Identification of Collagen IV Associated Proteins in Drosophila Using Genetics and Mass Spectrometry

Mayank S. Kapadia
Western Kentucky University, mayank.kapadia692@topper.wku.edu

Follow this and additional works at: http://digitalcommons.wku.edu/theses
Part of the Biochemistry Commons, Cell and Developmental Biology Commons, and the Molecular Biology Commons

Recommended Citation
http://digitalcommons.wku.edu/theses/1631

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
IDENITIFICATION OF COLLAGEN IV ASSOCIATED PROTEINS IN DROSOPHILA USING GENETICS AND MASS SPECTROMETRY

A Thesis
Presented to
The Faculty of the Department of Biology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science

By
Mayank S Kapadia
August 2016
IDENTIFICATION OF COLLAGEN IV ASSOCIATED PROTEINS IN DROSOPHILA USING GENETICS AND MASS SPECTROMETRY

Date Recommended 07/21/16

Dr. Ajay Srivastava, Director of Thesis

Dr. Sigrid Jacobshagen

Dr. Michael Smith

Dean, Graduate Studies and Research Date 07/28/16
ACKNOWLEDGMENTS

First, I would like to acknowledge Dr. Ajay Srivastava, my thesis advisor for the time and resources that he invested towards the completion of this research. I am grateful for his patience, consideration, and ability to carefully guide me in the right direction. It is his guidance that allowed this project to occur and result in a fruitful outcome.

I also would like to thank the other members of my thesis committee, Dr. Sigrid Jacobshagen and Dr. Michael Smith. These individuals have challenged me to the core to become a better biologist and have been an excellent source of information throughout this program and for that I want to thank them. Additionally, special thanks to Dr. Jessica Ross for her continued support in training and advising me with various techniques used during the research.

I also like to extend my thanks to Naomi Rowland of the WKU Biotechnology Center for training and allowing me to use various instruments within the facility. Furthermore, I would like to thank my lab mates for all the time spent together and for cooperation in troubleshooting problems. I would also like to thank all the members of the Biograds who had arranged various trips and events to ease the stress and creating unforgettable memories. Thank you to Jessica Dunnegan, Melanie Redden, and the rest of the biology office staff for always having such a positive attitude.

Finally, I want to thank my family and friends for their support and continuous encouragement throughout my years of study and the process of earning my Master’s degree.

This work was funded by KBRIN-AREA grant funded through a parent grant from the National Institute of General Medical Sciences of the National Institutes of
Health (5P20GM103436-13), RCAP 1-S grant (14-8050), Western Kentucky University Biology Department, and WKU Graduate Research Grants.
# TABLE OF CONTENTS

1. Introduction ................................................................................................................................. 1
   1.1 Overview of Basement Membrane .......................................................................................... 1
   1.2 BM Major Components ........................................................................................................ 3
   1.3 Components Associated with BM during Development ..................................................... 11
   1.4 Biogenesis of Collagen IV .................................................................................................... 14
   1.5 Surf-4 and its Role ................................................................................................................ 15
   1.6 Diseases Associated with BM Components ........................................................................ 16
   1.7 *Drosophila* as a Model Organism ...................................................................................... 17
   1.8 Purpose of the Study ............................................................................................................ 20

2. Materials & Methods .................................................................................................................... 21
   2.1 Fly Stocks ............................................................................................................................ 21
   2.2 Isolation Buffers ................................................................................................................ 21
   2.3 Stock Buffers ....................................................................................................................... 22
   2.4 Freshly-made Buffers ........................................................................................................ 22
   2.5 Gels and Membranes Used in the Study ............................................................................ 23
   2.6 Antibodies Used in the Study ............................................................................................ 23
   2.7 Protein Staining Solutions .................................................................................................. 24
   2.8 Sample Preparation ............................................................................................................ 24
   2.9 Protein Purification ............................................................................................................... 24
      2.9.1 Isolation and detection of proteins ............................................................................ 24
         2.9.1.1 RIPA buffer extraction process ............................................................................ 25
         2.9.1.2 Lysis buffer extraction process .......................................................................... 26
         2.9.1.3 2X Laemmli buffer extraction process ............................................................... 26
   2.10 Cross-linking ..................................................................................................................... 26
2.11 Co-immunoprecipitation ................................................................. 27
2.12 Coomassie Staining ......................................................................... 28
2.13 Mass Spectrometry and Data Analysis ............................................. 28
2.14 Protein Trap Studies ......................................................................... 28
3. Results ........................................................................................................ 31
   3.1 Selection of Isolation Buffer ............................................................... 31
   3.2 Standardizing the Formaldehyde Cross-linking Conditions .............. 33
   3.3 Co-immunoprecipitation of Cross-linked Protein Complex ............... 36
   3.4 Mass Spectrometric Analysis of Complexes Containing Collagen IV .... 38
   3.5 Surf-4 and its Localization in Drosophila Third Instar Larvae ............. 48
   3.6 Surf-4 RNAi Knockdown Study ............................................................ 50
      3.6.1 Lac Z Staining .............................................................................. 51
   3.7 Surf-4 Over-expression Study .............................................................. 57
4. Discussion ...................................................................................................... 59
   4.1 Future Directions ............................................................................... 63
      4.1.1 In-vivo Protein Knockdown Studies ............................................ 63
      4.1.2 Characterization Studies for other Mass Spectrometry Candidates .... 64
      4.1.3 Repetition of the Mass Spectrometry Analysis .............................. 64
5. References .................................................................................................... 65
LIST OF FIGURES

