FWHM in the MS/MS mode. A total of 10 MS/MS scans were obtained per MS cycle.

2.3.4 Protein Identification, Selection, and Analysis

Using the proteomics software PEAKS8 (Bioinformatics Solutions Inc., Waterloo, ON), de novo sequences derived from MS/MS data were matched against the quagga mussel foot transcriptome with tyrosine hydroxylation to Dopa set as a variable modification. Parent ion and fragment ion mass tolerances were set to 5 PPM and 0.01 Da respectively and hits were manually confirmed by inspecting the spectra. In PEAKS, the identification probabilities of the protein and peptide matches are indicated by the formula -10LogP. Our acceptance criteria for identifying novel proteins in PEAKS8 are peptide LogP $\geq$ 15, Protein -10LogP $\geq$ 50, and de novo average local confidence (ALC) score $\geq$ 80%, with at least two peptide spectra identification, unless otherwise stated. Protein percent coverage is the amount of the protein identified by supporting spectra. Signal peptides on proteins were identified using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/). Proteins with signal peptides were examined for protein homology using NCBI Protein BLAST (Basic Local Alignment Search Tool) using no adjustment method to compensate for amino acid composition and using filtering low complexity regions[82]. Sequences were searched against the NCBI non-redundant (nr) database, which consists of sequences from Swiss-Prot, Protein Information Resource (PIR), Protein Data Bank (PDB), and NCBI protein reference sequence (RefSeq) databases.

To identify proteins with potential characterization as adhesive proteins in the context of mussel-to-substrate adhesion, proteins were analyzed through BLAST-P to evaluate each protein’s sequence similarity to protein sequences of other biological organisms[82]. To score the significance of the match, the associated confidence level, Expect value (E), is assigned by BLAST to the predicted homology. An E-value is given to each match between a protein query submitted by the user and an existing protein sequence within the BLAST database. The lower the E-value, the closer the query sequence matches the database protein sequence as assigned by BLAST.

2.4 Results & Discussion

There currently exists extensive proteomic data on quagga mussel proteins secreted both via artificial induction and natural secretion. There also exists novel data on quagga mussel byssal protein identity and composition. However, there is a knowledge gap of the distribution of all proteins (byssal or otherwise) among the footprint and plaque. Relative quantification of naturally secreted proteins reveals the enrichment levels of proteins at the footprint and plaque, providing information beyond the identities of proteins found to be artificially induced and naturally secreted. This technique and investigation is instrumental to the discovery of a novel adhesive protein as such a protein is predicted to be enriched at the adhesive interface. However, other parameters such as absolute abundance of putative protein candidates, their predicted molecular weight range, and detection with LC-MS/MS in parallel experiments are taken into consideration.
2.4.1 Quagga Mussel Protein Distribution

Further processing of 618 quagga mussel proteins found in the footprint, plaque, and induced protein sources includes thorough analysis of unique and shared protein hits among each source, whose presence in each sector would, in addition to their predicted role through BLAST, offer clues to their potential role as members of a unique or ubiquitous source. Using a custom R script, all proteins were separated into seven separate sectors depending on their identity as being solely plaque proteins, solely footprint proteins, solely induced proteins, or proteins shared between one or more of the other sets.

![Figure 2.2](image)

**Figure 2.2** – Area-proportional Venn diagram in the center, representing 385 unique proteins sourced via artificial induction and natural secretion. Numbers shown in each sector represent the number of proteins localized uniquely to the respective sector. The induced family of proteins is tinted yellow while the footprint and plaque proteins, which together compose the naturally secreted family of proteins, are shown in blue. Breakdown of the ten most ubiquitous protein categories is indicated in separate pie charts. Pie charts referring to proteins represented in more than one sector are highlighted for emphasis. Colour scheme follows identical labelling as Figure 2.3.

Naturally Secreted Proteins

Among naturally secreted proteins, there are a selection unique to the adhesive interface/footprint and to the plaque. As a consequence of the experimental design wherein footprint and plaque proteins may contaminate the other’s sample during the collection process, an overlap (dark blue sector in Figure 2.2, N = 41) between footprint and plaque proteins is expected. Likewise, proteins found in both the footprint and plaque sectors are also present when artificially induced. These 61 proteins may be candidates...
already existing in their final cleaved form when present in storage within the mussel secretory granules; thus, they are the same forms of the proteins which are secreted by the mussel when required naturally for adhesion.

Artificially Induced Proteins

Proteins found solely in the induced sector (yellow sector in Figure 2.2, N = 102) constitute the most populous category, at almost double the size of the next highest sector. These proteins were artificially induced to be secreted directly from the mussel foot by the means of a KCl injection into the ventral groove of the mussel foot organ. Proteins released from the mussel were collected from the surface of the foot as they were secreted. This opens three possibilities for proteins found exclusively in the induced samples. Firstly, they contain premature and/or mature thread proteins as they are proteins which are eventually secreted by the mussel to be crosslinked as thread and plaque components. Secondly, these proteins may be components produced in response to KCl depolarization. Saline injections into the foot depolarize the mussel pedal ganglion, turning on all foot processes[32]. Unlike natural thread formation which is a well-coordinated on-off process, saline injections produce a sustained “on” process until ion balance is restored[32]. Thus, this unusual and lingering response triggered by KCl depolarization may trigger the secretion of proteins that would not ordinarily be present in natural mussel secretory conditions. Thirdly, collection of mussel proteins directly from the surface of the mussel foot after artificial induction introduces the possibility of inadvertent collection of proteins that may be naturally present on the surface of the mussel foot, such as mucus-like proteins already existing on the foot surface or proteins used by the mussel to prime the surface of the substrate prior to interaction.

2.4.2 Determination of Putative Protein Roles

Homology of quagga mussel proteins was estimated with BLAST with results separated into two streams based on E-value. BLAST results with an E-value $< 10^{-6}$ were considered to be confident, reliable matches of homology and therefore, likely function. BLAST results with an E-value $> 10^{-6}$ were considered unreliable matches - therefore, the function of the protein could not be positively inferred on the basis of sequence homology. Each of the 385 proteins found within more than one sector was further sorted on the basis of not meeting or exceeding the E-value threshold, and by the identity or absence of the match as indicated by BLAST, following the workflow as shown in Figure 2.3.
Figure 2.3 – 385 proteins sourced via artificial induction and natural secretion were sorted into a number of categories, sequentially narrowed by predicted or known roles based on confidence to existing proteins via BLAST results.

**E-value < 10^-6 (Reliable) Matches**

Proteins with reliable BLAST matches and strong homology to proteins within the NCBI nr database were separated into three categories: extracellular, cellular, and hypothetical. Extracellular proteins include those with functions outside the cell, including but not limited to cytokines, extracellular proteinases, enzymes, and proteins within the extracellular matrix (ECM) such as collagen. Proteins with a basic cellular role (tubulin, myosin, actin, etc.) were grouped as cellular proteins. Confident hypothetical proteins in this category are proteins whose top BLAST result is homologous to a protein of unknown function. This category of proteins with strong homology marked in black in Figure 2.3 were carried forward as potential byssal protein candidates (Dbfp proteins). It is likely that these proteins indeed have homology to hypothetical proteins which are unrelated to mussel adhesive activity, or mussel activity at all. It is also possible that these proteins were considered hypothetical because they are indeed novel in the sense that they had not been previously characterized in the context of being fingerprinted against a full mRNA database of quagga mussel foot proteins.

**E-value > 10^-6 (Unreliable) Matches**

BLAST results in this category represent proteins which do not have existing confident matches to proteins in the database. BLAST was able to identify protein queries directly to a match within the NCBI nr database with the higher E-value indicating uncertainty of the match. As a consequence of being a novel, uncharacterized, or unknown protein, these results return without homology to any protein, or return as an unconfident match. These proteins are sorted into the “uncertain match” category with proteins that do not have strong homology in Figure 2.3. These proteins may be truly novel while also being byssal components (D. bugensis foot proteins - Dbfp’s), may be novel and unrelated to mussel byssal activity, or may be proteins that have a well-characterized match in the NCBI nr database but
due to short sequence length or an incomplete database for that particular sequence, resulted in a match with a high E-value. Thus, these proteins are returned to the pool of potential byssal proteins. Their identification as byssal proteins is accomplished by first searching for matching protein accession numbers between the query sequences and existing known Dbfp accession numbers. As a catchall method, the query sequences within this pool were aligned to known Dbfp proteins as established by Rees et al. (manuscript in progress) with COBALT. There is a second category of proteins within this “unreliable” sector that includes proteins which return the same “hypothetical” result as in the E-value < 10^{-6} group; however, even the hypothetical nature of the match is uncertain as they are not confident matches. Lastly, there are BLAST results denoted as NA (non-annotated), which refer to proteins without homologs - even those which return a result of “hypothetical”, “non-characterized”, or “unknown” - as it is unclear if they encode actual proteins. Submitting these protein sequences into BLAST returns no result.

Putative Roles of 385 Quagga Mussel Proteins

All proteins found within Figure 2.2 were searched through BLAST for their top homolog. The majority of these matches had confident and significant BLAST results and were thus assigned to a distinctly labeled category (e.g. cellular protein, mucus component, structural protein, etc). Proteins with unreliable matches (E > 10^{-6}) were categorized together as “unknown” proteins. Not all proteins separated into their associated categories were given a match description that was precisely annotated as an existing protein. For example, some matches returned prefaced as “predicted” proteins (e.g. “PREDICTED: kielin/chordin-like protein isoform X3 [Lingula anatina]”). In the aforementioned example, despite their annotation as a predicted protein, the match was sorted into the “kielin/chordin-like” category of proteins. Similar generalizations were applied towards proteins that were described as “galaxin” or “galaxin-like”. All matches that were established in these categories were confident (below the E < 10^{-6} threshold) - thus, for the purposes of streamlining protein matches in text and figures, “-like” and “predicted” proteins are hereon pooled together. However, it is prudent to acknowledge that this spread of homologs and predicted proteins as classified within the BLAST database is a demonstration of the ambiguity and variation both within the BLAST database and the BLAST Blocks Substitution Matrix (BLOSUM) used for scoring identifying and scoring protein sequence alignments, and the peptides and proteins which compose the query matches. An analysis of major categories of the most ubiquitous proteins found is present in Appendix 2.A.

2.4.3 Byssal Proteins: All Sectors

Among 71 proteins identified as Dbfp’s from 385 unique proteins, each sector from Figure 2.2 with the exception of the footprint-only sector was represented, likely owing to the ubiquity of 18 different Dbfp transcripts appearing uniquely in sources depending on their putative role. The absence of Dbfp’s existing solely within the footprint sector is unusual for two reasons: (1) that with a large selection of Dbfp’s isolated from naturally secreted and induced proteins, at least one byssal protein would be expected to be represented within each sector; and (2) that the footprint sector is an expected sector for potential adhesive proteins, some of which would intuitively be found specifically at the adhesive interface. However, it is not a completely unfeasible result as protein collection itself - isolating footprint proteins separately from plaque proteins - is not a seamless method. As a consequence of the adhesive
nature of the interactions between footprint and plaque, it would be expected that a selection of footprint proteins remains intertwined within the plaque sample, and vice versa - thus resulting in the collection of plaque proteins even with the intention of isolating a pure footprint sample. The very nature of the proteins, one or both of which participate in an adhesive environment and thus might adhere to each other, combined with the physical collection method itself, ensure that an exceedingly sensitive method of sequencing such as LC-MS/MS will make it difficult to reveal pure protein components in footprint-only or plaque-only sectors.

**Figure 2.4** – Dbfp proteins separated by representation across footprint, plaque, and induced protein samples. Values in parentheses represent total peptide counts for each Dbfp within each respective sector. Footprint and plaque Dbfp breakdowns shown outside of the venn diagram indicate total Dbfp counts within entire footprint and entire plaque sectors. Repeated Dbfp’s appearing within different sectors represent transcripts of Dbfp homologs unique to each sector (e.g. Dbfpα, Dbfpβ, etc.)

Arranging Dbfp peptide count within each sector by decreasing order in Figure 2.4 amplifies the significance of more abundant Dbfp’s, particularly in sectors with high representation of Dbfp’s. Abundance within each sector is denoted by total peptide count of each Dbfp, shown in parentheses. Due to the large number of Dbfp hits among the footprint and plaque sectors, total peptide counts for entire footprint and plaque sectors are also shown in Figure 2.4. It is unusual that there are no Dbfp’s present within the footprint-only sector as one would expect that byssal proteins potentially responsible for adhesion would be concentrated at the adhesive interface. However, the high total Dbfp peptide count represented in footprint proteins overall (N = 883) indicates that many byssal proteins are indeed involved in footprint function. Their high presence within the plaque and induced sectors indicates the following possibilities: (1) that no single footprint byssal protein interacts alone within the adhesive interface and/or (2) that the protocol by which footprint proteins are isolated prevents the collection of pure byssal proteins within the footprint. The latter option is likely a consequence of footprint byssal proteins’ tight struc-
tural integration with plaque proteins when naturally secreted, preventing collection of footprint byssal proteins alone. That is, footprint byssal proteins which truly have a footprint-specific role may simply not be identified by the protocol used to collect the proteins, and instead are found in a shared sector as they are removed while the plaque is scraped off.

For these reasons, it is most logical to consider proteins within the overlapping sector of the footprint and plaque. This is the most thorough way to ensure that true footprint-only proteins - Dbfp’s or not - are considered. Similarly, plaque proteins with a currently unknown, but likely significant role in adhesion due to their proximity to the interface and their likely inadvertent collection by the razor blade method - and thus no longer appearing in the plaque-only sector, but rather within a sector that includes an overlap between footprint and plaque - points to an overlap between sectors as a fair and thorough starting point for potential adhesive proteins. Further justification for focused analysis on the footprint+plaque overlap is explained in Appendix 2.C.

2.4.4 Relative Protein Quantification of Footprint+Plaque Proteins

Represented in Figure 2.5a are all 155 proteins as individual points ranked by fold change ratio of relative abundance at the adhesive, where abundance is determined by the spectral count supporting presence of a certain peptide fingerprinted to the full protein transcript. Proteins with a fold change ratio greater than 1 are more abundant at the footprint - 48% of proteins fit this criteria. A value of 1 indicates equal abundance at the footprint and plaque (0%), while proteins with a fold change ratio below 1 are more abundant at the plaque (52%). Of the top 30 proteins (Table A1) among the 155, two are Dbfp proteins (Dbfp18 and Dbfp7 respectively). Eight proteins within these top 30 were also found among the 233 proteins present in the artificially induced samples, indicated by asterisks in Figure 2.5b. Two of these proteins were Dbfp hits while six were sorted as structural and unknown proteins. The vast majority (22/30, 73%) of the top 30 proteins were only naturally secreted products. An analysis of major categories of the most ubiquitous proteins found within this footprint and plaque quantification is present in Appendix 2.B.
Chapter 2. Quagga Mussel Adhesive Protein Identification

Figure 2.5 – (a) 155 byssal proteins ranked in order of decreasing relative abundance at the adhesive interface compared to the bulk plaque. Values are displayed as fold change ratio. Black arrows indicate Dbfp proteins; labels indicate the specific Dbfp. (b) Top 30 proteins abundant at the adhesive interface relative to the bulk plaque, arranged in decreasing relative abundance at the adhesive interface. Proteins are grouped by families based on similar characteristics as determined by BLAST. Proteins with an asterisk (*) were also found among the list of 233 induced proteins in a separate LC-MS/MS experiment. Colour scheme follows identical labelling as Figures 2.2 and 2.3.

The presence of naturally secreted proteins which were also found in the artificially induced samples (asterisks in Figure 2.5b) indicates a number of possibilities: (1) these proteins have very similar forms and roles both when naturally secreted by the mussel as well as when held within secretory granules; (2) artificially induced protein sample collection resulted in the simultaneous collection of proteins which were not related to, and did not originate from the secretory granules, but were collected from the surface of the mussel foot because artificially induced protein samples rise to the surface of the foot upon secretion, thus leading to proteins that were falsely categorized as “induced” when in fact were merely structural, foot surface, or foot priming proteins; and (3) the collection of naturally secreted proteins occurred in a manner that resulted in the simultaneous collection of foot surface or foot priming proteins, which may have been falsely categorized as “naturally secreted” when in fact they were unrelated to and did not originate from the mussel’s natural byssal protein secretion process. Implicated in points (2) and (3) above is that if there is a protein which is in fact related to foot structure or foot priming,
it may appear in the all-intersect or footprint+induced or plaque+induced sector, when in fact it has a role in neither naturally secreted or artificially induced byssal protein contexts.

The converse is also possible - a protein is truly present both in the naturally secreted and artificially induced sectors, but only appears in the induced sector because the crosslinking of proteins during natural secretion inhibits its detection via LC-MS/MS. Lastly, proteins may be truly only found in an artificially induced sample - for example, those secreted directly from the secretory granules within the foot were responding to a substance (ex. KCl) that is not normally present in the mussel’s natural environment, and particularly not in direct contact with the foot. Saline injections into the foot depolarize the mussel pedal ganglion, turning on all foot processes[32]. Unlike natural thread formation which is a well-coordinated on-off process, saline injections produce a sustained “on” process until ion balance is restored[32]. Thus, this unusual and lingering response triggered by KCl depolarization causes proteins to be secreted that would not be present when the mussel is thriving in its natural environment.

Proteins found on the foot (grouped as “foot surface proteins”) with potential roles in foot priming, mucosal secretions, or mussel proteins used to prime the surface for attachment may not necessarily be mutually found within the naturally secreted and artificially induced samples. Foot surface proteins collected among artificially induced proteins are not necessarily consistent with foot surface proteins collected from naturally secreted experiments as a consequence of physical dissection of the mussel and thorough probing of the mussel foot in the former, and natural mussel foot probing in the latter. Thus, within Figure 2.5, it would be possible to find naturally secreted proteins that may not be directly involved in adhesion, but rather have a role related to foot surface preparation/foot priming without the typical clue of being found within the artificially induced family of proteins, and are thus not marked by an asterisk.