Figure 1. Basement membrane structure and composition........................................... 2
Figure 2. Type IV collagen self-assembly. .................................................................. 5
Figure 3. Structure of Laminin and Nidogen/Entactin............................................. 8
Figure 4. Multiple interacting partners of HSPGs. .................................................. 10
Figure 5. Lifecycle of Drosophila melanogaster....................................................... 19
Figure 6. Schematic representation of the experimental setup. ............................... 30
Figure 7. Comparison of proteins isolated using RIPA and Laemmli buffers........... 32
Figure 8. Cross-linking using different concentrations of formaldehyde. ............... 34
Figure 9. Co-immunoprecipitation results of the Actin-GFP and Viking-GFP cross-linked samples................................................................. 37
Figure 10. Coomassie stain on co-immunoprecipitated samples............................... 39
Figure 11. Expression pattern of Surf-4 and Collagen IV in different tissues of Drosophila melanogaster................................................................. 49
Figure 12. Defective phenotypes when Surf-4 is knocked down............................... 54
Figure 13. Defective phenotypes when Surf-4 is knocked down............................... 55
LIST OF TABLES

Table 1. Comparison of the mass spectrometry result between Actin-GFP and Viking-GFP .......................................................................................................................... 40

Table 2. Proteins found common in both rounds of mass spectrometry analysis and unique to Viking-GFP sample .............................................................................................................. 46

Table 3. Generated phenotype description when Surf-4 (BDSC-57471) is knocked down using different RNAi drivers. .................................................................................................................. 52

Table 4. Phenotype description when Surf-4 (VDRC-GD5883) is knocked down using different RNAi drivers. .............................................................................................................................. 56

Table 5. Phenotype description when Surf-4 (19202) is overexpressed using Gal4 drivers. .................................................................................................................................. 58
Metastatic cancer cells invade and spread to other locations by disrupting the basement membrane (BM). The membrane plays a major role during the normal development of an organism as well. In order to understand the invasion mechanism it is important to know about the interactions occurring between the proteins of the BM during normal development. This study concentrates on isolating and identifying the major factors associated with collagen IV, a major component of BM, during the third instar larval development of *Drosophila*. Western blot and mass spectrometry analysis revealed that collagen IV associates with various growth factors, signaling molecules, and proteins that may play a role during the development of *Drosophila*. Co-localization and knockdown studies performed on a single protein found through mass spectrometry suggested a possible role of this protein in the development of *Drosophila*. Further analysis of this proteins’ function will provide new insights into its developmental role and its potential role in collagen IV transport.
1. INTRODUCTION

1.1 Overview of Basement Membrane

Basement membrane (BM) is a thin membrane-like structure located at the basal side of endothelial or epithelial cells [1, 2]. It acts as a substratum for the cells to adhere on and perform various cellular functions such as migration, proliferation, and differentiation [2]. It is 50-100nm thick and composed of large insoluble materials connected with each other to form a dense mesh-like structure [3]. Major proteins that are involved in the formation of this dense structure include type IV collagen, perlecan (a heparan sulphate proteoglycan), laminin, and entactin/nidogen [1, 2] (Figure 1). Type IV collagen accounts for approximately 50% of the overall molecular mass of the BM [4]. Laminin and type IV collagen each form their own suprastructure, which are then bridged together by entactin/nidogen and perlecan to form a stable and dense sheet-like structure. Other components that are also associated with BM include collagen XV, agrin, collagen XVIII, BM90, and secreted protein acidic and rich in cysteine (SPARC) [4]. Though BM found in various tissues consists of similar structure, molecular composition of minor components makes it unique at various locations within an organism [4, 5].
Figure 1. Basement membrane structure and composition  The figure provides a graphical view of the interaction of the BM and cells. The magnified view shows the mesh-like network formation with major components of the BM. (Modified from http://www.bioon.com/book/biology/mboc/mboc.cgi?action=figure&fig=19-56.htm).
1.2 BM Major Components

**Type IV collagen:** Type IV collagen is a non-fibrillar 540-kDa protein with unique repeating sequence of Gly-X-Y; where X and Y represents hydroxyllysine and hydroxyproline, respectively. The protein self-assembles to form a network-like structure providing BM its stability [4, 6]. Its structure consists of an α-chain polypeptide with three distinct parts - an amino terminus called 7S domain, a triple helical domain with Gly-X-Y sequence, and a carboxyl terminus called NC1 (non-collagenous) domain. The central triple helical domain consists of approximately 1,400 amino acids with 22 classical Gly-X-Y sequences while the NC1 domain is made up of 230 amino acids. The formation of network-like structure initiates with three monomer α-chains coiled to form a trimer called a Protomer (Figure 2). In mammals, six different types of α-chains (α1-α6) have been identified so far, which can form trimers in 56 different combinations [3]. These combinations have similar domain structure and show 50–70% homology in their amino acid sequence [7]. Two protomers then associate with each other through their NC1 domain to form dimers, which in turn form tetramers through linking of the 7S domain. These interactions form the nucleus for a type IV collagen scaffold. Through end-to-end and lateral associations between tetramers a type IV collagen suprastructure is formed [3].
Monomer (Single \(\alpha\)-chain)

Triple-helical domain (Gly-X-Y motif)