2.4.5 Relative Protein Quantification: Candidate Selection

Two previously-identified quagga mussel byssal proteins (Dbfp’s) were present among the top 30 proteins - Dbfp18 and Dbfp7. The presence of Dbfp’s found to be enriched at the adhesive interface is expected, particularly given the high spectral count and representation of Dbfp’s among the induced and naturally secreted proteins shown in Figure 2.4. Dbfp’s are prime candidates for adhesive capability due to their ability to satisfy a number of criteria expected from a novel adhesive protein: Dbfp’s are currently unknown, poorly characterized proteins, are unique to the activity of the quagga mussel, and may be suited to specialize in functions exclusive to the quagga mussel’s unique needs in a traditionally difficult (turbulent, wet, and salinic) environment for adhesive function.

Table A1 displays proteins ranked by decreasing relative enrichment at adhesive interface compared to the bulk plaque. Closer analysis of the proteins present within Table A1 reveals that despite the distinct fold change ratio present, the spectral counts for proteins ranked closer to the halfway point of the list rise dramatically. Some of the higher spectral counts are in fact found in the bottom half the top 30 proteins, revealing that abundance (total spectral count) should be considered in tandem with enrichment at the adhesive interface (fold change ratio) in the determination of a top adhesive candidate.

Ranking proteins and protein clusters with fold change ratio values greater than 1 (i.e. enriched at adhesive interface) in descending order of abundance reveals the ranking shown in Table A2. There are several fold change ratio values in Table A2 which fall below 1 - for example, Rank 12 has a fold change
ratio of 0.72. However, as a cluster of proteins, all accession numbers under comp52765_c1 are included as a group. Most relevant to note is the vast difference in spectral count between the top and bottom halves of proteins sorted by abundance. The only protein ranked highly by enrichment at the interface and by abundance is Dbfp7, having a rank within the top 17% in the former and the top 5% in the latter.

Primary emphasis in the search for an adhesive protein is placed on its abundance at the adhesive interface. However, it is difficult to ascertain the potential significance of a protein candidate on its enrichment at the adhesive interface alone. Abundance is an important metric to consider as it is an indication that the protein’s enrichment is not merely due to a non-bysal component that happens to be collected as a byproduct of surface priming. Assessing spectral count is a useful proxy for absolute abundance at the footprint and plaque as opposed to sole enrichment at the footprint, as a protein candidate with higher absolute abundance may have a more significant role compared to a protein found more relatively enriched at the interface, but lacking overall abundance and presence within collected protein samples. A protein that is both enriched and abundant is more likely to have a distinct role within the naturally secreted environment that extends beyond chance collection as an unintended consequence of procedure.

Although relative abundance is a useful proxy to identify proteins found abundantly at the interface, it is not a foolproof method to classify interfacial proteins as adhesive proteins. Protein presence at the adhesive interface is a confident indicator of its presence as an interfacial protein, but elucidating its role (adhesion, structural, or otherwise) requires further investigation. However, given the extensive evidence of a distinct family of proteins identified with MALDI and the presence of a consistent electron-dense layer remaining on the substrate surface upon detachment of the plaque, the identity of a protein at this interface would be a top candidate for a protein with interfacial-specific role, one of which may be adhesion.

The top 30 most abundant proteins in Table A2 were plotted against their respective fold change ratios, shown in Figure 2.6. There is a weak positive correlation between the two variables, with two large exceptions: the two proteins with the accession number comp53519_c2, and Dbfp18. As Dbfp7 is both the second-most abundant and enriched Dbfp, it is emphasized in Figure 2.6. Other Dbfp proteins present are also highlighted and labeled accordingly. Due to the prevalence of Dbfp representation among this comparison, the molecular weights, fold change ratios, and total spectral counts of each Dbfp match identified are listed in Table 2.1.
In the identification of novel byssal proteins from induced byssal thread and plaque samples, three criteria were used by Rees et al.[81]): (1) the protein contains Dopa, (2) it contains a signal peptide, and (3) the E-value of its top BLAST match is greater than $10^{-6}$, with the implication that the protein is novel in some manner. Although each criteria is important, the presence of Dopa is the most significant as it is not a typical endogenous post-translational modification. Thus, its presence is an affirmative sign that the protein is of byssal origin and is more likely to have a role in adhesion that non-Dopa containing proteins. The presence of a signal peptide is neither a necessary nor sufficient indicator that a protein is truly secreted, is not enough to confidently include or exclude secretion[83, 84], and thus has lower priority in terms of significance. E-value, though a fair determinant of predicted function, demands a binary decision of protein novelty for a non-discrete range of values.

Rees et al. performed an analysis of larger byssal proteins which had not met the criteria to be labelled as a byssal protein[81], of which 5 of 19 proteins previously sorted as non-byssal, non-Dbfp proteins were discovered in the writing of this paper to in fact contain Dopa in either the artificially induced or naturally secreted samples (Table 2.2). However, none of these five Dopa-containing, high molecular...
weight proteins had a top BLAST match above the 10^{-6} cutoff, although the E-values across the top matches for the 5 proteins ranged from 10^{-15} to 10^{-102}. Their inability to pass the signal peptide and E-value thresholds had initially excluded these proteins as potential byssal candidates. However, upon re-evaluation, the presence of Dopa alone was enough of a qualification to classify a candidate as a novel byssal protein, as Dopa does not appear natively or endogenously within other mussel tissues outside of the byssus.

**Table 2.2** – Large and abundant proteins identified by LC-MS/MS by Rees et al.[81] that did not originally meet acceptance criteria to be listed as a novel quagga mussel byssal protein, but were re-classified as potential byssal candidates due to the presence of Dopa within peptides. Bolded transcripts are both highly enriched and abundant at the footprint-plaque interface.

<table>
<thead>
<tr>
<th>Library Transcript</th>
<th>Artificially Induced</th>
<th>Naturally Secreted</th>
<th>Fold Change Ratio</th>
<th>Total Spectrum Count</th>
<th>MW (kDa)</th>
<th>Signal Peptide</th>
<th>BLAST E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>comp53519_c2_seq1-2R</td>
<td>N</td>
<td>Y</td>
<td>5.8</td>
<td>88</td>
<td>124.6</td>
<td>N</td>
<td>10^{-102}</td>
</tr>
<tr>
<td>comp56544_c0_seq1-2R</td>
<td>Y</td>
<td>N</td>
<td>3.4</td>
<td>41</td>
<td>149.2</td>
<td>Y</td>
<td>10^{-76}</td>
</tr>
<tr>
<td>comp56420_c14_seq2-1F</td>
<td>N</td>
<td>Y</td>
<td>2.1</td>
<td>25</td>
<td>209</td>
<td>Y</td>
<td>10^{-15}</td>
</tr>
<tr>
<td>comp48282_c0_seq2-2R</td>
<td>N</td>
<td>Y</td>
<td>0.9</td>
<td>23</td>
<td>68.8</td>
<td>Y</td>
<td>10^{-52}</td>
</tr>
<tr>
<td>comp53523_c0_seq1-1F + comp53523_c1_seq2-0R</td>
<td>N</td>
<td>Y</td>
<td>0.6</td>
<td>278</td>
<td>130.3</td>
<td>Y</td>
<td>10^{-64}</td>
</tr>
</tbody>
</table>

Among proteins identified in Figure 2.6 which present with both moderate fold change ratios and total spectrum counts, five proteins ranked equally or higher on both metrics than Dbfp’s. These include the points labelled in blue, green, and purple in Figure 2.6, which refer to comp53519_c2, comp56544_c0, and comp56420_c14 respectively. Of these five separate proteins, two share the same component number - comp53519 and comp56544 both contain two transcripts each. Transcripts belonging to these three component numbers were discovered to be proteins initially eliminated as byssal protein candidates in high molecular weight LC-MS/MS protein analysis. Although these proteins did not appear to have Dopa when sourced from artificially induced byssal material[81], a further analysis revealed that the PTM analysis of the transcript belonging to comp56544 did indeed detect Dopa in artificially induced sample processing, while four other proteins (comp53519_c2, comp56420_c14, comp48282_c0, comp53523) were found to contain Dopa in naturally secreted byssal protein samples.

It is worthwhile to note three of these larger molecular weight proteins (comp53519_c2, comp56420_c14, comp48282_c0) newly classified as potential byssal proteins presented with high fold change ratios (2.1 to 5.8) in relative quantification of naturally secreted footprint and plaque proteins, indicating their increased relative abundance in the footprint as compared to the bulk plaque, and thus a putative role in adhesive behaviour. These proteins were also among the top 30 footprint-enriched proteins when sorted by abundance (spectral count), ranking 2, 6, 13 in Table A2. These three proteins are identical to those found to rank similarly with Dbfp proteins in abundance and enrichment in Figure 2.6. We suspect that one of these newly byssal-classified large molecular weight proteins found abundantly in the interface may be Dbfp0. Rzepecki & Waite (1993) identified a >200 kDa NBT-positive (i.e. Dopa-containing) protein exclusive to quagga mussels[51]. Due to the paucity of Dbfp0 sequence data, it is not possible to exclusively identify which of these abundant, enriched, Dopa-containing proteins is Dbfp0. However, by molecular weight alone, comp56420_c14 (MW = 209 kDa) may be Dbfp0. This has yet to be confirmed; however, these newly classified Dopa-containing proteins with high enrichment at the adhesive interface may potentially indicate their role in adhesion, but absolutely indicate their role within the byssus as
foot protein which have yet to be fully classified.

**IgGFe-binding Protein (comp53519_c2)**

The abundance of comp53519_c2 in Table A2 is significant as the spectral counts are almost double those of the next highest accession number. These proteins BLAST to IgGFe-binding protein with a significant E-value of $10^{-101}$. IgGFe-binding proteins are related to mucosal activity that the mussel may use to assist in priming the surface of a substrate for adhesion. Its high abundance may be a consequence of mucosal components playing a large role at the adhesive interface with a role in adhesion, or alternatively, due to mussel production of mucosal secretions which are deposited at the footprint, on the layer of the substrate. The high spectral counts, far beyond other proteins found within the plaque and footprint samples, may have appeared so highly abundant because of the footprint collection protocol used. During footprint and interfacial protein collection, a large proportion of the glass slide is thoroughly scraped with a razor blade that spans the width of the slide, far exceeding the surface area actually taken up by each footprint. All contents on the edge of the blade are retrieved and bundled together as a “footprint” sample. It is likely that mucosal components produced by and released from the foot in the mussel’s effort to probe the substrate surface before and during adhesion were simultaneously scraped off and collected in this process.

**Dbfp18**

Proteins enriched at the footprint ranked in descending order of abundance (spectral count) reveals Dbfp18 as the most enriched protein present at the adhesive interface, with a fold change ratio of 10.9x. The molecular weight of Dbfp18 corresponds well with peaks observed in MALDI spectra of induced plaques and threads, which include a 4.3 kDa peak that has a notably higher intensity in the plaque spectra. This 4.3 kDa peak was also observed in spectra produced from upturned plaques and plaque footprints, indicating it may be present at the adhesive interface. However, it likely has a structural role due to its observation in the thread spectra[56]. In addition, its relative intensity (%) is significantly lower (20%) than a family of proteins appearing at a molecular weight range of 6-8 kDa (80%-100%)[56]. Its low spectral count (8) ranks Dbfp18 below three other Dbfp proteins by spectral count. Thus, although the relative quantification indicates that Dbfp18 is an attractive candidate for a protein demonstrating abundance at the adhesive interface, its low abundance relative to other Dbfp proteins and the low relative intensity of MALDI peaks at the adhesive interface by Rees et al.[56] eliminates its candidacy.

**Dbfp8**

The high spectral count of two Dbfp8 matches within the relative quantification experiment in tandem with the fair interface enrichment value (1.6-2x) point to Dbfp8 as a fair adhesive candidate. Summation of the spectral count of each Dbfp8 match reveals a total Dbfp8 spectral count of 92. However, the molecular weight range (1.6 - 2.0 kDa), though narrow, is far higher than the primary MALDI-TOF peaks of the adhesive interface and bulk plaque as performed by Rees et al.[56].
Dbfp6

Presenting with a fold change ratio (1.1x) and spectrum count (N = 7) among the lowest of other Dbfp proteins in Table A2, Dbfp6 also does not rank highly among the 30 other proteins ranked by abundance and relative enrichment. Its molecular weight (17.6 kDa) does not coincide with MALDI-TOF peaks within the induced thread and plaque[56].

Dbfp7

Macroscopically, Dbfp7 appears to be the top protein candidate which plays a strong role at the adhesive interface between the mussel plaque and substrate, both in terms of relative enrichment as well as abundance. Peaks in the 6-8 kDa range have been found prominently in MALDI-TOF spectra of fresh footprints laid down by quagga mussels, which represent the protein composition at the adhesive interface[56]. Notably, a 8.1 kDa peak in the MALDI-TOF footprint spectra was also observed as a plaque-specific protein in MALDI-TOF experiments comparing individual thread and plaque spectra[56]. Past experiments have demonstrated the presence of 6.1-6.7 kDa peaks consistent with Dbfp7 present in both the thread and plaque components of the quagga mussel in MALDI-TOF spectra[56] while also displaying significantly higher intensity in the induced plaque spectra. It is not unreasonable that Dbfp7, as a member of this molecular weight cluster, plays a structural or cohesive role in the attachment of quagga mussel plaques to the substratum. Individual Dbfp7 hits in Figure 2.6 suggest that individual Dbfp7 matches rank consistently below Dbfp8 in spectral count and average to a fold change ratio slightly higher than that of Dbfp8; however, combining total spectral counts for all separate matches to Dbfp8 and Dbfp7 (Table 2.1) reveals that Dbfp7 spectral counts (N = 204) far exceed those of Dbfp8 (N = 92), while matching the target molecular weight range of 6-8 kDa much more closely than Dbfp8. Dbfp7’s presence and high rank in abundance and enrichment within the sector of overlap between footprint and plaque proteins is logical, given that a protein with a strong role in adhesion at the adhesive interface would be found within an overlap between plaque and interface.

Despite SDS-PAGE gels and MALDI of proteins extracted from natural secretion of the quagga mussel footprint prominently featuring a protein family in the 6-8 kDa range, theoretical molecular weights of Dbfp7 sequences are between 9.7 and 14.8 kDa, consistently and notably higher than the molecular weights represented by proteins characterized by SDS-PAGE and MALDI (Table 2.3). LC-MS/MS bottom-up analysis of peptides fingerprinted to Dbfp7 revealed that all six transcripts of Dbfp7 which are represented in the relative quantification experiment feature a consistent bisection of the Dbfp7 transcript into two fragments, cleanly separating one set of Dbfp7 peptides which are consistently enriched in the footprint from those enriched in the plaque. Figure 2.7 highlights this observation. Fragments situated prior to the bisection were named the N-terminal fragment; those situated after were named the C-terminal fragment.
Figure 2.7 – LC-MS/MS peptide analysis shows variable localization of Dbfp7 peptide enrichment. Six transcripts of Dbfp7 present in the relative quantification experiment are represented by their sequence numbers (seq). All peptides appearing before the red line represent peptides enriched in the footprint; those appearing after represent peptides enriched in the plaque.

This observation was extrapolated to all other fingerprinted Dbfp7 sequences. Theoretical molecular weights of the N-terminal fragments (5.4-8.0 kDa) match the molecular weight range observed with SDS-PAGE and MALDI observations of naturally secreted footprint quagga mussel proteins. Molecular weights of full fragments in addition to the separate N- and C-terminal fragments are shown in Table 2.3. The larger of the fragments - in each case, the N-terminal fragment - is enriched at the footprint and adhesive interface, which coincides with the molecular weight range observed in SDS-PAGE and MALDI-TOF data. We hypothesize that the presence of these fragments is the result of post-translational processing. Although the full transcripts are translated and secreted by the mussel, a form of in vivo cleavage or other post-translational process occurs to permit consistent preferential selection and localization of Dbfp7 fragments separately to the footprint and plaque.
Table 2.3 – Theoretical molecular weights of full Dbfp7 sequences and N- and C-terminal fragments found in naturally secreted footprint and plaque samples.

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Full Length (kDa)</th>
<th>N-Fragment (kDa)</th>
<th>C-Fragment (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dbfp7 seq1</td>
<td>14.9</td>
<td>7.8</td>
<td>4.9</td>
</tr>
<tr>
<td>Dbfp7 seq3</td>
<td>9.7</td>
<td>5.8</td>
<td>2.6</td>
</tr>
<tr>
<td>Dbfp7 seq5</td>
<td>14.2</td>
<td>7.2</td>
<td>5.4</td>
</tr>
<tr>
<td>Dbfp7 seq6</td>
<td>12.0</td>
<td>8.0</td>
<td>2.6</td>
</tr>
<tr>
<td>Dbfp7 seq7</td>
<td>12.6</td>
<td>5.8</td>
<td>4.9</td>
</tr>
<tr>
<td>Dbfp7 seq8</td>
<td>12.3</td>
<td>5.4</td>
<td>4.9</td>
</tr>
<tr>
<td>Dbfp7 seq9</td>
<td>13.2</td>
<td>6.3</td>
<td>4.9</td>
</tr>
<tr>
<td>Dbfp7 seq10</td>
<td>12.7</td>
<td>6.0</td>
<td>4.9</td>
</tr>
<tr>
<td>Dbfp7 seq11</td>
<td>9.4</td>
<td>5.5</td>
<td>2.6</td>
</tr>
<tr>
<td>Dbfp7 seq13</td>
<td>10.3</td>
<td>6.3</td>
<td>2.6</td>
</tr>
<tr>
<td>Dbfp7 seq14</td>
<td>9.9</td>
<td>5.9</td>
<td>2.6</td>
</tr>
<tr>
<td>Dbfp7 seq15</td>
<td>11.1</td>
<td>7.1</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Bottom-up analysis of protein tryptic digests as performed through LC-MS/MS does not permit the identification of the original sequence fragments from cleavage products, while the relative quantification experiments reveal novel information not only regarding Dbfp7’s enrichment but also of post-translational processing which consistently influences 6-8 kDa fragments of Dbfp7 to be enriched at the footprint. The absence of the C-terminal fragments (2.6-5.4 kDa) as observed with SDS-PAGE and MALDI implies that there is preferential presence of the N-terminal fragment in the natural secretion of Dbfp7 at the adhesive interface. Thus, utilizing relative quantification as a metric for the identification of potential adhesive proteins has been fruitful in identifying not only a likely candidate for adhesive activity, but also in uncovering an unforeseeable nuance within post-translational processing or cleavage of this protein, which with further analysis and investigation may reveal other expansive roles of Dbfp7.