NC1 domain

\[\sim 230 \text{ aa} \quad \sim 1,400 \text{ aa}\]

NC1 interactions promote type IV protomer formation

Protomer (a trimer of \(\alpha\)-chains)

Dimer

Type IV collagen tetramer

7S domain

NC1 hexamer

7S domain
Figure 2. Type IV collagen self-assembly. The figure shows a graphical representation of Type IV collagen protein. The protein is made up of α-chain polypeptide consisting of three domains – an amino terminal 7S domain, a triple helical domain with Gly-X-Y repeats, and a carboxyl-terminal non-collagenous (NC)-1 domain. The formation of collagen IV network initiates with three α-chains forming a trimer, also called Protomer, by connecting their NC1 domains. Two type IV collagen protomers then connect with each other through their C-terminal to form Dimers (NC1 hexamers). The next step involves the interaction of the glycosylated amino-terminal 7S region of the four protomers to form tetramers. These interactions form the nucleus of a type IV collagen, which then evolves into a type IV collagen suprastructure by end-to-end and lateral associations of the protomers. (Modified from [3]).
**Laminin:** Laminin is a non-collagenous protein with a size of 850-kDa [8]. Its structure consists of three polypeptide chains forming three short arms and one long arm. The arms are labeled α, β, and γ based on the sequence and protein domain organization. A total of eleven genes code for the eleven chains of laminin (α1-5, β1-3, and γ1-3). The α-chain is 400-kDa in size, while β- and γ-chains are 200-kDa in size each [5, 8, 9]. The laminin structure resembles a three-pronged fork, with C-termini of all three arms/ chains acting as a handle for the fork [4, 10] (Figure 3). The α-chain C-terminus of laminin is 865-900 residues longer than the β- and γ-chains due to the presence of the laminin globular domains, also called G domains [9]. The presence of G-domain at the C-terminus helps laminin to interact with the proteins of the plasma membrane of cells; while the short arms at the N-termini are involved in interactions with other proteins of the BM [9]. Figure 3 shows various interacting sites on laminin for various proteins. *In vitro* analysis has shown that the arms interact in a domain-specific manner to form a network structure. It has also been found that collagen IV and laminins contain specific information within their amino acid sequence which guide them to initiate the self-assembly process [6]. Aumailley and colleagues showed that laminin can also interact directly with collagen IV via its short and long arms but these interactions are of low-affinity [9]. During the formation of BM, collagen IV acts as a scaffold while laminin forms the centerpiece with entactin/nidogen acting as a bridge connecting the two proteins [11, 12].

**Entactin/Nidogen:** Entactin is a 150-kDa tyrosine-sulphate glycoprotein. Studies of recombinant entactin/nidogen have revealed that the protein consists of three G domains (G1-G3). The G1 and G3 domains represent the N-terminus and C-terminus,
respectively; while G2 domain separates the two termini by two rod-like structures on each side [10, 12] (Figure 3). G2 domain consists of five cysteine-rich repeats with the presence of an RGD (Arg-Gly-Asp) sequence that promotes cell attachment. The C-terminal globular domain of entactin/nidogen strongly binds to laminin at the vicinity of the two arms and also interacts with collagen IV [1, 13].
Figure 3. Structure of Laminin and Nidogen/Entactin. The laminin short arm (N-terminus) is involved in interaction with various BM proteins; while the long arm (C-terminus) is involved in interaction with cellular receptors. The C-terminus of nidogen/entactin interacts with the γ-chain of laminin (via G3) and also with collagen IV (via G2), forming a bridge between the two network proteins.
**Heparan Sulphate Proteoglycans:** Heparan sulphate proteoglycans (HSPGs), 400-450kDa in size, are glycoproteins with one or more covalently attached heparan sulphate (HS) glycosaminoglycan (GAG) chains [10]. There are three subfamilies of HSPGs, which include membrane-spanning proteoglycans (syndecans, betaglycan, and CD44v3), glycosphatidylinositol-linked proteoglycans (glypicans), and extracellular matrix secreted proteoglycans (agrin, collagen XVIII, and perlecan) [14]. The presence of HS chains makes these glycoproteins very negatively charged and facilitates binding to a large number of proteins such as growth factors, receptor tyrosine kinases, chemokines and interleukins, enzymes and inhibitors, and ECM and plasma proteins [14]. HSPGs function by interaction of their HS/GAG chains with the signaling molecules or directly with different core receptor proteins. For instance, HS interacts with the fibroblast growth factors (FGFs) and their receptors resulting in a complex formation. This complex helps in lowering the concentration of FGFs required to initiate signaling and also in extending the duration of receptor response. In addition, HSPGs interact with cell-surface receptors like integrins to facilitate cell attachment, movement, and spreading. For example, fibronectin of ECM interacts with HS chains of syndecans and integrins to initiate cell movement and formation of focal adhesion [14–18]. Studies showed that all ECM proteins have binding domains for HS chains. For instance, HSPGs interact with laminin through nidogen/entactin to form a ternary complex [19]. Figure 4 provides a schematic representation of multiple interacting partners of HSPGs.
**Figure 4. Multiple interacting partners of HSPGs.** The figure shows HSPGs (Black S-shaped strings) interacting with laminin, entactin/nidogen or collagen IV through their HS chains (Yellow). It also interacts with receptor tyrosine kinase (a), integrins (b), and growth factors (c).
1.3 Components Associated with BM during Development

In addition to the conserved core components, BM is made up of various molecular components that differ between organisms, between tissues within an organism, and also with developmental age. These components perform various activities such as fill space between cells, act as a barrier between tissues, navigate migratory cells, provide signals to alter cell behavior, and sequester biologically active compounds such as growth factors [20].