2.4.6 Caveats of Experimental Design

Protein Collection

Protein extraction from mussels is exceptionally difficult on account of a number of factors that affect their quality in purity (plaque vs footprint protein) and quantity collected including mussel activity, fragility of mussel attachment, footprint and plaque proteins being closely ingrained and possibly nestled within each other, size of mussel footprint, and the extraction method. These factors have contributed to the difficulty in obtaining protein natively from mussels and underscores the importance of developing a recombinantly-driven expression method. However, the laborious experimental design used herein was imperative in the identification of naturally secreted mussel proteins that are specifically natively localized to the adhesive interface.

Ultimately, the limiting factor of extractions is most often the mussels themselves. Mussels must be active to produce fresh threads and plaques. Activity is not easily gauged. It has been found from experience
that younger and smaller mussels are more prone to produce fresh byssal material on a day-to-day basis, which includes manual detachment from the substrate and placement ventral side-down for 12 hours. Further processing includes successfully collecting each plaque by ensuring it is not pre-emptively peeled off or detached during mussel handling. Due to the fragility of these 12-hour old threads and plaques, the entire mussel may be inadvertently detached and the plaques removed, as the mussel needs to be lightly rocked on the substrate to check for adhesion.

Once attached, the plaques are shaved off the substrate with a razor blade in order to reveal the footprint, revealing a very thin layer of protein. The absolute amounts of protein left behind are not only sparse but also not entirely pure in their intended source - for example, we expect that some footprint protein samples are scraped with the plaque and that due to the intrinsically adhesive nature of some footprint proteins, that some plaque proteins remain ingrained among the footprint. This presents an issue for sample purity and may result in an underestimation of purely footprint or purely plaque proteins. This underscores the importance of careful analysis of proteins within overlapping sectors of the protein localization (Figure 2.2), as these proteins may in fact be sourced to only the footprint or plaque, but appear to be shared. Conversely, it is important to note that an adhesive protein may truly have a role in both the footprint and plaque, as this may strengthen its intended function and may therefore be found in both footprint and plaque sectors.

Abundance as a Proxy for Adhesive Character

Spectral count of top candidate peptides as a benchmark for identifying potential adhesive proteins was the primary method used in this work that helped establish Dbfp7 as a leading adhesive candidate. This was based on the premise that an adhesive protein and/or its fragments would most likely be (1) enriched at the interface, (2) abundant at the interface, (3) in the target 6-8 kDa molecular weight range, and (4) contain Dopa. Abundance at the interface is a unique proxy for identifying potential adhesive proteins that, although useful for identifying proteins presenting with a high spectral count at the interface, is not necessarily indicative of a protein that is exclusively or primarily involved in adhesion. Although this is a caveat of our quantitative protein experiments, it does not preclude the utility and sensitivity of performing relative quantification of proteins found among multiple samples. However, the weaknesses of using protein spectral count and abundance cannot be overlooked as they influence the interpretation of adhesive candidates identified by this method.

Selecting a top adhesive protein candidate by its spectral count at the plaque-substrate interface assumes that the protein must be present among the highest of other proteins in the interfacial vicinity. However, the molecular mechanism of adhesion, exact distribution of a potentially adhesive protein, and the importance of enrichment of the protein at the interface are still unknown factors. It is possible that protein(s) involved in adhesion may not necessitate abundance at the interface. For example, a monolayer may be sufficient for adhesion, either to reduce the metabolic load required by the mussel or to provide the minimum amount of adhesive required to streamline plaque-substrate contact. Additionally, the premise of this relative quantitative experiment centers on LC-MS/MS, a technique which identifies proteins by their ability to be trypsinized, ionized, and deflected. Analyte ionization is influenced by analyte electronegativity, proton affinity, and ionization energy, making it difficult to ascertain that a truly abundant protein will, too, manifest with a high spectral count via LC-MS/MS. Thus, a protein’s lack of high spectral count in relative quantification results does not necessitate its absence at the level of
sample collection. Conversely, the presence of a high spectral count is a positive indicator of the protein’s abundance. Thus, the spectral counts and abundance levels of proteins identified via LC-MS/MS are true positives; however, low or absent spectral counts are not necessarily indicative of protein absence as a consequence of experimental procedure, thus resulting in false negatives manifesting as absent protein candidates.

2.5 Conclusion

We have developed a novel procedure to isolate naturally secreted footprint and plaque proteins as well as artificially induced and minimally crosslinked whole thread and plaque proteins from the quagga mussel. We report for the first time the identities, putative roles via BLAST searches, and localization of naturally secreted plaque, footprint, and artificially induced quagga mussel protein samples as identified with bottom up proteomic sequencing via LC-MS/MS techniques. For the first time, we report novel data demonstrating naturally secreted proteins found to be enriched at the footprint (adhesive interface) relative to the bulk plaque. These conclusions are a culmination of an investigation into quagga mussel adhesive protein candidature, bringing together the database of quagga mussel foot protein sequences together with bottom-up sequencing data of footprint, plaque, and thread proteins produced natively by the mussel. Selection of recently identified *D. bugensis* foot proteins within the naturally secreted protein samples in tandem with existing data of molecular weight characterization (MALDI-TOF, SDS-PAGE) indicating the significance of a protein family in the 6-8 kDa range, have been instrumental in the determination of the most likely protein candidate responsible for adhesion. This candidate, Dbfp7, is simultaneously the only protein candidate which shows the presence of Dopa, is both enriched and abundant at the adhesive interface, and matches the molecular weight range seen in previous experiments of proteins localized to the adhesive interface after investigation of Dbfp7 peptide fragmentation favoring footprint enrichment. Separately, reanalysis of novel quagga mussel proteomic data has also resulted in the identification of additional byssal candidates, one of which may represent Dbfp0, the largest quagga mussel byssal protein to be identified.

2.6 Acknowledgements

The authors gratefully acknowledge Paul Taylor and Jonathan Krieger of the SickKids Proteomics, Analytics, Robotics Chemical Biology Centre (SPARC BioCentre) at the Peter Gilgan Centre for Research and Learning (PGCRL) in Toronto for LC-MS/MS and relative protein quantification analysis and guidance.
Appendix

2.A  All Venn Diagram Sectors: Other Protein Roles

2.A.1  Structural Proteins

Present in each sector and shown in green in Figure 2.2 are proteins with significant BLAST matches to known structural proteins. Structural proteins are abundantly present among the 385 proteins but are more highly represented among the overlapping sectors of footprint, plaque, and induced proteins. This is a logical finding as structural proteins (and their precursors, in the case of induced proteins) are expected to be found among all sources. Structural proteins may have a large role in encouraging strength of the attachment structure (plaque), but it is not unfeasible that proteins at the adhesive interface utilize ECM qualities as part of an interfacial role.

Papilin was a top BLAST match to proteins within the footprint-only, plaque+induced, and induced-only sectors, with each match belonging to the same component number (comp49340). Papilin is a primary constituent of basement membranes with homologies to ADAMTS group metalloproteases (a disintegrin and metalloproteinase with thrombospondin motifs)\[85, 86, 87\], which has particular significance as a protease enzyme whose catalytic mechanism involves a metal, the significance of which is further elucidated below. Although six separate sequences had very strong homology (E \(< 10^{-104}\)) to papilin, they belonged to the same component number and appeared equally distributed among the footprint, plaque, and induced protein sectors. This underscores the hesitation that must be held when isolating one role by localization completely. It would be expected that as a basement membrane component, papilin would have a structural role that is emphasized in the plaque component to promote structural integrity of the byssus to the substrate as well as to encourage movement of potential footprint and adhesive proteins into the interfacial space between the plaque and substrate. Kielin/chordin-like proteins, seen abundantly in the footprint+plaque relative quantification experiment, are also highly represented among each sector.

2.A.2  Mucin-like Proteins

Mucin-like components are the top BLAST matches for 8 of the 385 proteins across all sectors shown in Figure 2.2 except for the induced-only sector. It is surprising that there are no mucus-like components within the induced-only sector as this is the most likely region to find proteins which are on the surface of the foot, and the collection of induced proteins was the only region from which residue from the foot organ could have been inadvertently included. 7 of the 8 transcripts with BLAST matches to mucin-like proteins have very low E-values, one of which is 0 (an exact match, especially for a large protein such as mucin), and the others which reach \(10^{-160}\). One transcript displays homology to a protein with a top BLAST match to IgGFc-binding protein. Structurally, the predicted amino acid sequence based on the cDNA of IgGFc binding protein (Fc\(\gamma\)BP) in human colonic epithelial cells contains 12 repeats of a 400-amino acid cysteine-rich unit that resembles that which is found in mucin\[88\]. Fc\(\gamma\)BP is found to be widely distributed in mucus cells and secreted mucins, while also being expressed on mucosal surfaces\[88, 89\].
Both of the query transcripts have the highest molecular weights of the proteins that were found abundantly in the adhesive interface, which shares a characteristic with mucin, as it is a similarly high molecular weight family of proteins. CGLCGN (Figure A1) a motif in Fc\(\gamma\)BBP that is conserved in MUC2, MUC5AC, MUC5B (oligomeric mucus gel-forming protein in humans) and prepro-von Willebrand factor (associated with platelet adhesion and platelet-to-platelet cohesion during thrombus formation\cite{90}) is present in both query transcripts with good alignment among the query sequence, the top BLAST hit, as well as the IgGFc binding protein found in humans. Although the query transcripts are not cysteine-rich (as is typical of mucin and glycosylated proteins) and are instead most abundant in lysine, glycine, and glutamate, both transcripts align with the C8 domains present in their respective top BLAST hits and pertinent native proteins. This domain contains eight conserved cysteine residues, which are found in disease-related proteins including zonadhesin and mucin. Both transcripts also contain trypsin inhibitor-like (TIL) domains that are a cysteine-rich family containing trypsin inhibitors found in many extracellular proteins.

The alignment between query transcripts and the mucin-5AC-like BLAST hit displays strong alignments at the regions covered by the thrombospondin type-1 (TSP-1) superfamily. These TSP-1 repeats, which were first identified in thrombospondin proteins, are composed of glycoproteins that function at cell surfaces and in the extracellular matrix milieu\cite{91}. Lastly, the query transcripts align strongly to the amino acids associated with the TIL cysteine-rich domain. Thus, the presence of mucin-like characteristics across all three transcripts such as high molecular weight (>100,000 kDa) and the abundance of cysteine-rich domains that align in these regions permits the possibility that in addition to having a mucosal role, these proteins may be involved in the formation of structural networks that rely on extracellular activity.
2.A.3 Redox Proteins

A number of transcripts are associated with proteins that have potent metal-binding and redox capabilities (metal-binding protein and metallothionein 4 respectively). Found within the induced-only sector are seven strong ($E < 10^{-107}$) BLAST matches from the same component number (comp54245) whose top hits unanimously return myeloperoxidase-like proteins, which are involved in oxidative stress and inflammation within humans[92]. Peroxidase presence within the shell matrix of Lottia gigantea, a commonly found species among BLAST searches of quagga mussel proteins, suggests that peroxidase is involved in bimineral-hydrogel formation within the matrix framework assembly[93, 94]. Such strong representation of myeloperoxidase-like proteins within the induced proteins suggests a more likely role of involvement in shell-to-mussel attachment and foot organ maintenance.

The presence of BLAST hits linked to oxidoreductase-modifying activity that has been shown to occur in marine mussels is a positive indicator of similar activity within quagga mussels. Two transcripts with amine oxidase matches with very low E-values ($10^{-170}$) suggest that some proteins function to catalyze oxidative cleavage of alkylamines. However, it is imperative to note that not every match with a redox quality will be related to quagga mussel adhesion, even though the presence of Dopa is a distinctive characteristic to mussel adhesion. For example, filter-feeding bivalves such as mollusks are frequently exposed to polycyclic aromatic hydrocarbons (PAHs), a form of ubiquitous environment pollutants, and have thus evolved methods to combat oxidative stress[95]. Support for these protective methods may in fact be the intended role of oxidative proteins.

2.A.4 Mussel Proteins

The only significant match within the induced-only sector to a true mussel protein was to transgelin, an actin-binding protein within Mytilus coruscus, a marine bivalve mollusc. As a member of the calponin family, transgelin contains a positively charged region required for actin binding[96]. Other Mytilus species with similar calponin domains and/or calponin-homology domains have been found exclusively in the myostracum of the Mytilus galloprovincialis shell[97]. It is hypothesized that the potential role in calponin-actin interaction may contribute to myostracum-muscle attachment in the mussel. With a low E-value ($10^{-90}$), it is likely that the presence of this protein is in fact unrelated to mussel adhesion, but rather may be involved with shell interactions within the mussel. A separate transcript within the induced-only sector featured a top BLAST match to an insoluble matrix shell protein within Ruditapes philippinarum, the Manila clam, with a slightly higher E-value of $10^{-39}$. The presence of these two proteins within the induced-only sector indicates the high likelihood of their role as byproducts of inclusion via the collection method used for artificially induced proteins. As the mussel was required to be splayed open, it is not unfeasible that traces of nacreous components were incorporated along with induced proteins.

Several proteins held confident matches to proteins related to a mussel species (e.g. M. galloprovincialis - comp42073 in the induced-only sector). However, due to the fact that the top BLAST match was a hypothetical protein, it was categorized as such as opposed to a mussel protein. Similarly, found within the plaque-only sector were three transcripts for comp46484 with BLAST matches to Mytilus californianus foot protein 14 - however, the E-value was above the threshold at $10^{-4}$. Despite this, as
this was the only BLAST result which displayed a match to an existing mussel protein, it was included in Figure 2.2 for its remarkability and close match to the E-value threshold of $10^{-6}$.

2.A.5 Protease Inhibitors

Protease inhibitors prevent degradation of proteins that may be imperative for mussel function, whether for general metabolism, cellular, or adhesive functions. Protease inhibitor proteins and domains have been found in other mollusk shell matrices[98, 99], suggesting a protecting system for extracellular proteolysis, biomineralization processes contributing to the mechanical properties of mollusk shell nacre[100], and within the extrapallial fluid (EPF), potentially contributing to the innate immune system in mussels[101]. Protease inhibitor proteins are found among sectors featuring overlaps, particularly between footprint+induced and footprint+plaque sectors. Protease inhibitors within the quagga mussel may thus have a role beyond shell formation, explaining their presence beyond the induced sector and suggestive of a role in the inhibition of crosslinking or degradation in processes related to adhesion, which are concentrated to the footprint and plaque components. The majority of protease inhibitor BLAST matches have low E-values ($10^{-19}$) and feature Kunitz-type domains. These domains are 50-70 amino acids in length, ubiquitous in nature, and function as serine protease inhibitors[102]. Four BLAST matches were made to Kunitz-like protease inhibitors and Kunitz domain-containing proteins, with E-values ranging from $10^{-13}$ to $10^{-29}$.

2.A.6 Tyrosinase

Tyrosinase catalyzes both the enzymatic hydroxylation of monophenols (e.g. tyrosine) into o-diphenols (e.g. Dopa), and o-diphenols into o-quinones[103, 104]. Along with other members of the type-3 copper protein superfamily, tyrosinase, tyrosinase-related proteins, and catechol oxidases are responsible for pigmentation, immunity, oxygen transport, and wound healing[105, 106, 107, 108]. It is found within vertebrates to regulate pigment synthesis[105, 106] and in invertebrates to support innate immunity through melanin and pigmentation processes, as melanin is able to physically encapsulate pathogens[107]. In mollusks, tyrosinase has been found to contribute to shell pigmentation via cross-linking and quinone-tanning, an ubiquitous and natural process also known as sclerotization by which structural protein-rich materials in nature are transformed from pliable to tough via oxidation of Dopa-containing proteins[109]. Although the Mytilus periostracum, the outermost shell of the mussel, is known to contain Dopa-containing proteins[110], this is also the case for the byssal components including the plaque and threads[109, 111, 112].

Among the 385 proteins, tyrosinase-like proteins with E-values ranging from $10^{-10}$ to $10^{-39}$ were found within the all-intersect sector, which is in accord with the known roles of tyrosinase within mussels and invertebrates. Tyrosinase proteins may be present within induced proteins because of mussel immunity or pigmentation processes within the foot organ. They may also be secreted from the secretory granules during the artificial induction process as a consequence of the need for the mussel to prepare for oxidation of threads and plaques during the natural secretion process. This explains the presence of tyrosinase within the naturally secreted footprint and plaque samples. It would be expected that tyrosinase-like proteins are instrumental to the quinone-tanning and cross-linking process linked to Dopa-containing proteins during the process of adhesion in spite of that fact that Dopa content within quagga mussels
is significantly lower than that of their marine counterparts. Oxidation of Dopa to dopaquinone with polyphenol oxidases and tyrosinases effectively eliminates mussel foot protein adsorption to surfaces. The ubiquitous presence of tyrosinase proteins within the quagga mussel footprint, plaque, and induced proteins, and specifically in a mussel with lower Dopa mol% than marine mussels suggests that a truly Dopa-independent method of adhesion may have a larger role than expected. It also suggests that the activity of tyrosinases might be to favor a non-Dopa method of adhesion that is so ubiquitous among freshwater mussels.

2.A.7 Cellular

Very strong BLAST matches to cellular proteins were found in the majority of sectors, with E-values ranging from $10^{-26}$ to 0. Cellular proteins include actin, tubulin, ribosomal proteins, and calmodulin. Presence of these proteins is detected in part to the exceptionally sensitive LC-MS/MS method used to sequence proteins.