A cell makes contact with the BM components through specialized receptor molecules present on its membrane. This contact allows the cell to adhere to the BM and in turn allows the BM to control the behavior of the cell [20]. It is also known that BM is constantly remodeled by the cells within and around it, by proteolytic enzymes, and through deposition or degradation of the BM components at various developmental stages. As a result, important changes in the cell–cell and cell–BM interactions occur generating new signals from the cell surface. This, in turn, affects gene expression and influences critical cellular behaviors such as proliferation, survival, differentiation, and motility [20, 21].

Integrins: Studies have found that cells bind to specific BM components using a variety of receptors present on their plasma membrane. These receptors include proteoglycans, lectins, and integrins [20]. For instance, α6β4 is an integrin–laminin receptor that anchors epithelium to the BM, forming a rigid structure called hemidesmosome [22, 23]. Integrins, when associated with the BM and a cell, bind with various signal-transducing molecules, including focal adhesion kinases, which on
activation phosphorylate various targets that control cell survival, differentiation, and proliferation [20].

**Proteoglycans:** Proteoglycans are found to be involved in promoting cell–BM adhesion. Most proteoglycans form hydrated gels by consuming water to fill space in the BM, and some, specifically the proteoglycan heparan sulfate, bind to a variety of growth factors, concentrating them in the BM and preventing their diffusion to other parts of the body [20]. Heparan sulfate chains also interact with proteins such as chemokines, morphogens, and enzymes [24]. Studies in *Drosophila*, mouse, and zebrafish have shown that the heparan sulfate chains on the HSPGs bind with various signaling molecules such as Fibroblast Growth Factors (FGFs), Hedgehogs (*Hh*), Wingless (*Wnt*) [25], Transforming Growth Factors β (TGFβs) [26], Hepatocyte Growth Factor (HGF) [27], and Heparin Binding-Epidermal Growth Factor (HB-EGF) [28]. It was also found that *Drosophila* contains genes coding for two glypican members, division abnormally delayed (*Dally*) and Dally-like (*Dly*) [29, 30]; one syndecan [31, 32]; and one perlecan [33]. Over-expression studies of glypican and syndecan in tissue culture cells restrict the response of cells to FGF [34, 35]. Perlecan, one of the largest members of the HSPG family, has been linked with signaling by heparin-dependent growth factors FGF2, Vascular Endothelial Growth Factor (VEGF), and Sonic Hedgehog (SHH) in the mammalian system [36, 37]. Perlecan is also involved in the maintenance of epithelial cell polarity by interacting with the BM receptor dystroglycan [38]. During the first instar brain development in *Drosophila*, perlecan’s homolog *trol* (terribly reduced optic lobes) regulates the activity of Hedgehog and Branchless (an FGF homolog) to control the onset of stem cell proliferation [24]. HSPGs are also linked with growth factor signaling
pathways such as wingless (wg/Wnt) and decapentaplegic (Dpp/TGF-β). Both of these pathways take actively part along with Hh and Ras-MAPK pathways in the development of *Drosophila* eye disc and/or second instar brain [39, 40]. In addition, HSPGs are responsible for regulating Dpp movement in *Drosophila* wings [37]. Studies by Jackson and his co-workers showed that Dally is required for normal Dpp signaling during imaginal disc development. Using a Dally mutant, they also showed that Dally is responsible for altering the response of a cell to Dpp [41].

**Collagen IV:** Collagen IV is found to be associated with Dpp and plays a vital role during *Drosophila*’s early embryonic development [42]. Dpp is a bone morphogenetic protein (BMP) signaling molecule that belongs to the TGF-β superfamily of growth factors. It is responsible for cell fates at different developmental stages of *Drosophila* and regulating the levels of Dpp signaling is vital during the development of *Drosophila*. For instance, gradients of Dpp form the anterior-posterior axis of the wing and proximal-distal axis of the leg [43, 44]. Similarly, alterations in the Dpp signaling may lead to several human diseases, including skeletal disorders, vascular diseases, and cancer [45, 46]. Collagen IV binds Dpp to inhibit the signaling of morphogen to the distant cells, thereby inhibiting the number of germline stem cells (GSC) [42, 47]. In an embryo, collagen IV promotes interaction of Dpp with heteromeric receptor complexes to form Dpp/Scw–Tsg-Sog complex. Disruption of this complex leads to reduced target gene expression and positive feedback, further decreasing subsequent signaling [42].

All these data indicate that the components of the BM play vital roles during *Drosophila* development.
1.4 Biogenesis of Collagen IV

The complete process of collagen IV, from production to its incorporation into the BM, is complex. The process starts with α chains undergoing extensive posttranslational modifications in the endoplasmic reticulum. Initially, prolines and lysines in the X and Y positions of the tripeptide repeats are hydroxylated to form hydroxyproline and hydroxylysines, respectively [48]. Hydroxilation of prolines at 3’ or 4’ position provides stability to the triple helix; while hydroxilation of lysine is followed by O-linked glycosylation in the ER. The triple helix is then properly folded and trimerized by chaperones and enzymes [49]. The post-translational modifications cause collagen IV protomers to be about 300nm long, which makes the transport of protein from ER to Golgi apparatus impossible into the COPII vesicles of 60-80nm in diameter. Studies performed in Drosophila showed that collagen IV is initially produced in soluble form and packaged into the COPII vesicle with the help of COPII cargo adaptor protein called TANGO1 (Transport And Golgi Organization) [50]. Once the protein reaches the Golgi apparatus it is further modified and transported out of the plasma membrane to form BM. Collagen IV accounts for 50% of the total BM mass and it is important that the protein is continuously supplied in the appropriate form. Hence, during the embryonic development of Drosophila, the requirement for collagen IV is fulfilled by hemocytes; however, the increasing need for the protein is fulfilled by the fat body during the later stages of development [51]. Literature review also showed that surf-4, an early secretory protein is involved in the transport of soluble proteins. This study provides with the insights on the role of surf-4 during the development of Drosophila.
1.5 Surf-4 and its Role

Three membrane organelles, namely, Endoplasmic Reticulum (ER), Endoplasmic Reticulum Golgi Intermediate Compartment (ERGIC), and Golgi apparatus constitute the early secretory pathway in higher eukaryotic cells [52]. Cycling proteins play an important role in the maintenance of these organelles by participating in proper trafficking of proteins in anterograde and retrograde directions. COPII vesicles are involved in anterograde mode of protein transport from ER to ERGIC and pleomorphic vesicles transport the cargo from ERGIC to Golgi apparatus. COPI vesicles perform the retrograde mode of transport from ERGIC or Golgi apparatus to the ER and they are mostly involved in the recycling of membrane proteins [52, 53]. Transport between these membrane organelles during the retrograde and anterograde modes is facilitated by transmembrane cargo receptors and inactivation of cargo receptors have been linked to several human diseases. For example, inactivation of ERGIC-53 leads to inefficient secretion of blood coagulation factors V and VIII resulting in provoking bleeding disorder. One of the cargo receptor that was identified in yeast to perform ER-to-Golgi apparatus transport is Erv29p. This receptor is required for efficient packaging of the glycosylated α-factor pheromone precursor into the COPII vesicles at the ER exit site [53]. The receptor is conserved among eukaryotes and the mammalian orthologue was found to be surf-4, a type of housekeeping protein. Surf-4 is part of the Surfeit locus which encodes six surf genes (Surf-1 to Surf-6). All six genes differ in their amino acid sequence and have unique features such as overlapping genes, bidirectional transcriptional promoter, and CpG islands at the 5’ end [54–56]. Drosophila melanogaster Surfeit genes are located on chromosome 3R. However, unlike in mouse
the Surf-4 gene in *Drosophila* is not in close proximity to other Surfeit genes but located upstream from a gene encoding a homolog of a yeast seryl-tRNA synthetase protein [57, 58]. The Surf-4 gene encodes an integral membrane protein of 30-kDa mass with dilysine motifs at its C-terminal making it an ER resident protein [59]. The protein consists of 270 amino acids in *Drosophila* and between 250-270 amino acids in mouse, human, yeast, and *C. elegans* and is conserved between these species [58]. Knockdown studies performed by Mitrovic and her colleagues [53] showed no major effect on total protein secretion suggesting that the protein acts as a cargo receptor for a specific set of proteins.

1.6 Diseases Associated with BM Components

According to the literature, each individual component of the BM plays a vital role during the development of an organism (*Drosophila*). Hence defects in the genes expressing these components might result in severe abnormalities in an organism. Some of the diseases due to defects in the BM components are as follows:

a) Alport syndrome: It is an inherited disorder due to mutations in the *COL4A4*, *COL4A5* gene. Mutation in any of these genes prevents the proper production/assembly of collagen IV network in the BM of kidney. This results in improper filtration of waste products from blood causing renal failure [60].

b) Knobloch syndrome: It is a rare autosomal recessive developmental disorder due to a mutation in the *COL18A1* gene. The mutation results in defects in collagen XVIII protein, which plays an important role in determining the retinal structure as well as closure of the neural tube. Mutations in the gene can also lead to occipital encephalocele and severe ocular alterations [61, 62].
c) Congenital muscular dystrophy: It is an autosomal recessive disorder due to a defect in the \( \textit{LAMA2} \) gene. Mutation in this gene results in defects in the \( \alpha2 \) chain of laminin. The complete and near-complete deficiency of \( \alpha2 \) laminin chain results in severe hypotonia at birth or within the first few months of birth [63].

d) Schwartz-Jampel syndrome: It is a rare autosomal recessive skeletal dysplasia associated with myotonia. It is caused by a mutation in the \( \textit{HSPG2} \) gene which encodes perlecan. The disorder results in short stature, osteochondrodysplasia, myotonia, and a characteristic face with a fixed facial expression, blepharophimosis, pursed lips, low-set ears, and myopia [64].