2.B Relative Protein Quantification: Proteins of Interest

2.B.1 Rhinophore Proteins

The protein most abundant at the interface has a top BLAST result to a hypothetical rhinophore protein. The native protein associated for functional domain analysis includes *Aplysia G* proteins, an IP3 receptor, and a phospholipase C, which are expressed and localize to the olfactory sensory epithelium of the rhinophore[113]. Rhinophore genes are known to be upregulated following pheromone stimulation[114]. Rhinophore proteins are specialized sensory organs, act as a finely tuned nose, and play a key role in sensory detection and chemoreception in molluscs[114]. *Aplysia* have no acoustic or visual sense; thus water-soluble chemical cues are imperative to behavioral events such as defence, courtship, and aggregation[114]. Aligning with the second half of the query sequence of the top BLAST hit (hypothetical rhinophore protein) is an overlap with the DUF3011 superfamily. This bacterial family of proteins has no known function. When aligned with the alpha subunit of the guanine nucleotide-binding protein in rhinophore proteins of *A. californica*, the query sequence does not have strong amino acid matches, and most importantly, strong alignments were absent in functional regions of the protein, including binding and interaction sites. However, the identification of genes upregulated following pheromone exposure is a strategy to better understand the genetic basis of smell in molluscs and can be applied to the identification of receptors responsible for chemicals secreted in response to a need for adhesion. This is particularly important for marine molluscs which rely heavily on olfactory cues. Thus, even if this protein does not play a role in the adhesion of proteins, it may be conducive to the production and mediation of cues responsible for protein binding.

2.B.2 Structural Proteins

Represented in Figure 2.5b as green bars are known structural proteins, which include ECM proteins and kielin/chordin-like proteins. This category includes a large umbrella of ECM proteins which con-
tribute to structure including papilin, collagens, and non-collagenous glycoproteins such as tenascin. Glycosylated proteins provide structural support for cells, define tissue borders, and mediate intercellular interactions[115]. Molecular processes of cells and the availability of extracellular signals are highly mediated by ECM proteins, particularly by their functions as adhesive substrates which track migratory cells and contribute concentration gradients for cell migration[115, 116].

Kiellin/chordin-like proteins are highly represented within this footprint+plaque overlap. The majority of query sequences with BLAST matches to kiellin/chordin-like proteins are cysteine-rich. The native human protein associated with this BLAST result is the kiellin/chordin-like protein isoform 1 precursor in Homo sapiens. These kiellin/chordin-like proteins do not have conserved domains for analysis. However, the native KCP protein in Homo sapiens displays multiple superfamilies, one of which is a cysteine-rich (C8) domain. Another domain featured is the trypsin inhibitor-like (TIL) cysteine rich domain, which is found in many extracellular proteins. The regions that showed matching amino acids between the query sequence and this native kiellin/chordin-like protein only covered the regions featuring the Von Willebrand factor type C (VWC) superfamily. Named after VWC modules and also known as chordin-like cysteine-rich (CR) repeats, these regions have been found in approximately 200 ECM proteins[117]. These repeats are 58-75 residues in length and are defined by a consensus sequence of 10 cysteines with characteristic spacings[118]. Commonly associated with modulating growth factor signaling, these chordin CR repeats are part of the von Willebrand factor-C domain and are involved in maintaining homeostasis. Query transcripts of proteins found within the footprint, plaque, and induced samples display good alignment and matching amino acids in regions that are marked by the VWC superfamily region, both when aligned with the top BLAST match as well as the native kiellin/chordin-like protein isoform found in humans.

2.B.3 Mucus-like proteins

Mucus components present with a fold change ratio of 5-6x at the adhesive interface, two transcripts of which also display the highest spectra counts among all proteins analyzed in this relative quantification experiment. Mucus within humans is secreted from goblet cells and is imperative to intestinal mucosal defence[89]. Mucin, the primary macromolecule of which mucus is composed, is an integral component of the glycans within mucus. Certain mucin glycan motifs favor the adhesion of particular microorganisms[119]. Mucus components are likely to be found in sectors which involve protein interaction directly with the surface of the mussel foot, which may produce mucus-like substances to prime the surface of a substrate prior to adhesion as a native coating to keep the surface of the foot moist and prepared for potential interactions, or as a natural method to prevent bacterial invasion in natural environments. The presence of mucus components within the plaque and footprint sectors is logical - prior to secreting threads and plaques naturally, the mussel also actively probes the surface of the substrate with its foot prior to adhesion, thereby leaving mucus-like residues on the surface of the slide to be inadvertently collected during footprint and plaque collection.

2.B.4 Redox proteins

Presenting with a fold change ratio of 3.8x enrichment at the interface, myeloperoxidase-like proteins are involved in oxidative stress and inflammation within humans[92]. Peroxidases are iron proteins which
catalyze the oxidation of amines and phenols by hydrogen peroxide[93] and have a breadth of functions. Peroxidase has been found to catalyze Dopa reactions within Lymnaea stagnalis[120] and is involved in melanin biosynthesis in Sepia officinalis ink glands[121]. The unique structural domain of peroxidase, featuring two histidines and one calcium-binding site, suggests that peroxidase functions to support protein structure in the heme environment[93, 121]. This redox protein is also present within artificially induced proteins, implying a thorough role of myeloperoxidase-like protein within mussel activity and metabolism.

2.B.5 Plasminogen-like Proteins

One transcript was found with BLAST matches found most strongly associating with plasminogen, an enzyme responsible for dissolving blood clots. Plasminogen is converted to its active form (plasmin) within the liver, where one of its roles is to promote migration of pathogenic bacteria through the human extracellular matrix[122]. Plasminogen has many ligand binding sites and domain interaction sites. The query transcript aligns well to the regions marked by the ligand binding site in both the top BLAST and native protein. Kringle domains are well-established in plasminogen and are modules that correspond to structural and folding domains which mediate binding[123]. Kringles of proteases (such as plasminogen) are thought to be evolved from a common ancestral protein binding molecule. It is not unreasonable that the mediation of protein-protein interactions is a putative role of this query transcript.

2.C Justification for Targeted Footprint+Plaque Overlap Search

Thorough plaque byssal protein integration with footprint byssal proteins ensures that what would be a pure footprint sample is instead “tainted” with plaque proteins which are well-ingrained among the adhesive interface. Following this logic, the presence of many unique proteins within the plaque sector is questionable, as surely there are footprint proteins tightly incorporated when separating footprint and plaque once secreted and set down by the mussel. The presence of proteins unique to the plaque-only sector can be explained by multiple factors - firstly, that the size of the plaque and amount of protein within one single plaque sample far outweighs that of a single footprint sample. Perhaps a more keen observation to make is that the number of plaque-only proteins is, too, an underestimate of the true number of byssal proteins likely to participate in adhesion. We are more likely to observe plaque-only proteins over footprint-only proteins because the relative area and region in which plaque proteins are concentrated in (i.e. well above where the interfacial interaction exists between footprint and plaque), and thus, where the footprint is separated from the plaque, is much larger. It is likely that due to their close interactions with footprint proteins, there are some plaque Dbfp proteins which are truly unique to the plaque that inadvertently end up within the overlap sectors, specifically the footprint+plaque and all-intersect sectors. Thus, it would not be prudent to assume that a protein’s sole presence within the plaque sector precludes itself of a role in adhesion, for proteins are collected in the interface in a relatively crude manner via a razor blade separating footprint from plaque. This actually ensures that we are including the possibility that proteins nearer to the plaque-substrate interface, but not necessarily solely within the footprint itself, are collected and considered as proteins which are indeed not exclusive to the plaque.
The presence of distinct Dbfp’s within the sectors which include an overlap with induced proteins requires a separate explanation, as the sources of these proteins not only differ in localization (footprint/plaque versus foot) but also in the temporal state (not cross-linked) and trigger (natural vs artificial) for secretion. Overlaps between footprint and induced sectors indicate that proteins held within the secretory granules of the mussel foot (in preparation for release) hold closely similar forms to Dbfp’s which are secreted naturally. At this point, we are unable to determine if the proteins are spliced differently when held within the secretory granules versus for natural secretion, due to the bottom-up proteomic approach. It is nevertheless useful to be able to confirm that Dbfp’s within the footprint+induced sector are found consistently at the interface and within the mussel foot. Overlap between mussel proteins within the plaque+induced sector follows the same logic. It is not surprising that there are such a high number of peptide counts and Dbfp’s present within the all-intersect sector, as artificially induced proteins collected and sequenced did not account for separate artificially induced thread, plaque, or footprint proteins.

Dbfp’s found solely within the induced sector may have an exclusive role within the secretory granules, exist uniquely on or within the mussel foot, may be a form of Dbfp which exists uniquely as a pre-secretion byssal protein, or may be a Dbfp which composes a pre-crosslinked thread protein. However, the presence of Dbfp7 and Dbfp10 within at least one other sector of Figure 2.4, combined with the low peptide counts (total 5) in the induced-only sector, lends itself to the explanation that Dbfp7 and Dbfp10 are unlikely to have a sole role within the induced sample of proteins. Dbfp distribution among the entire footprint and entire plaque sector indicates that Dbfp2 and Dbfp7 are the most prominent Dbfp’s within the naturally secreted sectors overall. However, while Dbfp2 was found in approximately equal abundance in the entire footprint and entire plaque sectors (238 peptides), Dbfp7 was slightly more abundant within the footprint sector (251 peptides). In addition, Dbfp7 but not Dbfp2 had reasonable representation within the footprint+plaque sector (77 peptides), which holds particular relevance as one would expect that a potential adhesive protein would have prominence as a uniquely naturally secreted protein found within both the footprint and plaque sectors.

The absence of Dbfp proteins unique to the footprint-only sector suggests that the most logical regions of overlap are regions where footprint and plaque proteins may both be found - namely, the all-intersect and footprint+plaque sectors. This rationale is compatible with the experimental caveat which outlines the difficulty in isolating “pure” footprint and “pure” plaque proteins as a consequence of close interactions between proteins and the crude method of separating plaque from footprint. For these reasons, further analysis of relative abundance of proteins within these sectors will provide insight into the identities and distribution of proteins found in both the footprint and plaque sectors.
Table A1 – Top 30 naturally secreted proteins ranked by decreasing relative enrichment at adhesive interface compared to the bulk plaque. Byssal proteins are bolded.

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Fold Change Ratio</th>
<th>Spectral Count</th>
<th>BLAST Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>comp42807_c0_seq2-2R</td>
<td>16.2621</td>
<td>2</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp44473_c0_seq2-2F</td>
<td>11.4611</td>
<td>4</td>
<td>Rhinophore</td>
</tr>
<tr>
<td>comp12176_c0_seq2-2R</td>
<td>10.8976</td>
<td>8</td>
<td><strong>Dbfp18</strong></td>
</tr>
<tr>
<td>comp50443_c0_seq4-0F</td>
<td>9.8897</td>
<td>5</td>
<td>Structural</td>
</tr>
<tr>
<td>comp50443_c0_seq7-0F</td>
<td>8.6762</td>
<td>2</td>
<td>Structural</td>
</tr>
<tr>
<td>comp43807_c0_seq3-2F</td>
<td>7.9649</td>
<td>2</td>
<td>Cellular</td>
</tr>
<tr>
<td>comp50443_c0_seq2-2F</td>
<td>7.7704</td>
<td>2</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp42807_c0_seq1-2R</td>
<td>7.2657</td>
<td>4</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp58428_c0_seq11-0F</td>
<td>7.1483</td>
<td>2</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp52859_c0_seq1-1R</td>
<td>6.4856</td>
<td>8</td>
<td>Mucus component</td>
</tr>
<tr>
<td>comp41116_c0_seq1-2F</td>
<td>6.249</td>
<td>9</td>
<td>Structural</td>
</tr>
<tr>
<td>comp53519_c2_seq1-2R</td>
<td>5.7961</td>
<td>88</td>
<td>Mucus component</td>
</tr>
<tr>
<td>comp53519_c2_seq2-0R</td>
<td>5.7749</td>
<td>92</td>
<td>Mucus component</td>
</tr>
<tr>
<td>comp44551_c0_seq1-2R</td>
<td>4.9042</td>
<td>2</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp36226_c0_seq1-1F</td>
<td>4.8766</td>
<td>7</td>
<td>Structural</td>
</tr>
<tr>
<td>comp30100_c0_seq2-2F</td>
<td>4.6203</td>
<td>14</td>
<td>Other</td>
</tr>
<tr>
<td>comp56420_c11_seq1-2F</td>
<td>4.2106</td>
<td>10</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp56420_c6_seq1-1F</td>
<td>4.2045</td>
<td>2</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp37515_c0_seq1-1F</td>
<td>4.1639</td>
<td>5</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp57624_c1_seq1-1R</td>
<td>4.0258</td>
<td>2</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp31275_c0_seq2-2R</td>
<td>3.999</td>
<td>17</td>
<td>Structural</td>
</tr>
<tr>
<td>comp53519_c2_seq3-0R</td>
<td>3.9721</td>
<td>6</td>
<td>Plasminogen</td>
</tr>
<tr>
<td>comp31287_c0_seq1-2F</td>
<td>3.8896</td>
<td>2</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp51774_c0_seq1-2F</td>
<td>3.7852</td>
<td>7</td>
<td>Redox</td>
</tr>
<tr>
<td>comp31275_c0_seq1-2R</td>
<td>3.6139</td>
<td>17</td>
<td>Structural</td>
</tr>
<tr>
<td>comp52765_c1_seq12-1F</td>
<td>3.5545</td>
<td>37</td>
<td><strong>Dbfp7</strong></td>
</tr>
<tr>
<td>comp30932_c0_seq1-0F</td>
<td>3.5513</td>
<td>6</td>
<td>Structural</td>
</tr>
<tr>
<td>comp41375_c0_seq1-1R</td>
<td>3.4629</td>
<td>3</td>
<td>Other</td>
</tr>
<tr>
<td>comp56544_c0_seq1-2R</td>
<td>3.4237</td>
<td>41</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp53074_c1_seq8-0R</td>
<td>3.3348</td>
<td>4</td>
<td>Unknown</td>
</tr>
</tbody>
</table>
### Table A2 – Top 30 naturally secreted proteins abundant at the adhesive interface ranked by decreasing abundance by total spectral count. Byssal proteins are bolded.

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Fold Change Ratio</th>
<th>Spectral Count</th>
<th>BLAST Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>comp53519_c2_seq2-0R</td>
<td>5.7749</td>
<td>92</td>
<td>Mucus component</td>
</tr>
<tr>
<td>comp53519_c2_seq1-2R</td>
<td>5.7961</td>
<td>88</td>
<td>Mucus component</td>
</tr>
<tr>
<td>comp56544_c3_seq1-2R</td>
<td>3.2697</td>
<td>50</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp44862_c1_seq3-2F</td>
<td>1.5904</td>
<td>49</td>
<td>Dbfp8</td>
</tr>
<tr>
<td>comp44862_c1_seq1-2F</td>
<td>2.0411</td>
<td>43</td>
<td>Dbfp8</td>
</tr>
<tr>
<td>comp56544_c0_seq1-2R</td>
<td>3.4237</td>
<td>41</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp52765_c1_seq14-1F</td>
<td>1.9349</td>
<td>39</td>
<td>Dbfp7</td>
</tr>
<tr>
<td>comp52765_c1_seq12-1F</td>
<td>3.5545</td>
<td>37</td>
<td>Dbfp7</td>
</tr>
<tr>
<td>comp52765_c1_seq3-0F</td>
<td>2.1062</td>
<td>34</td>
<td>Dbfp7</td>
</tr>
<tr>
<td>comp52765_c1_seq9-1F</td>
<td>1.8793</td>
<td>34</td>
<td>Dbfp7</td>
</tr>
<tr>
<td>comp52765_c1_seq11-1F</td>
<td>2.7556</td>
<td>33</td>
<td>Dbfp7</td>
</tr>
<tr>
<td>comp52765_c1_seq13-1F</td>
<td>0.7292</td>
<td>27</td>
<td>Dbfp7</td>
</tr>
<tr>
<td>comp56420_c14_seq2-1F</td>
<td>2.0673</td>
<td>25</td>
<td>Other</td>
</tr>
<tr>
<td>comp35860_c0_seq1-2R</td>
<td>1.3645</td>
<td>19</td>
<td>Dbfp15</td>
</tr>
<tr>
<td>comp31275_c0_seq2-2R</td>
<td>3.999</td>
<td>17</td>
<td>Structural</td>
</tr>
<tr>
<td>comp31275_c0_seq1-2R</td>
<td>3.6139</td>
<td>17</td>
<td>Structural</td>
</tr>
<tr>
<td>comp30100_c0_seq2-2F</td>
<td>4.6203</td>
<td>14</td>
<td>Other</td>
</tr>
<tr>
<td>comp53709_c0_seq1-2F</td>
<td>1.0014</td>
<td>13</td>
<td>Other</td>
</tr>
<tr>
<td>comp53132_c0_seq2-2R</td>
<td>1.3048</td>
<td>11</td>
<td>Tyrosinase</td>
</tr>
<tr>
<td>comp56420_c11_seq1-2F</td>
<td>4.2106</td>
<td>10</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp41116_c0_seq1-2F</td>
<td>6.249</td>
<td>9</td>
<td>Structural</td>
</tr>
<tr>
<td>comp12176_c0_seq2-2R</td>
<td>10.8976</td>
<td>8</td>
<td>Dbfp18</td>
</tr>
<tr>
<td>comp52859_c0_seq1-1R</td>
<td>6.4856</td>
<td>8</td>
<td>Mucus component</td>
</tr>
<tr>
<td>comp56544_c8_seq1-1R</td>
<td>3.1639</td>
<td>8</td>
<td>Other</td>
</tr>
<tr>
<td>comp36226_c0_seq1-1F</td>
<td>4.8766</td>
<td>7</td>
<td>Structural</td>
</tr>
<tr>
<td>comp51774_c0_seq1-2F</td>
<td>3.7852</td>
<td>7</td>
<td>Redox</td>
</tr>
<tr>
<td>comp56420_c10_seq1-1F</td>
<td>1.6145</td>
<td>7</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp19599_c0_seq1-2R</td>
<td>1.4558</td>
<td>7</td>
<td>Unknown</td>
</tr>
<tr>
<td>comp31337_c0_seq1-0F</td>
<td>1.0726</td>
<td>7</td>
<td>Dbfp6</td>
</tr>
<tr>
<td>comp53519_c2_seq3-0R</td>
<td>3.9721</td>
<td>6</td>
<td>Plasminogen</td>
</tr>
</tbody>
</table>
Chapter 3

Dbfp7 Purification and Characterization

Judith Ng, Magdalena Wojtas, Kate Li, David Rees, Emma Master, Jayachandran Kizhakkedathu, Eli Sone

Chapter 3 is in preparation for a manuscript to be submitted for peer-reviewed journal publication. All work was performed by Judith Ng. Magdalena Wojtas’ contribution was a high level of assistance and guidance with protein purification. Kate Li’s contribution was the analysis of Dbfp7 sequence breakdown. David Rees’ contribution was the collection and purification of protein samples which led to the identification of Dbfp7 as a top candidate. Emma Master and Jayachandran Kizhakkedathu are collaborators.

3.1 Abstract

The quagga mussel (Dreissena bugensis) has the remarkable activity to selectively attach to a plethora of synthetic and organic surfaces via the byssus, a proteinaceous anchor which is composed of proteins originating within the mussel foot organ, secreted as a liquid and rapidly crosslinking to form multiple threads, each terminating distally in a flattened, protein-rich adhesive pad known as the plaque. The identity, structure, and role of these proteins has piqued interest due to their unique ability to display resilient adhesive qualities in turbulent, wet, and ionic conditions. Our group has identified a leading quagga mussel protein candidate (Dreissena bugensis foot protein 7 - Dbfp7) responsible for this adhesive activity. For the first time, we purified Dbfp7 directly from the mussel phenol gland and demonstrated its purity and identity with MALDI-TOF, SDS-PAGE, and LC-MS/MS. Preliminary circular dichroism, QCM-D, and atomic force microscopy experiments were performed to elucidate protein adsorption activity and protein structure. Dbfp7 as purified from the phenol gland assumes a primarily $\beta$-sheet and disordered structure in solution and adsorbs to SiO$_2$, TiO$_2$, and Au substrates with thicknesses of 9-18 nm. This work has been a significant step forward in determining putative candidates for novel adhesive proteins responsible for this adhesive activity, as well as in establishing methods to purify quagga mussel.
proteins of interest, and adds on to the significant progress made by our laboratory in understanding freshwater mussel adhesion.

3.2 Introduction

The quagga mussel (*Dreissena bugensis*) and the closely related zebra mussel (*Dreissena polymorpha*) are invasive freshwater mussel species indigenous to the river basins of the Black, Baltic, and Caspian seas, known for their rapid spread in the waterways of eastern North America due to their ability to adhere to a variety of surface chemistries underwater[2, 3, 55]. The mussels’ ability to carpet virtually all underwater surfaces has resulted in serious environmental and economic consequences, incurring costs from the need for mussel removal, retrofitting equipment, prevention of biofouling, adhesion control, and lost time[2, 124]. Freshwater mussels foul by attaching opportunistically to an extensive breadth of surfaces by the means of a proteinaceous anchor known as the byssus, a bundle of threads composed of fibrous extracellular structures which terminate in adhesive pads known as plaques, many of which contain the post-translationally modified amino acid 3,4-dihydroxyphenylalanine (Dopa)[51]. Dopa has been well-studied in marine mussels e.g. *Mytilus* spp., *Perna* spp.) to be the primary enabling mechanism behind the attachment of byssi to almost any submerged surface[125], due to its high concentrations within mussel foot protein-3 (Mfp-3) and Mfp-5 (20 and 30 mol% respectively)[11]. Dopa, a post-translational hydroxylated form of tyrosine, is able to bind non-covalently to metal oxide surfaces via hydrogen bonding[45] and form cation-mediated complexation cross-links such as cation-π interactions[22]. Upon oxidation to Dopa-quinone by catechol oxidase, the catechol group is available to form covalent cross-links with Dopa and other residues[51].

Although there exists extensive literature on the molecular mechanism, involved proteins, and their composition and distribution within the mussel plaque and thread of marine mussels, the mechanism behind the impressive adhesive capacity of freshwater mussels has not been well-characterized. Distinctly, overall Dopa-containing protein concentrations are lower within freshwater mussels: 0.6% for zebra mussels[51] and 0.1% for quagga mussels[52], indicating that a novel mechanism of adhesion is potentially responsible for the significant adhesive abilities of freshwater mussels. In order to determine the relevant proteins and associated mechanism for adhesion, it is critical to investigate the proteins at the adhesive interface between the substrate and mussel plaque, as these proteins are likely to have a significant role in mediating this unique interaction which persists in underwater and turbulent conditions, a setting which is typically averse to adhesive performance[8]. There has been significant progress in the understanding of zebra mussel[54] and quagga mussel[56] byssal protein spatial variation as performed by our group, in addition to the identification of novel zebra mussel byssal proteins[53] and the development of a *de novo* assembled quagga mussel foot transcriptome (manuscript in progress).

This extensive proteomic investigation serves to elucidate both sequence and localization data regarding proteins found abundantly at the plaque-substrate interface. We built on this background of quagga mussel protein composition and spatial variation by determining for the first time, with a novel protein collection protocol and tandem mass tag technique, the relative enrichment and abundance (spectral count) of proteins found within the quagga mussel plaque and adhesive interface. The top protein candidate hypothesized to play a major role in adhesion was *Dreissena bugensis* foot protein 7 (Dbfp7), as determined through extensive spatial analysis techniques including MALDI-TOF and relative protein
quantification between the mussel bulk plaque and its associated adhesive footprint. Dbfp7 was purified natively from the phenol gland of the quagga mussel by dissection, homogenization, solubilization in denaturing agents, and reverse-phase chromatography.

Understanding the interaction of proteins with solid surfaces is critical to the applications within the biomaterial field as protein adsorption is the first step in the integration of an implanted device, or in the case of an adhesive, in its interaction with tissue[126, 127]. Fundamentally, protein-surface interactions involve both protein binding and folding; thus, studies into this phenomena contribute to the body of knowledge on protein biophysics as a whole[128]. As an adhesive candidate, Dbfp7 is projected to play a role in interfacial interactions on substrates. Thus, protein adsorption experiments and models provide an unprecedented view into the behaviour of Dbfp7 behaviour on a surface. Optical characterization techniques such as circular dichroism (CD) provide data on the secondary structure of the protein in solution. Surface characterization techniques, most pre-dominantly quartz crystal microbalance with dissipation monitoring (QCM-D), reveal highly sensitive (in the ng cm\(^{-1}\) range) time-resolved adsorption data by which an adsorbed mass, \(\Delta m\), can be deduced by measured changes in the resonant frequency, \(\Delta f\) of an oscillating quartz sensor. QCM-D is particularly useful for its superiority to optical methods such as ellipsometry in its ability to determine the rigidity of an adsorbed film from layer hydration, viscosity, and shear-wave propagation[129]. Traditional analytical techniques to assess protein molecular weight (SDS-PAGE and MALDI-TOF) were also performed to confirm protein sizes and fragments. Liquid chromatography tandem-mass spectrometry (LC-MS/MS) of purified Dbfp7 was performed to evaluate sample purity and to determine sequences of proteins obtained from purification directly from the phenol gland.

### 3.3 Methods

#### 3.3.1 Protein Extraction

Dbfp7 was extracted from the foot organ of *D. bugensis* mussels which were collected from the origin of the Credit River in Mississauga, Ontario and immediately transferred to holding tanks with circulating seawater at 12°C. The phenol glands were dissected in groups of 25 mussels and homogenized in 2 mL of 5% acetic acid containing 10 \(\mu\)M of protease inhibitors leupeptin and pepstatin (Sigma-Aldrich, St. Louis, MO). Tissues were thoroughly homogenized in a chilled ground glass tissue grinder (Kontes, Vineland, NJ). The homogenate was centrifuged at 20,000xg at 4°C for 40 minutes, producing supernatant S1 and pellet P1. P1 was isolated, homogenized in 2 mL of 5% acetic acid, 8M urea and centrifuged at 20,000xg at 4°C for 40 minutes, producing supernatant S2 and pellet P2. The contents and purity of supernatant S2 were assessed using SDS-PAGE. S2 was retained for purification by high-performance liquid chromatography (HPLC) with a C-18 column (250 x 4.6 mm, XBridge BEH130, 130Å) (Waters, Milford, MA). The proteins were eluted with an aqueous acetonitrile (Caledon) gradient containing 0.1% trifluoroacetic acid (TFA) (Sigma-Aldrich, St. Louis, MO) at 1 ml/min. The column eluate was monitored for absorbance at 280 nm and 220 nm, with 1 ml fractions strongly absorbing at 220 nm individually lyophilized and re-suspended in 5% acetic acid. Polyacrylamide gel electrophoresis with 12% polyacrylamide was performed to assess the purity of the protein fractions. The gels were run at a constant voltage of 165V. Two sets of gels were prepared for each sample to detect protein bands.
by following the associated staining and destaining procedures with Coomassie Blue R-250 (Bio-Rad, Hercules, CA) and ProteoSilver Silver Stain (Sigma-Aldrich).

3.3.2 Amino Acid Analysis

Amino acid analysis was performed at the SickKids Proteomics, Analytics, Robotics Chemical Biology Centre (SPARC BioCentre) at the Peter Gilgan Centre for Research Learning at the Hospital for Sick Children, Toronto, Canada. Briefly, samples were dried in pyrolyzed borosilicate tubes in a vacuum centrifugal concentrator and subjected to vapor phase hydrolysis by 6M HCl with 1% phenol at 110°C for 20-24 hours under a pre-purified nitrogen atmosphere. After hydrolysis, excess HCl was removed by vacuum and hydrolyzates were washed with a re-drying solution of methanol, water, triethylamine (2:2:1) and derivatized at room temperature for 20 min with solution of methanol, water, triethylamine, and phenylisothiocyanate (PITC) (7:1:1:1). After evaporation and re-drying, the derivatized sample was dissolved in sample diluent (pH 7.4) and an aliquot was injected into the column, running on a modified PICO-TAG gradient. Columns were operated at a temperature of 48°C. The derivatized amino acids were detected at 254 nm. The Waters Acquity UPLC system (Milford, Massachusetts, USA) employed consists of a Binary Solvent Manager Module, a Sample Manager Module, a TUV Detector Module and a Waters Acquity UPLC BEH C18 column (2.1 X 100 mm). Data was collected, stored and processed using Waters Empower 3 Chromatography software.

3.3.3 MALDI-TOF MS Spectrometry

MALDI-TOF spectrometry was performed at the Analytical Facility for Bioactive Molecules (AFBM) at The Hospital for Sick Children (SickKids), Toronto, Canada. Two MALDI matrices were used: 10 mg ml\(^{-1}\) of sinapinic acid (SA) and \(\alpha\)-cyano-4-hydroxycinnamic acid (CHCA), dissolved in 50% water and 50% acetonitrile. A 3:4 ratio of matrix and Dbfp7 were suspended in water at a concentration of 0.5 \(\mu\)g \(\mu\)l\(^{-1}\). Samples were vortexed and 1 \(\mu\)l was spotted onto a MALDI plate. The samples were air-dried and inserted into the MALDI-TOF/TOF 4800 (Applied Biosystems, Foster City, CA). The machine was used in linear mid-mass positive ion operating mode with a source voltage of 20 kV and a detector voltage of 2.19 kV. Laser power was set to 6800 units. The number of average shots for each spectrum was 400.

3.3.4 Liquid Chromatography Tandem Mass Spectroscopy (LC-MS/MS)

LC-MS/MS of extracted peptides was performed at the SickKids Proteomics, Analytics, Robotics & Chemical Biology Centre (SPARC BioCentre) at the Peter Gilgan Centre for Research Learning at the Hospital for Sick Children. The digested peptides were loaded onto a 100 \(\mu\)m ID pre-column (Dionex) at 4 \(\mu\)l/min and separated over a 50 \(\mu\)m ID analytical column (C18 2 \(\mu\), Dionex). The peptides were eluted using a 0 to 35% acetonitrile gradient with an EASY-nLC 1000 nano-chromatography pump (Thermo Fisher, Odense Denmark). Data was acquired at 70,000 FWHM resolution in the MS mode and 17,500 FWHM in the MS/MS mode. A total of 10 MS/MS scans were obtained per MS cycle.
3.3.5 Protein Identification, Selection, and Analysis

Using the proteomics software PEAKS8 (Bioinformatics Solutions Inc., Waterloo, ON), de novo sequences derived from MS/MS data were matched against the quagga mussel foot transcriptome with tyrosine hydroxylation to Dopa set as a variable modification. Parent ion and fragment ion mass tolerances were set to 5 PPM and 0.01 Da respectively and hits were manually confirmed by inspecting the spectra. In PEAKS, the identification probabilities of the protein and peptide matches are indicated by the formula \(-10\log P\). Our acceptance criteria for identifying novel proteins in PEAKS8 are a peptide \(10\log P \geq 15\), Protein \(-10\log P \geq 50\), and de novo average local confidence (ALC) score \(\geq 80\%\), with at least two peptide spectra identification, unless otherwise stated. Protein percent coverage is the amount of the protein identified by supporting spectra. Signal peptides on proteins were identified using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/). Proteins with signal peptides were examined for protein homology using NCBI Protein Basic Local Alignment Search Tool (BLAST) using no adjustment method to compensate for amino acid composition and using filtering low complexity regions[82]. Sequences were searched against the NCBI non-redundant (nr) database, which consists of sequences from Swiss-Prot, Protein Information Resource (PIR), Protein Data Bank (PDB), and NCBI protein reference sequence (RefSeq) databases.

3.3.6 Circular Dichroism

CD spectra of surface-immobilized Dbfp7 were collected using a Jasco J-800 spectropolarimeter and a 0.1-cm-path-length quartz cell. Solution samples were prepared at a concentration of 500 \(\mu\)g ml\(^{-1}\) in 20 mM acetate buffer (pH 5.0). The spectra were obtained over a wavelength range of 190-260 nm, using a continuous scanning mode with a response time of 1 s with 0.5 nm steps, a bandwidth of 1.5 nm, and a scan speed of 50 nm/min. To improve the signal-to-noise ratio, an average of 100 scans was obtained. The temperature of the sample compartment was kept at 25°C by means of a water chamber. Spectra were corrected by subtracting the background (20 mM acetate buffer spectrum) from the sample spectrum. Secondary structure content was calculated using CONTINLL[130].

3.3.7 Quartz Crystal Microbalance

The 5-MHz piezoelectric quartz crystals coated with silicon, titanium, and gold were purchased from Q-Sense (Västra Frölunda, Sweden). The crystals were ozone-treated for 10 min prior and after immersion in 2% SDS purified water (Milli-Q, Millipore) for 30 min. The cleaned surfaces were then washed thoroughly with purified water and dried with nitrogen gas. Measurements by quartz crystal microbalance with dissipation (QCM-D) were performed with a Q-Sense E4 QCM (Västra Frölunda, Sweden) instrument with four-channel peristaltic pump (Ismatec SA, Glattburg, Switzerland) instrument with controlled flow. The samples were dissolved in degassed 20 mM acetate buffer (pH 5.0) at a concentration of 100 \(\mu\)g ml\(^{-1}\). On attaining stable resonant frequencies, each protein was pumped through the measurement chambers at a constant flow rate of 0.1 ml/min. The instrument recorded changes in both resonance frequency and dissipation of the sensor. The fundamental resonance frequency was 5 MHz, and the overtones 15, 25, 35, 45, 55, and 65 MHz were recorded.
3.3.8 Atomic Force Microscopy

SiO$_2$ was coated with 150 µg ml$^{-1}$ Dbfp7 in 0.1 M acetate buffer (pH 5.0) and placed into a 100% humidity environment for 24 hours to allow protein adsorption while preventing sample evaporation. Samples were dipped three times in 0.1 M acetate buffer (pH 5.0) and three times in MilliQ. All surfaces were rinsed with MilliQ in a squirt bottle to remove salt residue, and dried individually with argon. Atomic force microscopy (AFM) measurements were performed on a multimode, Nanoscope IIIa controller (Digital Instruments, Santa Barbara, CA) equipped with a fluid cell. Cantilevers were V-shaped silicon nitride with a tip radius of 5-40 nm and a spring constant of 0.06 N m$^{-1}$ as quoted by the manufacturer. All measurements were taken at an approach rate of 1.0 Hz.

3.4 Results & Discussion

Homogenization of mussel phenol glands reveals a selection of byssal proteins present within the soluble fractions of S1 and S2. Dbfp's as identified by Rees et al.[56] were abundantly present within the supernatants of S1 and S2, the latter of which Dbfp7 is the most prominent byssal protein. Each supernatant represents the proteins that were soluble in their respective buffer following homogenization and centrifugation. Prevalent in S1 are higher molecular weight proteins which include previously known $D$. bugensis foot proteins Dbfp0, Dbfp1, Dbfp2, and several novel proteins identified in a previous colleague’s research (Figure 3.1). S2 contains many of the same proteins with an emphasis on novel proteins found at 40 kDa and 50 kDa. Dbfp7, presenting as a doublet at approximately 6 kDa, was present both in a previous colleagues gel of artificially induced quagga mussel byssal proteins (Figure 3.1b) as well as S2 (Figure 3.1a). S2 was carried forward for further purification to isolate Dbfp7 which appeared as a distinct doublet, identical in appearance and intensity to the Dbfp7 doublet in Figure 3.1b and in high intensity relative to other proteins within S2.
Figure 3.1 – PAGE gels of separated *D. bugensis* proteins. (a) Sequentially isolated fractions (Supernatants S1, S2 of the original homogenized aliquot of *D. bugensis* phenol glands in extraction buffer (5% acetic acid with protease inhibitors). Gel was run on 12% Bis-Tris gel and visualized with Coomassie blue. The band of interest is highlighted in red. (b) Induced quagga mussel byssal visualized with silver stain, performed by a colleague. Protein bands marked as novel indicate quagga mussel proteins not previously identified by researchers who had isolated crosslinked, Dopa-containing byssal proteins.