1.7 \textit{Drosophila} as a Model Organism

The overall goal of this research is to understand the mechanisms that cancer cells use to metastasize to different parts of the body. During the metastasis process, the Matrix Metalloproteases(MMPs) released by the cancer cells break the basement membrane at three different locations – one around its localized mass, another at the entry site into the blood stream, and finally at the exit site of the blood stream. As the current study is concentrated only on one protein of the BM, collagen IV, \textit{Drosophila melanogaster} proves to be a perfect model organism. \textit{Drosophila} has only two genes which code for collagen IV – \textit{Viking (Vkg)} and \textit{Collagen at 25C (Cg25C)} when compared to other model organisms like rats and mouse which consists of multiple genes coding for the same protein [65–67]. Knockdown studies of these genes in \textit{Drosophila} have displayed embryonic lethality and absence of the protein has resulted in aberrant shapes in several organs during the larval stage [65, 66]. This study uses transgenic flies consisting of \textit{Viking} gene tagged to \textit{GFP} gene which upon expression produces a GFP-
tagged collagen IV protein. In addition, *Drosophila* has a completely sequenced genome and significant homology with the human genome exists, making it an ideal organism for understanding human biology and disease processes. Further, nearly 75% of the disease-related genes in humans have functional orthologs in the fly [68, 69]. Though the overall identity of nucleotide sequence between the fly and mammal is only approximately 40%, functionally they are 80-90% identical [68]. The fly also has a rapid life cycle producing hundreds of genetically identical offspring within 10 to 12 days at 25°C (Figure 5). Each stage of the fly growth cycle; i.e., from embryo to the adult; can be used as a model system to understand the effects caused due to mutations. The study uses late third instar stage as it helps in studying the developmental and physiological processes. In addition, during the late third instar stage the wing imaginal disc undergoes immense morphological changes which at the later stage develops into an adult fly wing. [69, 70].
**Figure 5. Lifecycle of Drosophila melanogaster.** The lifecycle starts with adult flies mating and female flies laying eggs in the culture medium. The eggs undergo embryogenesis and hatch to form the first instar larvae in a day. The larvae continuously feed on the medium and molt to form second instar larvae in 24 hours. The process repeats for another 24 hours and again the larvae molt to produce third instar larvae. At this stage, larvae eat voraciously and finally enter the pupal stage where they undergo pupation for another 3-4 days. The pupae then metamorphose to produce adult flies on the 4th day of entering pupation and the lifecycle continues. Adopted from [http://morphologicallydisturbed.weebly.com/the-biology.html](http://morphologicallydisturbed.weebly.com/the-biology.html).
1.8 Purpose of the Study

Malignant tumors spread to different parts of the body by disrupting the basement membrane. To understand the mechanism involved in this disruption process, it is important to identify the makeup of BM. However, BM is a complex structure with continuous modification, remodeling, and degradation involving interaction of several components during the development of an organism [71]. Therefore to understand the invasion mechanism, it is important to first know what components interact to assist in the function of BM during normal development. This study uses Drosophila melanogaster’s late third instar larval stage to identify components/proteins that interact with type IV collagen of the BM through mass spectrometry analysis. Once these interacting partners and their relationship with collagen IV is defined, then similar studies would be performed in the cancer-bearing Drosophila to identify the interactive partners at the third instar larval stage. This has the potential to identify new basement membrane partners which may have a role in the cancer metastasis process.
2. MATERIALS & METHODS

2.1 Fly Stocks

Transgenic flies expressing Actin-GFP and Viking-GFP were obtained from the Srivastava lab stock collection and cultured at 25°C in the standard Drosophila cultured medium (LabExpress). Viking-GFP is a fusion protein which produces a GFP tagged collagen IV; whereas Actin-GFP produces a GFP protein wherever actin promoter is expressed. The study used transgenic Actin-GFP and Vkg-GFP third instar larvae for the isolation experiments. We also used transgenic flies for the Surf-4 gene tagged with GFP for protein trap studies. For knockdown and overexpression studies of the Surf-4 gene we used various drivers, RNAi lines, and over-expression lines. These lines were obtained from the Bloomington Drosophila Stock Center. Details of these fly stocks and their genotypes can be found in Tables 3 and 5.

2.2 Isolation Buffers

The following isolation buffers were used:

1. RIPA (50mM Tris HCl, pH 8.0; 150mM Sodium chloride; 1% NP40; 0.5% Sodium deoxycholate; 0.1% SDS; 1mM EDTA; and 1 tablet Roche’s Protease inhibitor cocktail/10ml solution) [72]

2. Lysis buffer (0.5M Urea, 0.01% SDS, 2% Triton-X 100, 2mM PMSF (Phenylmethanesulfonyl fluoride), and 1 tablet Roche’s Protease inhibitor cocktail (Catalogue no. 05892970001) in 1X PBS/10ml solution) [73]

3. 2X Laemmli buffer (65.8mM Tris-HCl, pH 6.8; 26.3% (w/v) Glycerol; 2.1% SDS; and 0.01% Bromophenol blue) [74]
4. MBL’s IP lysis buffer (50mM Tris-HCl pH 7.2, 250mM NaCl, 0.1% NP-40, 2mM EDTA, 10% Glycerol, and 1 tablet Roche’s Protease inhibitor cocktail (per 10ml) in 1X PBS; store at 4°C)

2.3 Stock Buffers

All buffer recipes mentioned below are for 1 liter, unless specified.

1. 10X PBS (8mM Na$_2$HPO$_4$, 2mM KH$_2$PO$_4$, 137mM Sodium Chloride, 27mM Potassium Chloride when diluted to 1X working concentration; pH 7.4) – Applied Biosystems (Catalogue no. 70011069)

2. 10X Running buffer (SDS-PAGE) (25mM Tris-HCl, 192mM Glycine, 0.1% w/v SDS when diluted to 1X working concentration; pH 8.3) – Bio-Rad (Catalogue no. 1610734)

3. 10X Transfer buffer (Western Blotting) (25mM Tris-HCl, 192mM Glycine, 10% w/v Methanol, 20% w/v SDS when diluted to 1X working concentration)

4. 10X TBS (Western Blotting) (50mM Tris-HCl, 150mM Sodium chloride when diluted to 1X working concentration; pH 7.6)

2.4 Freshly-made Buffers

The concentrations of freshly made buffers used in this study are provided below.

1. 0.0-2.0% Formaldehyde in 1X PBS (Freshly prepared, keep at RT in dark)

2. 1.25M Glycine in 1X PBS (store at 4°C)

3. 1X PBS (pH 7.4)

4. 1X Running buffer
5. 1X Transfer buffer with 20% SDS and 10% methanol (store at 4°C – can be re-used)

6. 1X TBST (Tris-buffered Saline with Tween® 20)

7. 5% dried skim milk in 1X TBST (store at 4°C)

8. 2X Laemmli buffer with β-mercaptoethanol (950μl 2X Laemmli with 50μl BME – 1ml solution)

9. Clarity™ Western ECL Substrate (Bio-Rad Catalogue no. 1705061)

2.5 Gels and Membranes Used in the Study

1. 4–20% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad catalogue no. 4561093)

2. 7.5% Mini-PROTEAN® TGX™ Precast Gel (Bio-Rad catalogue no. 4561023)

3. Immun-Blot® PVDF membrane (Bio-Rad catalogue no. 1620239)

4. Precision Plus Protein™ Dual Color Standards (Bio-Rad catalogue no. 1610394)

2.6 Antibodies Used in the Study

1. Primary antibody (anti-GFP rabbit polyclonal antibody – Life Technologies A-6455) (1:2,000 dilution was prepared by mixing 1μl of antibody with 2,000μl of 5% milk in 1X TBST)

2. Secondary antibody (Goat anti-rabbit polyclonal HRP conjugate – Jackson ImmunoResearch Laboratories, Inc. Catalogue No. 111-035-047) (1:20,000 dilution was prepared by mixing 1μl of antibody with 20,000μl of 5% milk in 1X TBST)

3. MBL’s anti-GFP mAb agarose beads (Catalogue no. D153-8)
2.7 **Protein Staining Solutions**

All solution recipes are for 1 liter, unless specified.

1. Gel-fixing solution – (50% Ethanol in deionized water, 10% acetic acid)
2. Coomassie stain – (Coomassie Brilliant Blue R250 2.5gm; Glacial acetic acid 100ml, Methanol: Deionized water (1: 1 v/v) 900ml)
3. Destaining solution – (Glacial acetic acid 100ml, Methanol: Deionized Water (1: 1 v/v) 900ml)
4. Ponceau S stain – 30ml (0.3ml glacial acetic acid, 0.033gm Ponceau S, 30ml Deionized water)

2.8 **Sample Preparation**

_Viking-GFP-tagged flies were cultured using standard culture medium (LabExpress Fly Food B) and incubated at 25°C. The late third instar larvae were collected on the 5th/6th day of culture. Actin-GFP-tagged flies were cultured in the same way and used as control. The use of Actin-GFP larvae as a control aided in eliminating any non-specific detection of proteins, which might have bound to GFP during the normal development._

2.9 **Protein Purification**

2.9.1 *Isolation and detection of proteins*

Three different buffers (RIPA, Lysis buffer, and 2X Laemml) were tested for efficient isolation of collagen IV and GFP proteins from the third instar larvae (Figure 6). I was interested in detection of a band at 250-kDa for collagen IV and at 27-kDa for GFP proteins in Western blot analysis.
2.9.1.1 RIPA buffer extraction process

Ten wandering third instar larvae of Actin-GFP and Vkg-GFP each were collected in separate microcentrifuge tubes and 500µl of freshly made RIPA buffer was added. Lysate was prepared by homogenizing the larvae using pestle and centrifuged at 12,000g for 10 minutes at 4°C. The supernatant was collected in a fresh microcentrifuge tube and stored at -20°C until used. SDS-PAGE analysis was performed by mixing 60µl of sample with 60µl of sample buffer (2X Laemmlli with β-mercaptoethanol). After mixing, the samples were boiled at 95°C for 10 minutes and 30µl was loaded to the 4-20% Mini-Protean TGX precast polyacrylamide gradient gels for optimal resolution. Three microliters of Bio-Rad’s Precision Plus Protein Dual Color Standard was used as ladder. The gels were run at 120 Volts in 1X running buffer for approximately 1 hour 20 minutes. The proteins were then transferred to the PVDF membrane using Western blot wet transfer method with transfer carried out at 40mAmp at 4°C for 18 hours. Once transfer was complete, the PVDF membrane was air dried and presence of protein was confirmed by staining with Ponceau S stain. After successful detection of proteins bands the membrane was washed with Deionized water until stained was completely washed. The membrane was incubated with 5% milk TBST solution for an hour on a rotary shaker at room temperature. The membrane was then incubated with primary antibodies (1:2000) overnight on a shaker at 4°C. Subsequently, the membrane was washed 4 times with 1X TBST for 10 minutes each. After washing, the membrane was incubated with secondary antibodies (1:20,000) for 1 hour at room temperature. The membrane was washed 4 times with 1X TBST for 10 minutes each and then incubated with ECL reagent (1:1 ratio of Clarity Western Peroxide and Western Luminol/Enhancer Reagents kit) for a
few second and observed for chemiluminescence in imager. The desired image was captured using attached camera.