### 3.4.1 Phenol gland purification of mussel adhesive proteins

S2 separation via reverse-phase (RP) high performance liquid chromatography (HPLC) revealed clear separation of Dbfp7 from larger molecular weight proteins (Figure 3.2b). Despite the abundance of larger molecular weight proteins present in S2, HPLC elution products were within the 6-18 kDa range with the sole byssal proteins eluting at 32-36% acetonitrile. No other proteins, byssal or otherwise, at a molecular weight higher than 18 kDa were present within any HPLC fractions even with an acetonitrile/water gradient spanning from 0-100%, indicating the lower relative hydrophobicity of this 6-18 kDa family of proteins. HPLC peaks shown in Figure 3.2 were the only significant peaks at either 220 nm or 280 nm. Elution fractions from 24-26 min, the only fractions presenting peaks with this method, were shown for the purpose of confirming the absence of proteins eluting outside the 38-48 min time frame.
Figure 3.2 – RP-HPLC chromatogram of partially purified quagga mussel phenol gland extract Supernatant-2 (S2). (a) RP-HPLC chromatograms were detected at 220 nm (blue line) and 280 nm (black line). A gradient of acetonitrile was employed. Highlighted regions in yellow represent fractions that were further analyzed by gel electrophoresis. Fractions pooled from 38-48 min correspond to Dbfp7 elution. (b) SDS-PAGE of representative column fractions in (a) coinciding with HPLC peaks. Ladder with stained protein standards is located in the left lane with MW standards whose apparent molecular masses are given in kDa. Numbers above the lanes correspond to elution time and fraction number in (a). Sections marked as A and B represent the same sections highlighted as A and B in (a).

3.4.2 Seasonal Purification

Quantification of Dbfp7 purification from the mussel phenol gland correlates negatively with time elapsed from mussel removal from its native conditions (i.e. Lake Ontario). Monitoring Dbfp7 protein amounts obtained from consistent extractions over time through chromatographic peak integration and comparisons with mussel lifespan within the artificial water environment reveals that Dbfp7 purification efficiency is maximal within a 24-hour period after removal of the mussels from their natural environment. Represented in Figure 3.3 are quantifications of Dbfp7 purification efficiency per extraction (25 mussels) decreasing linearly with each hour between mussel transfer and mussel death in the process of purifying Dbfp7.
Figure 3.3 – Dbfp7 purification efficiency is maximal within 24 hours of mussel retrieval from native conditions. (a) Chromatographic peak area integration versus time between mussel obtainment and mussel detachment for Dbfp7 extracted directly from the phenol gland and separated by HPLC; (b) for Dbfp7 extracted directly from the phenol gland, pre-purification with a 50 kDa syringe filter, and separated by HPLC.

It is important to note that although the time between mussel transfer and mussel death is a primary contributor to the sharp decrease in Dbfp7 - despite identical purification procedures, mussel health, and mussel size - a large change in ambient water temperature experienced by the mussel may affect mussel metabolism and thus the regulation of byssal protein production. However, comparisons of extractions presented in Figure 3.3 represent proteins obtained during the winter (water temperature = 4°C) versus the summer (water temperature = 22°C). Extraction quantifications shown in Figure 3.3a represent the first time in the laboratory that mussels were collected during the winter, leading to speculation that perhaps elevated Dbfp7 levels were a result of sudden removal from frigid ambient temperatures to spring-like temperatures. However, unlike their zebra mussel cousins, quagga mussels continue feeding throughout the winter, reducing the probability of a scenario where mussels produce or receive cues to ramp up and elevate production of byssal proteins, hypothetically in preparation for adhesive behaviour upon detection of warmer waters. Optimal temperatures exist for freshwater mussel spawning and larval development; however, tests of zebra mussel adaptability to water temperature changes indicate that mussels can tolerate rapid increases of up to 10°C and rapid decreases of up to 15°C before mortality occurs. Most significantly, extractions performed on mussels retrieved from both warm and cool waters demonstrate this distinct decrease in Dbfp7 recuperation from the phenol gland. It would be expected that a decrease merely due to the cold-to-warm water transition would preclude evidence of this same trend in mussels obtained from warm waters.

In addition, represented in both Figure 3.3a and b are a linear decrease in hours following mussel removal and detachment from their natural substrate of choice in native waters, representing an increased likelihood of a marked increase in protein obtained in the first 24 hours not only from the change in native water environment, but also from manual detachment of the byssus. All mussels collected for lab experiments had been previously steadfastly attached by many threads and plaques to a substrate (plastic ladder, metal chain, etc.) It is likely that the mussel recognizes the detachment of its long-held byssus, which is approximated to be 6 months+ at the very shortest. Byssal protein production may then correlate to the response of mussel cues which incite the mussel to produce byssal or adhesive
proteins in preparation for the search for the next anchoring location or substrate, as freshwater mussels attach to any stable substrate in the water column or benthos at water speeds of less than 2 m s$^{-1}$[131]. Experiments are undergoing to determine these multiple contributing factors, including the time from removal of the mussels from their native waters, the presence or absence of a tank circulation pump to mimic the presence of turbulent waters better mirroring native mussel conditions, and whether mussels which have adhered to the tank surface on their own (i.e. are actively producing adhesive byssal proteins) will have higher amounts of Dbfp7.

### 3.4.3 Size and secondary structure of native Dbfp’s

Eleven Dbfp7 variants were identified in LC-MS/MS performed on fractions purified directly from the phenol gland and separated via RP-HPLC. Coverage of peptides present strongly mirrors identical fragments shown in naturally secreted footprint and plaque proteins. Fragment identities and their predicted molecular weight can be used to infer an average molecular weight of proteins present within the purified protein sample, presenting a more accurate summary of true Dbfp7 sequence composition that has been demonstrated extensively in Chapter 2. To better understand downstream characterization of phenol gland-derived Dbfp7, molecular weight fragments were analyzed following fragmentation as identified through LC-MS/MS (Table 3.1), values which were consistent with banding shown in SDS-PAGE gels of purified proteins (6-8 kDa) as opposed to the theoretical molecular weight indicated by the full Dbfp7 sequence (10-14 kDa).

**Table 3.1 –** Theoretical molecular weights of full Dbfp7 sequences and N- and C-terminal fragments found in Dbfp7 purified natively from quagga mussel phenol glands. Fold change ratio data in relative quantification experiments of footprint versus plaque samples are also shown.

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Footprint:Plaque Fold Change Ratio (Naturally Secreted)</th>
<th>MW (kDa)</th>
<th>N-Fragment (kDa)</th>
<th>C-Fragment (kDa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dbfp7 seq1</td>
<td>N/A</td>
<td>14.9</td>
<td>5.9</td>
<td>4.1</td>
</tr>
<tr>
<td>Dbfp7 seq3</td>
<td>2.1</td>
<td>9.7</td>
<td>6.6</td>
<td>2.3</td>
</tr>
<tr>
<td>Dbfp7 seq6</td>
<td>3.6</td>
<td>12.0</td>
<td>8.0</td>
<td>2.3</td>
</tr>
<tr>
<td>Dbfp7 seq7</td>
<td>N/A</td>
<td>12.6</td>
<td>6.6</td>
<td>4.1</td>
</tr>
<tr>
<td>Dbfp7 seq8</td>
<td>N/A</td>
<td>12.3</td>
<td>5.4</td>
<td>4.1</td>
</tr>
<tr>
<td>Dbfp7 seq9</td>
<td>1.9</td>
<td>13.2</td>
<td>6.3</td>
<td>4.1</td>
</tr>
<tr>
<td>Dbfp7 seq10</td>
<td>N/A</td>
<td>12.7</td>
<td>5.9</td>
<td>4.1</td>
</tr>
<tr>
<td>Dbfp7 seq11</td>
<td>2.8</td>
<td>9.4</td>
<td>5.4</td>
<td>2.3</td>
</tr>
<tr>
<td>Dbfp7 seq13</td>
<td>0.7</td>
<td>10.3</td>
<td>6.3</td>
<td>2.3</td>
</tr>
<tr>
<td>Dbfp7 seq14</td>
<td>1.9</td>
<td>9.9</td>
<td>5.9</td>
<td>2.3</td>
</tr>
<tr>
<td>Dbfp7 seq15</td>
<td>N/A</td>
<td>11.1</td>
<td>7.1</td>
<td>2.3</td>
</tr>
</tbody>
</table>

In Dbfp7 samples purified directly from the phenol gland, Dopa was not detected in PTM analysis as performed with LC-MS/MS. Dopa residue values shown in Figure 3.4 represent Dopa identification from PTM analysis of naturally secreted footprint and plaque samples shown in Chapter 2. Dopa concentrations as identified through amino acid analysis (Table 3.2) of combined eluted fractions of
Dbfp7 are 0.17 mol%, on par with the levels of Dopa seen in quagga mussels overall (0.1 mol%) and Dbfp1 (0.6 mol%)\cite{51}. Dbfp proteins also demonstrate much lower Dopa residual composition values than Dpfp proteins (from Dreissena polymorpha, the zebra mussel).

<table>
<thead>
<tr>
<th>Seq3 (Dbfp7ζ)</th>
<th>LVLLVSCSSTP...</th>
<th>pI = 8.66</th>
<th>DOPA = 4 (4%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq6 (Dbfp7ε)</td>
<td>TPLGKWDPYGSSG</td>
<td>pI = 5.22</td>
<td>DOPA = 2 (2%)</td>
</tr>
<tr>
<td>Seq9 (Dbfp7γ)</td>
<td>TPLGKWDPYSY---</td>
<td>pI = 4.85</td>
<td>DOPA = 4 (3%)</td>
</tr>
<tr>
<td>Seq11 (Dbfp7ε)</td>
<td>TPLGKWDPYG---</td>
<td>pI = 8.83</td>
<td>DOPA = 1 (1%)</td>
</tr>
<tr>
<td>Seq13 (Dbfp7ε)</td>
<td>TPLGKWDPYG---</td>
<td>pI = 7.77</td>
<td>DOPA = 1 (1%)</td>
</tr>
<tr>
<td>Seq14 (Dbfp7ε)</td>
<td>TPLGKWDPYG---</td>
<td>pI = 8.80</td>
<td>DOPA = 2 (2%)</td>
</tr>
</tbody>
</table>

Figure 3.4 – Select (6/11) amino acid sequences of phenol gland-derived Dbfp’s shown with corresponding pI and content of Tyr (Y in red). Sequences represented are those found with a footprint:plaque fold change ratio present in relative quantification experiments.

Table 3.2 – Amino acid analysis performed on Dbfp7 purified natively from quagga mussel phenol glands.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Residue Weight</th>
<th>Count (pmol)</th>
<th>Mol% (residues/100)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASX</td>
<td>115.09</td>
<td>50.4</td>
<td>7.1</td>
</tr>
<tr>
<td>GLX</td>
<td>129.11</td>
<td>53.1</td>
<td>7.5</td>
</tr>
<tr>
<td>SER</td>
<td>87.08</td>
<td>60.9</td>
<td>8.5</td>
</tr>
<tr>
<td>GLY</td>
<td>57.05</td>
<td>153.6</td>
<td>21.5</td>
</tr>
<tr>
<td>HIS</td>
<td>137.14</td>
<td>18.2</td>
<td>2.6</td>
</tr>
<tr>
<td>ARG</td>
<td>156.19</td>
<td>42.1</td>
<td>5.9</td>
</tr>
<tr>
<td>THR</td>
<td>101.11</td>
<td>18.0</td>
<td>2.5</td>
</tr>
<tr>
<td>ALA</td>
<td>71.09</td>
<td>32.3</td>
<td>4.5</td>
</tr>
<tr>
<td>PRO</td>
<td>97.11</td>
<td>47.2</td>
<td>6.6</td>
</tr>
<tr>
<td>DOPA</td>
<td>179.17</td>
<td>1.2</td>
<td>0.2</td>
</tr>
<tr>
<td>TYR</td>
<td>163.17</td>
<td>44.8</td>
<td>6.3</td>
</tr>
<tr>
<td>VAL</td>
<td>99.13</td>
<td>22.9</td>
<td>3.2</td>
</tr>
<tr>
<td>MET</td>
<td>131.19</td>
<td>26.5</td>
<td>3.7</td>
</tr>
<tr>
<td>CYS</td>
<td>103.14</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>ILE</td>
<td>113.15</td>
<td>18.1</td>
<td>2.5</td>
</tr>
<tr>
<td>LEU</td>
<td>113.16</td>
<td>68.4</td>
<td>9.6</td>
</tr>
<tr>
<td>PHE</td>
<td>147.18</td>
<td>20.3</td>
<td>2.8</td>
</tr>
<tr>
<td>LYS</td>
<td>128.17</td>
<td>34.4</td>
<td>4.8</td>
</tr>
</tbody>
</table>

MALDI-TOF was performed on phenol gland-derived Dbfp7 with representative spectra from all fractions of the largest HPLC peak using sinapinic acid (SA) as the matrix, demonstrating a distinct family
of proteins within the 6-8 kDa range consistent with MALDI-TOF spectra on induced plaque and naturally secreted footprint samples [56]. SA was chosen over \( \alpha \)-Cyano-4-hydroxycinnamic acid (CHCA) as indicated by previous MALDI-TOF experiments performed on quagga mussel threads, plaques, and adhesive interfaces demonstrating increased background noise and preferential ionization of smaller proteins (<5 kDa) [56]. Spectra observed with SA, as expected from previous MALDI-TOF data of quagga mussel thread/plaques, showed more distinct peaks with higher relative intensity across proteins >5 kDa. To elucidate the differences in molecular weight distribution across individual HPLC fractions and the entire eluent from S2 purification via HPLC, MALDI-TOF was performed on combined purified products as well as individual 1 mL eluted fractions (Figure 3.5a and b, respectively). A detailed breakdown of the most prominent molecular weight families per fraction is shown in Figure 3.6. The intent of MALDI-TOF in this context was to determine the breakdown of polymorphs within the HPLC fractions in Figure 3.5. For example, if earlier fractions contained higher relative amounts of a 6 kDa protein versus a 7 kDa protein, the fractions would be pooled separately for independent downstream characterization.

**Figure 3.5** – Representative MALDI-TOF MS spectra of purified native Dbfp7 demonstrating the absence of consistent Dbfp7 polymorph separation in separate fractions by RP-HPLC. (a) MALDI-TOF spectra of all fractions pooled together, purified from RP-HPLC from 4000 to 10,000 kDa. The main molecular weight families represented are within the 6-8 kDa region. (b) Representative MALDI-TOF spectra of one fraction purified from RP-HPLC from 4000 to 20,000 kDa.

MALDI of combined fractions demonstrate high relative intensity of peaks in the 6-8 kDa range, with greatest abundance at 6.8 kDa. Individual fractions do not demonstrate consistent molecular weight distribution among time-resolved elution fractions which is expected for RP-HPLC separation by hydrophobicity. SDS-PAGE gels shown in Figure 3.6c show an increase in abundance of the lower molecular weight band of the Dbfp7 doublet, which is not represented in the distribution of MALDI-TOF fractions by molecular weight in Figure 3.6a. Given the unclear separation of different Dbfp7 polymorphs by RP-HPLC and the lack of a particular molecular weight family distinct among individual time-resolved fractions, further Dbfp7 characterization was performed on combined fractions from an entire HPLC injection.
Figure 3.6 – (a) Relative percent intensities of four most consistently present families of protein sizes in kilodaltons (kDa). (b) Representative HPLC spectra indicating fractions isolated from RP-HPLC for MALDI-TOF analysis. Each division indicates a fraction that was analyzed by MALDI-TOF, numbered for the time at which it was eluted, in minutes (44-53).

Purified Dbfp7 proteins presented with distinct SDS-PAGE banding between 6-8 kDa (Figure 3.2) and MALDI-TOF peaks in the 6-9 kDa range (Figure 3.6). The 6-8 kDa fragments are explained by the N-terminal fragments being found in relative quantification experiments to display disparate abundance levels between the footprint (adhesive interface) and bulk plaque. Identical fragmentation is present in LC-MS/MS analysis of natively purified Dbfp7 protein. The presence of purified Dbfp7 presenting as a doublet with distinct molecular weight distribution in the 6-8 kDa range indicates that this hypothesized cleavage of Dbfp7 proteins occurs well before natural mussel secretion of Dbfp7 - in fact, Dbfp7 is already cleaved when it is situated within the secretory granules of the mussel foot phenol gland in preparation for secretion, as indicated by the presence of distinct 6-8 kDa SDS-PAGE bands after purification by HPLC. This may prevent undesired non-specific protein binding interactions that may trigger undue inactivation of Dbfp7’s adhesive capability. SDS-PAGE slightly overestimates the molecular weight of the bands, with banding appearing at a 7-13 kDa range despite MALDI data showing distinct banding no higher than 8.5 kDa. This is likely due to deamidations present on each variant of Dbfp7 present in the purified extract. As the most ubiquitous PTM appearing multiple times on each variant, the higher molecular weight due to the addition of the -NH₂ group as well as altered folding and interaction with SDS-PAGE may result in a subtle increase in molecular weight in an interpretation solely based on SDS-PAGE gels of the purified protein.

LC-MS/MS bottom-up analysis of purified native Dbfp7 proteins indicates that both the N- and C-terminal fragments are secreted. Although the C-terminal fragment sizes range from 2.3-4.1 kDa, there is limited observation of this fragment in MALDI-TOF and SDS-PAGE. MALDI-TOF spectra of an
individual Dbfp7 fraction (Figure 3.5) reveals a peak of moderate intensity in the 4.1 kDa range. MALDI should be repeated from 0 kDa to reveal the presence of the 2.3 kDa peak. The absence of SDS-PAGE banding in the 2-4 kDa range can be attributed to the very small size of these peptides, affecting the gel fixing process and permitting them to be washed out of the gel during staining.

CD spectra for Dbfp7 in solution exhibited a mix of predominantly random coil and β-sheet structures, with minima between 200 and 205 nm as well as a slight increase leading to a plateau from 230-260 nm (Figure 3.7). Deconvolution with DichroWeb[130, 132] and the CONTIN-LL (Provencher & Glockner method)[133, 134] reveals Dbfp7 secondary structure breakdown as 6.5% α-helix (intra-chain hydrogen bonds), 39% β-sheets (inter-chain hydrogen bonds), 22% turns, and 33% unordered. Mean residue ellipticity (Δε) was used to represent ellipticity in Figure 3.7 under the assumption that both N- and C-terminal fragments of Dbfp7 are monomeric and do not aggregate under buffer conditions. LC-MS/MS of purified Dbfp7 proteins indicates that both N- and C-terminal fragments (average of 6.3 kDa and 3.1 kDa respectively) are present within mussel secretory granules in preparation for secretion. However, the much higher relative intensity of the 6-8 kDa molecular weight family in MALDI performed on purified Dbfp7, the absence of the 2-4 kDa fragments within SDS-PAGE through both Coomassie and silver staining, and the overwhelmingly higher abundance of the 6-8 kDa SDS-PAGE bands indicate that in spite of the narrow MALDI observation range and potential for peptide loss during fixation steps of SDS-PAGE, the 6-8 kDa fragment most strongly dominates in abundance. It can be inferred that the sample of Dbfp7 analyzed by CD primarily consists of the N-terminal fragment, which has an average molecular weight of 6.3 kDa across all eleven purified Dbfp7 variants identified through LC-MS/MS. This value was used in tandem with the average number of amino acid residues across Dbfp7 variants (N = 52.7) to determine the best mean residue weight (MRW) approximation of N-terminal Dbfp7 fragments, where MRW = molecular weight in Daltons / (number of residues−1). The calculated MRW for the N-terminal fragment of Dbfp7 is 121.9, which roughly approximates the typical mean residue of weight of 115 for most proteins of unknown molecular weight[135].

![Figure 3.7](image-url)  
**Figure 3.7** – CD spectra of 500 µg ml⁻¹ Dbfp7 in 20 mM acetate buffer showing mean residue ellipticity [θ] as a function of wavelength.

The conditions used for CD spectra collection were 20 mM acetate buffer (pH 5.0). Interference with solvent absorption in the UV region required non-absorbing buffers with low salt concentrations to permit measurements below 200 nm. Thus, for the purpose of maintaining the buffer at a low enough concentra-
tion to maintain acceptable high-tension values, Dbfp7 was re-suspended in an acetate buffer that was five times less concentrated than that used by other mussel protein secondary structure analyses[129, 136]. Low pH was selected to prevent oxidation of Dopa residues within solution. These conditions are a departure from physiological conditions, thus making it difficult to draw final conclusions on native conformations of Dbfp7 from this data. Investigation of Dbfp7 secondary structure in aqueous conditions mirroring those of optimal freshwater mussel growth (pH 7.4 to 8.0) will better represent mussel adhesive activity within physiological conditions[131]. However, this presents a substantial challenge due to the propensity for Dopa to oxidize at neutral-to-basic pH levels, and thus the loss of Dopa adhesive functionality.

### 3.4.4 Adsorption studies of Dbfp’s on SiO$_2$, TiO$_2$ and Au

AFM images show aggregated Dbfp7 evenly dispersed across the surface of the wafer, but critically do not demonstrate a consistent density of adsorption on the surface. SiO$_2$ was selected as the substrate for AFM imaging despite adsorbed thicknesses of protein on TiO$_2$ and Au demonstrating higher average thickness, as AFM optimization trials with TiO$_2$-coated substrates showed dense deposition on the surface on both protein-coated and non-protein coated surfaces, making it difficult to ascertain between artifacts and true Dbfp7 deposition. Figure 3.8 indicates that the height of the aggregated protein on the SiO$_2$ surface is approximately 15 nm. Positively, there are clear aggregates of expected protein size on the silicon wafer, demonstrating successful albeit inconsistent deposition onto the silicon surface. AFM aggregates on the section of the chip shown in Figure 3.8 range from 15-35 nm.

**Figure 3.8** – (a, b) Representative AFM images of 150µg ml$^{-1}$ Dbfp7 in 0.1 M acetate buffer (pH 5.0) adsorbed on a SiO$_2$-coated silicon wafer. (c) AFM image of untreated silicon wafer.

QCM measurements were performed for the first time on byssal proteins isolated from *D. bugensis*, permitting in-situ, real-time characterization of Dbfp7 adsorbed layers at the substrate-interface level. This technique has been increasingly used in elucidating biological materials and their interactions due to its capacity to simultaneously measure changes in resonant frequency ($\Delta f$) - changes in mass adsorbed

---

**Figure 3.8** – (a, b) Representative AFM images of 150µg ml$^{-1}$ Dbfp7 in 0.1 M acetate buffer (pH 5.0) adsorbed on a SiO$_2$-coated silicon wafer. (c) AFM image of untreated silicon wafer.
to the sensor surface, and energy loss or dissipation ($\Delta D$, dimensionless units, $\times10^{-6}$) of the system -
properties related to the structural, viscoelastic, and thin-film properties of the adlayer[137]. Generally,
a decrease in frequency is proportional to an increase in mass while an increase in dissipation correlates
to softer films. QCM-D measurements upon adsorption of Dbfp7 to SiO$_2$, TiO$_2$, and Au sensors followed
by rinsing in a pure buffer solution are displayed in Figure 3.8. The simultaneously-measured changes in
frequency and energy dissipation were obtained at the fundamental frequency in addition to its overtones
- $n = 1$ (MHz), $n = 3$ (15 MHz), and $n = 5$ (25 MHz) - as a function of time.

Figure 3.9 – (a) $\Delta f$ and $\Delta D$ for QCM-D measurements performed with Dbfp7 adsorption onto
SiO$_2$, TiO$_2$, and Au sensors. (b) Evolution of the change in the effective thickness, obtained from
the best fit between the viscoelastic model (obtained using Q-Soft), and the experimental data.

A cursory comparison of Dbfp7 films on each surface suggests that the most material was deposited on
Au sensors, and that the adsorbed film on Au is less rigid than on SiO$_2$ or TiO$_2$. Upon addition of
Dbfp7, mass uptake was observed on all surfaces, reaching a plateau at approximately 1 hour into each
experiment. This timeframe is representative of QCM-D experiments performed on mussel proteins such
as *Mytilus edulis* foot protein (Mefp)[129]. Experiments were extended to 24+ hour timepoints to ensure
consistency among all surfaces and to establish a confident indication of a plateau, representative of a
consistent layer formed by Dbfp7 on the sensor surface. Frequency drift of approximately 0.05 Hz per hour beyond reaching a plateau was observed on all surfaces. This observation is normal for QCM-D experiments and is likely caused by ambient temperature fluctuations, while simultaneous frequency and dissipation drift (not present in the data) may be due to viscosity changes as a result of evaporation from the solvent mixture.

The Sauerbrey equation represents a linear relationship between resonance frequency of an oscillating crystal and mass changes\[138, 139\]. Estimating thickness and mass adsorbed onto the crystal with the Sauerbrey relationship requires that the surface, approximated with a thicker crystal, must be similar to the quartz itself. The model assumes that the added layer can be approximated to the oscillating crystal itself - i.e. it is thin, rigid, and firmly attached to the crystal surface\[138, 139\]. Large changes in dissipation upon Dbfp7 adsorption across all surfaces indicate that the adsorbed film is not rigid as $\Delta D > 0$; thus the above criterion for usage of the Sauerbrey equation are not met and cannot be used as frequency dampening will occur, resulting in an overestimation of the mass of the attached layer\[138, 139\]. The adhering film is then characterized by applying a Voigt viscoelastic model, which does not assume that the mass couples 100% to the oscillatory motion of the sensor.

The thicknesses of adlayers adsorbed onto SiO$_2$, TiO$_2$, and Au are $14.8 \pm 2.6$ nm, $9.9 \pm 0.2$ nm, and $18.3 \pm 0.9$ nm respectively. Following rinsing of the sensor with pure buffer solution, thicknesses on each surface decreased by negligible amounts on Au (-0.02 nm) and SiO$_2$ (-0.6 nm), while the thickness of layer loss on TiO$_2$ was unobtainable via analysis with the viscoelastic model. Masses of adlayers onto SiO$_2$, TiO$_2$, and Au are 1.6 $\mu$g, 1.2 $\mu$g, and 0.9 $\mu$g respectively. Thicknesses of adlayers represented in QCM data closely approximate measurements performed on AFM for all measurements. Masses and thicknesses approximated by QCM-D give the hydrated mass, which includes the mass of molecules adsorbed in addition to the solvent trapped in between. Dissipation curves in Figure 3.9 reveal high dissipation with insignificant spreading across overtones with Dbfp7 on SiO$_2$ and TiO$_2$, indicating that Dbfp7 has likely arranged itself in a dense and ordered fashion while adsorbed in solution, resulting in little solvent trapped between layers. Conversely, Au demonstrates much higher dissipation values with significant spreading of overtones across each harmonic. Thus, adsorption of Dbfp7 onto Au substrates potentially maintains a more extended and sparse structure. Addition of crosslinking agents such as NaIO$_4$ to induce di-Dopa formation, or Cu$^{2+}$ and Mn$^{2+}$ to induce metal complex formation with Dopa, may reveal changes in viscoelasticity if Dopa plays a role in Dbfp7 adsorptive activity. Differences between thicknesses abstracted from AFM and QCM-D may also be attributed to differing concentrations and exact Dbfp7 protein composition between experiments. AFM data was collected on Dbfp7 purified in a separate extraction process from CD and QCM-D data. Minor differences in exact composition of Dbfp7 fragments and variants in tandem with variations in concentration, particularly techniques which impose adsorption loss via surface characterization, may lead to amplification of differences between models inferring surface thickness.

Isoelectric points (pIs) of QCM sensor surfaces greatly influence surface activity with adsorbed material. Gold has been reported to have pI = 4.5\[140], 5\[141], or 4-6\[142], with increased surface oxidation influencing a shift towards a more basic pI\[142]. Au QCM sensors in particular have been shown to have pI = 5.2\[143]. SiO$_2$ pI values range from 3.5-5.6 depending on the form measured, from fused silica glass\[144], native oxides on a Si wafer\[145], or for thick SiO$_2$ films grown by plasma-enhanced vapor deposition\[145]; however, SiO$_2$ QCM sensors have been shown to exhibit surface charge behavior with a pI of 3.9 which strongly approximates those of Si wafers with a native SiO$_2$ layer\[143]. Oxidized films
Chapter 3. Dbfp7 Purification and Characterization

of TiO$_2$ QCM sensors demonstrate a pI of 2.9\cite{143}, while pure Ti exhibits a basic pI\cite{143}. It has also been demonstrated that net charge at the sensor surface can produce short-range interactions between the sensor surface and adlayer which impede efficient packing of the protein layer\cite{143}. Film densities of proteins such as mucins have been shown to achieve maximal densities when the solution pH is close to the corresponding sensor pI. This has been theorized to be due to short-range interactions in the form of electrostatic repulsion as a result of net charge on the sensor surface, thus preventing efficient packing of protein layers\cite{143}.

Although eleven variants of Dbfp7 are present within purified protein sample, the fragments composing comp52765_seq14 were chosen as representative sequences due to their individual N- and C-terminal fragment pIs (pI = 9.26 and 3.93 respectively) most closely approximating the average pI of all eleven N- and C-terminal fragments, at pI = 9.26 and 3.95 respectively. Peptide and spectral counts for comp52765_seq14 were also among the highest of the eleven fragments. Presence within a buffer of pH 5.0 imparts a net charge onto Dbfp7 fragments in addition to the QCM sensor substrate surface, the predicted charges of which are shown in Table 3.3. QCM-D data indicates that layer thickness and film softness are both highest on Au samples, which is not consistent with the mucin packing patterns observed by Cuddy et al, who noted that a buffer pH (5.0) closely matching the sensor pI (for Au, 5.2) would mitigate electrostatic repulsion between the sensor surface and adlayer\cite{143}. Following this observation, although this coalescence of a denser Dbfp7 film on Au was not observed in our QCM-D experiments, the net charge of the N-terminal fragment within a buffer of pH 5.0 imparts a charge onto this dominant fragment. The oxide layers of the SiO$_2$, TiO$_2$ and Au sensor surfaces also have a net charge imparted as a result of both the surrounding pH of the bulk electrolyte in addition to the ionic strength. It is possible that the positively-charged and abundant N-terminal fragment undergoes electrostatic repulsion with Au, the only positively charged substrate surface, leading to an increasingly extended conformation that lends Dbfp7 to a “softer” viscoelastic film which, when adsorbed, imparts a higher thickness on the sensor surface. This phenomenon is not observed on SiO$_2$ and TiO$_2$ to the same extent, potentially due to the positively charged N-terminal fragments interacting attractively with negatively charged substrate surfaces, resulting in a flatter orientation.

**Table 3.3** – Dbfp7 fragment and substrate surface net charge at experimental pH conditions (pH 5.0).

<table>
<thead>
<tr>
<th>N-terminal Fragment</th>
<th>C-terminal Fragment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fragment pI</td>
<td>9.26</td>
</tr>
<tr>
<td>Fragment charge at pH 5.0</td>
<td>+3.2</td>
</tr>
<tr>
<td>SiO$_2$ (pI = 3.9)\cite{143}</td>
<td>Surface negatively charged at pH 5.0</td>
</tr>
<tr>
<td>TiO$_2$ (pI = 2.9)\cite{143}</td>
<td>Surface negatively charged at pH 5.0</td>
</tr>
<tr>
<td>Au (pI = 5.2)\cite{143}</td>
<td>Surface positively charged at pH 5.0</td>
</tr>
</tbody>
</table>

Future QCM experiments require a larger sample size of results which are under consistent conditions which were unattainable in this preliminary characterization of Dbfp7. Due to limitations of protein availability, the same solution was used across multiple QCM-D experiments and surfaces, leading to long-term, gradual protein loss through sensor and tubing adsorption. Improved optimization of experimental conditions will permit collection of consistent data that is better represented across a large sample size.
LC-MS/MS experiments performed on purified Dbfp7 proteins indicate approximately equal spectral counts for both N- and C-terminal fragments. The average pI for both fragments differs widely, with the N-terminal fragment significantly more basic (pI = 9.2) compared to the C-terminal fragment (pI = 3.9). Although SDS-PAGE and MALDI of purified proteins provide a preliminary indication of increased abundance of N-terminal fragments within purified protein fraction, it would be prudent to separately isolate the N- and C-terminal fragments in order to determine each fragment’s interactions with sensors, particularly since the N-terminal fragment has been found to be significantly more enriched at the footprint (adhesive interface) of naturally secreted proteins.

Although frequency change normalized by molecular weight is proportional to the number of protein molecules adsorbed into the sensor, and thus a measurement of the affinity to a certain surface, analysis of the density on the surface of the sensor allows for differentiation between the presence of more molecules compared to the presence of an extended layer, both of which contribute to the measured thickness on the sensor surface. Knowledge of density of the hydrated layer by the means of a tensiometer in tandem with the QCM-D technique measured at multiple harmonics would reveal thickness, shear viscosity, and shear elastic modulus data.

Although the net charge on the N-terminal fragment is positive at pH 5.0, the positive charges are distributed evenly among the protein (Figure 3.10). It is significant that the net charge of +3.2 is primarily due to singly-present positive charges introduced among many neutral residues. Thus, Dbfp7 presence on negatively-charged surfaces such as SiO₂ and TiO₂ may result in this N-terminal fragment undergoing preferential electrostatic attraction to the negatively charged surface. Conversely, repulsive electrostatic interactions with Au, a surface that is slightly positively charged at pH 5.0, may encourage the N-terminal fragment to self-assemble into an extended confirmation, particularly with the positive charge at the C-terminal end of the N-terminal fragment. Acting in tandem with the repulsive interactions between Dbfp7 and Au are the attractive electrostatic interactions occurring among the positively charged residues of Dbfp7 and the comparatively negatively charged SiO₂ and TiO₂ surfaces by virtue of departure from their native isoelectric points. It can be speculated that Dbfp7 maintains a “laid-down” formation on the surface of the sensor as opposed to a brush-like structure, owing to the formation a more rigid and thinner film in comparison to an Au-adsorbed layer.

![Figure 3.10](image_url) – Dbfp7 N-terminal fragment charge distribution in solution at pH 5.0.

The average diameter of the N-terminal fragment of Dbfp7 (6.3 kDa) is 2.4 nm. At this predicted diameter, it appears that at its greatest thickness (18.3 nm on Au), there are 7.6 layers of Dbfp7 molecules. This prediction was made on a model of the partial specific volume calculated to be occupied by a protein of mass \( M \) in Dalton, and thus the assumption of the minimum diameter of a smooth sphere.
that could contain the mass of a given protein\cite{146}. However, this is commonly an underestimate as proteins largely present with an irregular surface; thus, even those that appear approximately spherical will have a larger radius (and thus, height as adsorbed on a surface)\cite{146}. It is also important to note that protein adsorption on the surface of a substrate may not necessarily cause the protein to maintain a spherical or globular structure, thus further compounding thickness and layer predictions made on the basis of spherically-assembled Dbfp7.

3.5 Conclusion

Mussel adhesives have the ability to attach to substances including plastic, glass, metals, and Teflon\cite{28}, while its adhesive proteins successfully bind to living body substances such as porcine skin\cite{40} and mammalian cells\cite{29, 41, 42}. This renders the freshwater mussel an excellent model organism for developing a biomaterial glue that is biocompatible, biodegradable, and easily manipulated for use in medical and dental applications. Interest in the adhesive bonding of sessile freshwater organisms, such as mussels, has long been investigated in marine species, but their freshwater counterparts have not seen the same attention until recently. This work has been a significant step forward in determining putative candidates for novel freshwater mussel adhesive proteins responsible for this activity, as well as establishing methods to purify quagga mussel proteins of interest, and contributes to the significant progress made by our group in understanding freshwater mussel adhesion.

For the first time, a quagga mussel protein (Dbfp7) demonstrating potential for adhesive capability has been purified natively from the mussel phenol gland. Significant progress has been demonstrated in the optical and surface-sensitive characterization of Dbfp7. It has been shown to maintain a primarily $\beta$-sheet and unordered conformation in solution. Its adsorptive qualities differ by the surface on which it attaches. Thicknesses on SiO$_2$ demonstrate inconsistency likely due to the need for optimization of adsorption protocols and the inconsistency among protein concentrations and sample composition between AFM and QCM-D experiments. Dbfp7 adsorption patterns on surfaces of relatively increased hydrophobicity (TiO$_2$ and Au) demonstrated consistent thicknesses between multiple surface-sensitive experiments. Differences in adsorption thickness (ranging from 9-18 nm) may be the result of differing interactions of Dbfp7 with electrostatically charged surfaces. Dbfp7 amino acid composition presents with 29% content containing amino acids with ionizable R groups which alter the net charge of Dbfp7 upon exposure to different pHs. The sensitivity of Dbfp7’s net charge in tandem with the pI of the tested surfaces (SiO$_2$, TiO$_2$, and Au) is a speculative theory of Dbfp7 interaction onto surfaces.

These novel methods for purifying phenol gland-derived quagga mussel proteins are a strong step forward in the existing knowledge of the purification and analysis of standalone freshwater mussel adhesive candidates. This methodology can be extrapolated for use in the isolation of alternate Dbfp proteins and serve as a starting point in the purification of Dreissena polymorpha (zebra mussel) proteins. This large breadth of secondary structure, molecular weight, and adsorptive studies serve as valuable groundwork for future kinetic and thermodynamic analyses of Dbfp7-surface interactions, the study of which is imperative in the assessment and characterization of an adhesive protein candidate.
3.6 Acknowledgements

The authors gratefully acknowledge Paul Taylor and Jonathan Krieger of the SickKids Proteomics, Analytics, Robotics Chemical Biology Centre (SPARC BioCentre) at the Peter Gilgan Centre for Research and Learning (PGCRL) at the Hospital for Sick Children in for LC-MS/MS and relative protein quantification analysis and guidance. We also thank Reynaldo Interior for amino acid analysis sample analysis and guidance. We thank Prof. Walid Houry, Thiago Vargas Seraphim, and Elisa Leung for access, training, and guidance of CD facilities. We thank Prof. Emma Master and Thu Vuong for access, training, and guidance regarding QCM-D. We thank Leonardo Ermini and Hayley Craig-Barnes of the SickKids Analytical Facility for Bioactive Molecules (AFBM) at the Hospital for Sick Children for MALDI-ToF sample analysis and guidance. We thank Prof. Jayachandran Kizhakkedathu and Kai Yu for access, training, and guidance of ellipsometry and AFM equipment.
4.1 Future Work & Recommendations

It can be inferred from the evidence of separate N- and C-terminal fragments of Dbfp7 existing within mussel secretory granules prior to natural secretion of byssal threads and plaques, and the relative enrichment of the N-terminal fragment at the adhesive interface, that purification and characterization of the N-terminal fragment (average MW = 6.3 kDa) without larger molecular weight Dbfp’s present as indicated by silver staining (14-17 kDa) and LC-MS/MS indicating the presence of the C-terminal fragment would be fruitful. It is likely that multiple electrophoretic variants (at least 11) of Dbfp7 may be expressed by individual mussels as a response to different surface compositions, which has been established in Mcfp-3[11]. More finely-tuned separation protocols can be applied in future experiments to facilitate separation by other parameters, such as isoelectric point via two-dimensional gel electrophoresis and affinity to ion exchangers via ion exchange chromatography. This purification of isoelectric point as an exploitative strategy is particularly relevant to the N- and C-terminal fragments, which significantly differ in pI (9.26 and 3.93 respectively). Reverse phase high performance liquid chromatography (RP-HPLC) has demonstrated promising ability to separate Dbfp7 from other Dbfp’s within the sample. Inclusion of ion exchange chromatography (IEX) as a separate step in column purification may result in adsorptive protein loss in fraction collection and column binding. Thus, mixed-mode high performance liquid chromatography techniques combining reverse-phase and ion-exchange column - porous spherical silica particles with covalently-modified silyl ligands - can be used to isolate the N-terminal fragment (cation exchange column) versus the C-terminal fragment (anion exchange column)[147].

4.1.1 Dbfp7 Structural Characterization Optimization

Circular dichroism has been used extensively to investigate marine mussel protein structure and conformation in conditions. A vast selection of electrolyte solutions (e.g. NaCl, KF, GuCl) covering large pH ranges have been investigated; however, Small Angle X-ray Scattering (SAXS) experiments may suggest
β-turn and helix conformations in further detail. To provide a better indication of the perturbations to secondary structure between Dbfp7 in an aqueous versus adsorbed state, CD can be performed on quartz slides in the wet state to mimic adhesion between substrates in underwater conditions, the dry state to mimic the use of the adhesive in a biomedical context, or in an aqueous state more closely representing the conditions in which Dbfp7 is purified. CD performed on Dbfp7 in the “wet state” and “dry state” refer to methods of obtaining surface-immobilized Dbfp7 samples; in the former, by sandwiching Dbfp7 in solution between two quartz slides, and in the latter, by allowing Dbfp7 in solution to adsorb to each side of a quartz slide, and recording CD spectra of the quartz surfaces.

A breadth of parameters can be altered in the recording of CD spectra to monitor for conformational changes: temperature, denaturants, heat, or binding interactions[135]. The temperature-controlled cuvette holder allows for collection of spectra at precise temperature increments, followed by supplementary measurement of protein activity (e.g. enzyme activity, ability to bind to ligands). Presentation of CD spectra data in units of mean residue ellipticity ensures that spectral interpretation does not vary as a function of protein concentration or path length, unless Dbfp7 is not monomeric or undergoes aggregation at changing concentrations. The oligomeric state of Dbfp7 can be determined with ultracentrifugation, DLS, or native gradient electrophoresis[135]. Although CD spectra allow for determination of the secondary structure of the entire protein, X-ray crystallography and NMR permit determination of the secondary structure of specific residues[135]. Surface plasmon resonance (SPR) techniques, useful for their ability to detect the binding of molecules to a surface in real time without the use of labels[148], can be used to determine aggregation rates of Dbfp7 in solution and the effect of pH changes on multilayers formed on the initially adsorbed layer.

Expansions on techniques used in this study also serve to broaden the understanding of Dbfp7 structure and characterization. MALDI-TOF of purified Dbfp7 proteins should be performed from 0-20 kDa as lower molecular weight proteins are more likely to be visualized via MALDI as opposed to SDS-PAGE. Generating MALDI-TOF spectra in this low molecular weight range with α-Cyano-4-hydroxycinnamic acid is recommended due to its propensity to ionize peptides less than 10 kDa[149]. Tricine-SDS-PAGE of purified Dbfp7 proteins and pre-purified extracts (S1 and S2) in lieu of traditional tris-glycine gel systems permits superior separation of proteins from 1 to 20 kDa[150].

MALDI-TOF has been used extensively to assess the relative intensities of peaks from naturally secreted and artificially induced quagga mussel threads, plaques, and footprints[56]. Although these studies were hampered by limitations in minor crosslinking and sample contamination (e.g. with thread and plaque sample overlap), MALDI-TOF was an exceptionally powerful technique to isolate molecular weight families of interest at the adhesive interface. A separate technique known as time-of-flight secondary ion mass spectroscopy (TOF-SIMS) allows for inferences of the orientation of the protein, its state of unfolding, or whether the protein has denatured upon surface adsorption[148], in addition to identifying and distinguishing contributions from multiple types of proteins at the surface[151]. Via bombardment with ions, short peptides are removed from the top nanometer of an adsorbed protein layer in order to determine the composition of the top layer. Although MALDI-TOF has been used to assess purified Dbfp7 proteins in solution, sample preparation for MALDI-TOF requires airdrying of aqueous Dbfp7 prior to analysis, which does not have significant consequences for assessing molecular weight, but influences protein conformation. The ToF-SIMS technique is hampered by the limitation of the inability to relate traditional airdried MALDI-TOF results to the original, hydrated conformation of adsorbed proteins. Xia et al. developed a preservation technique of preparing adsorbed protein samples with
trehalose[148], known for its ability to inhibit protein folding during water removal and stabilizing protein conformation by forming hydrogen bonds to polar residues in the protein[152, 153]. Thus, phenolic-purified aqueous Dbfp7 proteins analyzed for MALDI-TOF need not be limited by potentially denaturing or conformation-modifying activity that may occur upon adsorption. TOF-SIMS of trehalose-preserved and unpreserved purified Dbfp7 samples can be compared to provide structural information of Dbfp7 adsorption. Conformational behaviour of Dbfp7 is likely to differ between experimentation performed in aqueous and adsorbed states, particularly due to its role as an interfacial adhesive and the utility of Dbfp7’s adhesive qualities and interactions in the solid state.

4.1.2 Dbfp7 Surface-Sensitive Characterization Optimization

Surface-sensitive methods such as QCM-D and AFM were used extensively in this study. To supplement these results, bulk-sensitive methods for aggregate characterization such as DLS to determine the hydrodynamic radius of Dbfp7 in tandem with SAXS experiments will generate high resolution results to investigate intra- and intermolecular aspects Dbfp7 assembly and potential factors controlling crosslinking rates of quagga mussel adhesive proteins. Dynamic light scattering (DLS) would also provide Dbfp7 monomer diameters in solution with excellent precision, which would provide a useful benchmark against Dbfp7 thicknesses and mass contribution of aggregates as predicted by adsorption behaviour via QCM-D and AFM. Light scattering measurements of Dbfp7 in free solution surrounded by different ion types may reveal changes in hydrodynamic radius, as has been demonstrated in Mefp-1[154]. Changes in the effect of salt type and concentration on Dbfp7 activity may suggest that byssal ionic composition is tailored by the mussel to achieve maximum adhesion and minimum curing times[155].

Kinetic and thermodynamic studies indicate that conformational changes may occur as a protein adsorbs to the surface[128]. Similar experiments of Pvfp proteins showed dominance of β-sheets and random coils when adsorbed on TiO$_2$ with a slow transition to β-sheets over time[32], rationalized by the activity of adsorption inciting an ordered structure in the form of β-sheets that may arise to stabilize the adsorbed structure, an observation further bolstered by Attenuated Total Reflection Infrared Spectroscopy (ATR-FTIR). A wide spectrum in deconvolution breakdown of Dbfp7 CD spectra among these contexts may reveal that Dbfp7 hardly has a secondary structure at all, which is consistent with Mefp-1 proteins.

Understanding the secondary structure content of surface-tethered Dbfp7 also lends information to interpretation of QCM-D data analysis. The Voigt viscoelastic model obtains adlayer thickness from the change in resonant frequency ($\Delta f$) upon interaction with the protein solution. Mass adsorbed onto the sensor is also extrapolated from the best-fit data of changes in resonant frequency obtained from the viscoelastic model. This leads to one weakness of the inability to use the Sauerbrey model to fit QCM-D data - due to the lack of rigidity in the adsorbed thin film, interpreting mass change with thickness is ambiguous without information regarding the density of the adsorbed film. An identical thickness of an adsorbed layer between two surfaces may not necessarily represent the same mass of protein adsorbed onto the surface, as thickness is not only a function of mass but also the orientation and conformation of the protein upon adsorption. Density data of the hydrated layer provides information on the layer in addition to viscosity and shear patterns. CD data of surface-immobilized proteins, particularly in the wet state, also contributes to an understanding of Dbfp7 activity upon adsorption to a QCM-D sensor. Convolution revealing a much lower contribution of unordered protein conformation and a concomitant increase in a contribution from α-helical or β-sheet structures may represent a significant difference in
protein activity upon adsorption to a surface.

Changes in pH conditions for surface-sensitive characterization methods performed with ellipsometry and QCM-D may yield information regarding how differences in Dbfp7 net charge and surface net charge at increased or decreased buffer pH influence Dbfp7 thickness and viscoelasticity upon adsorption. A caveat of increasing pH is that Dopa exhibits a propensity to autoxide and irreversibly crosslink at neutral-to-basic pH conditions\[45, 156, 157\]. Thus, incorporation of a Dopa-protecting group such as cyclic ethyl orthoformate (Ceof) prevents oxidation of the catechol hydroxyl groups\[158\]. Other methods of protecting groups for Dopa include shielding by hydrophobic amino acids\[136\] or temporary Fe\(^{3+}\) complexation\[25\].

QCM-D experiments which introduce a crosslinking agent in order to assess dissipation and surface adsorption changes upon interaction have been performed on Mefp-1 proteins\[32\]. Similar experimentation with autooxidation methods including NaIO\(_4\), other oxidants, or enzymatic methods (e.g. tyrosinase) on Dbfp7 will reveal changes in structure and conformation of Dbfp7 on the adlayer patterns in the infrared spectrum: amide I (~1620-1680 cm\(^{-1}\)) and amide II (~1520-1580 cm\(^{-1}\)). Infrared spectra differences upon adsorption to different surfaces (e.g. polystyrene or poly(o-methoxyaniline)) may reflect the influence of substrate film chemistry on Dbfp7 structure.

Atomic force microscopy (AFM) is a powerful technique to acquire direct measurements of adhesion between protein deposits and substrate surfaces. AFM has been used extensively to investigate whole mussel adhesive protein interactions with surfaces\[rank2002, 159\]. Possible deformation of the cantilever tip, unknown area of contact, and unknown source of contact forces have led to functionalizations of AFM tips with beads, Lee et al. utilized AFM to study interactions of a single Dopa residue in interaction with a metal oxide surface, as whole mussel adhesive interactions unavoidably include interactions from multiple amino acids and residues, and an unknown number of proteins on the AFM tip\[45\]. AFM has also been used extensively in the study of barnacle adhesion to examine both the morphology of the attachment structures within native footprints\[160\], the adhesive strength between the AFM tip and the footprint proteins on R-NH\(_3\) and R-CH\(_3\)-terminated glass\[160\], and adhesive force measurements of AFM probes functionalized with footprint deposits and their interaction with surfaces of differing wettability\[161\]. AFM could be used to observe naturally secreted quagga mussel footprint morphology, and to perform further analysis of Dbfp7 adsorption on surfaces via functionalization of the AFM tip with Dbfp7 and self-assembled monolayer formation of aqueous Dbfp7.

Repeats of Dopa quantification of Dbfp7 may reveal variability in Dopa concentration due to differences in the level of post-translational tyrosine hydroxylation. Identification of cumulative concentrations of Dopa and tyrosine at the onset of sample preparation will indicate if variability of Dopa concentrations in Dbfp7 are a result of post-translational hydroxylation occurring as a natural response by the mussel, or by oxidation of this residue during purification and handling. Identification of Dopa residues beyond LC-MS/MS post-translational modification detection via Arnow staining, UV spectroscopy, and nitro blue tetrazolium (NBT) staining will reveal evidence of Dopa among purified Dbfp7 samples. Staining for Dopa of pre-purified homogenates (S1 and S2) will give clues to Dopa-containing byssal proteins.
4.1.3 Recombinant Expression of Dbfp7

Select recombinantly purified Dbfp7 fragments have been successfully purified by a colleague in our group in an \textit{E. coli} expression system. This method allows for much higher protein collection and cost efficiency due to the highly scalable nature of bacterial expression systems. Native Dbfp7 purification directly from the quagga mussel phenol gland, though the most accurate method and that which is most representative of native Dbfp7 proteins, is an intensely laborious process which yields approximately 250 µg of Dbfp7 per 25 mussels after homogenization and purification via HPLC - although this value is highly dependent upon the amount of time that has passed between mussel collection and phenol gland dissection. An additional caveat of recent experiments is the absence of Dopa within the sequence fragment as a consequence of the inability to incorporate post-translational modifications in bacterial expression systems. A future goal is to compare recombinantly purified Dbfp7 protein activity with CD, QCM-D, ellipsometry, and AFM data against natively purified Dbfp7. A lack of Dopa will be a consideration in protein activity; however, its absence permits reaction conditions that may not be typically possible without sample modification in natively purified Dbfp7, such as neutral to high pH conditions.

4.1.4 Expansion of Quagga Mussel Byssal Protein Knowledge

Tremendous progress has been made in our group to identify the sequences of byssal proteins produced by the quagga mussel in addition to their spatial distribution in naturally secreted plaque and footprint structures. However, little remains known about the distribution of the foot tissues responsible for the synthesis and storage of Dbfp7. The paucity of data in this regard has been addressed within \textit{D. polymorpha}, in which Anderson & Waite\cite{162} raised polyclonal antibodies to a recombinant version of \textit{Dreissena polymorpha} foot protein 1 (Dpfp1)\cite{162}. Sections of zebra mussel foot exposed to these antibodies revealed Dpfp1 detection in a subset of byssal gland cells along the mussel ventral groove in addition to extracellular granules found within the ventral groove\cite{162}. The generation of specific antibodies to quagga mussel byssal proteins, particularly the N- and C-terminal fragments of Dbfp7 in addition to other byssal proteins projected to be involved in byssal thread synthesis, could be performed in similar immunolocalization studies. Cross-reactivity between polyclonal antibodies raised against foot-derived mytilid byssal precursors and other \textit{Mytilus} species byssal proteins has been demonstrated\cite{163} in addition to cross-reactivity between homologous proteins from other species\cite{111}. An investigation into quagga mussel byssal protein antibodies in this context would be of great interest. Lastly, isolation and histochemical tests of quagga mussel byssal gland cells for tyrosine and phenols would provide further evidence of Dopa-containing byssal precursor proteins within the mussel foot and phenol gland, as has been performed in zebra mussels\cite{6}.

4.2 Significance & Conclusions

Mussel adhesives have the ability to attach to substances including plastic, glass, metals, and Teflon\cite{28}, and its adhesive proteins successfully bind to living body substances such as porcine skin\cite{40} and mammalian cells\cite{29, 41, 42}. This renders the freshwater mussel an excellent model organism for developing
a biomaterial glue that is biocompatible, biodegradable, and easily manipulated for use in medical and dental applications. Interest in the adhesive bonding of sessile freshwater organisms, such as mussels, has been long been investigated in marine species, but their freshwater counterparts have not seen the same attention until recently. This work has been a significant step forward in determining putative candidates for novel adhesive proteins responsible for this activity, as well as in establishing methods to purify quagga mussel proteins of interest, and adds on to the significant progress made by our laboratory in understanding freshwater mussel adhesion.

Proteins composing the quagga mussel plaque body were compared to proteins produced at the adhesive interface. Extensive proteomic analysis and data on localization of proteins at the adhesive interface in quagga mussels was crucial to narrowing down a number of putative proteins involved in adhesion, to our top candidate. Purification of this novel protein was successfully achieved, guided by protocols for freshwater mussel proteins that had been purified in the past[12, 52]. Surface-sensitive and optical techniques provided preliminary data concerning adsorption of Dbfp7 as well as structural information in solution and on surfaces. Further investigation of Dbfp7 and byssal protein interaction with surfaces by studying changes in the thickness of an adsorbed layer over time, or the effect of surface chemistry on the nature of the adsorbed layer, are relevant questions that may yield information regarding the tenacity of this adhesive layer in nature. This information can be used in the design of both biomaterials for adhesive use in addition to the rational design of antifouling coatings which deter mussel attachment.
References


REFERENCES


REFERENCES


REFERENCES