2.9.1.2 Lysis buffer extraction process

The isolation of proteins using this buffer followed the same procedure as mentioned in Section 2.9.1.1; however, there is only one modification that was performed for this isolation process. The boiling step before loading samples to the SDS-PAGE was avoided as the isolation buffer contains urea. Boiling samples at elevated temperature (95°C) causes carbamylation and results in protein precipitation.

2.9.1.3 2X Laemmli buffer extraction process

The isolation of proteins using this buffer followed the same procedure as mentioned in Section 2.9.1.1.

2.10 Cross-linking

The cross-linking step was performed only with 2X Laemmli buffer as RIPA and Lysis buffers did not provide workable amounts of protein after extraction. To perform cross-linking, different concentrations (0.0 to 2.0% at increments of 0.2%) of formaldehyde were used to find a correct concentration for protein cross-linking. Four larvae each of Actin-GFP and Vkg-GFP were collected in 1.5μl microcentrifuge tubes and incubated with 500μl of different formaldehyde concentrations for 7 and 17 minutes on a vortex mixer. The samples were then centrifuged at 20,000g for 3 minutes at room temperature. This made a total of 10 and 20 minutes exposure to formaldehyde. The formaldehyde was quickly removed and 400μl of ice-cold 1.25M glycine was added to quench the reaction. The tubes were then centrifuged at 20,000g for 5 minutes at room
temperature. The supernatant was discarded and the quenching step was repeated with centrifugation carried out at 4°C. The glycine was removed [Protocol modified from Ref. 72] and 200µl of appropriate buffer (2X Laemmli) were added and larvae homogenization was performed using a pestle. The samples were centrifuged at 20,000g for 10 minutes at 4°C and supernatant was collected and stored at -20°C until used. SDS-PAGE and Western blot analysis were performed as mentioned in Section 2.9.1.1 to detect the presence of protein bands at or above 250kDa for the experimental sample and 27KDa for the Actin-GFP sample.

2.11 Co-immunoprecipitation

Once the optimal concentration of formaldehyde (0.4%) was determined, the co-immunoprecipitation step was performed. Two-hundred microliters of cross-linked sample was placed in a 0.5ml microcentrifuge tube and 20µl of anti-GFP monoclonal agarose beads were added. The tubes were then incubated on a rotary shaker for 2 hours at 4°C. All centrifugation steps were performed at 4°C, unless otherwise specified. The beads were pelleted by centrifugation at 2,500g for 10 seconds. The supernatant was discarded and to the pellet, 200µl of freshly-made ice cold IP lysis buffer was added. Tubes were gently inverted to mix the contents and then centrifuged at 2,500g for 10 seconds. This washing step was repeated twice and supernatant was collected in a fresh tube, which was later used to confirm the absence of high molecular weight protein complex in a Western blot analysis. After washing, the pellet was resuspended in 20µl of 2X Laemmli buffer containing β-mercaptoethanol and heated at 65°C for 5 minutes. The samples were quickly vortexed and centrifuged at 2,500g for 5 minutes. The supernatant was transferred to fresh tubes. This step was repeated one more time to ensure complete
protein complex isolation. SDS-PAGE and Western blot analysis were performed on the extracted and supernatant wash samples as mentioned in Section 2.9.1.1.

2.12 Coomassie Staining

Our interest was to find out the interactive partners of collagen IV and one of the best ways to identify these partners is to perform mass spectrometry analysis. The co-immunoprecipitated supernatant samples were loaded to a 7% SDS-PAGE gel and run at 120 Volts for 1 hour. The gels were carefully removed from the cast plates and kept in gel fixing solution for 30 minutes at room temperature with gentle agitation. The gels were incubated in Coomassie stain for 1 hour and then destained until the bands were clearly visible against a light background. The bands were carefully cut individually, collected in clean microcentrifuge tubes, labeled, and stored at -80°C.

2.13 Mass Spectrometry and Data Analysis

The individual bands obtained from the Coomassie staining step were sent for LC-MS/MS spectrometry analysis at the University of Kentucky core facility. Upon completion of the mass spectrometry analysis, data comparison between the genes from Actin-GFP and Vkg-GFP was performed to find the presence of unique genes. Flybase database was used to search the possible functions of these unique genes in the development of Drosophila and their role in the basement membrane. Details on the role of these unique genes during the Drosophila development are tabulated in the Results section 3.4.

2.14 Protein Trap Studies

The protein trap studies were performed on Viking GFP and Surf-4 GFP transgenic larvae. Ten third instar larvae of each were collected in ice-cold 1X PBS solution and
individual larva was dissected using no. 5 Dumont forceps. The dissected larvae were inverted inside-out according to the dissection protocol described in Ref. [75]. The individual larval tissues like wing disc, trachea, salivary gland, and fat body were carefully dissected and mounted as described in the paper. The mounted slides were viewed under Zeiss fluorescence microscope for GFP expression in mentioned tissues.
Figure 6. Schematic representation of the experimental setup. The figure provides a brief design of the overall experimental workflow.

Took 10 third late instar larvae (Actin-GFP & Viking-GFP)

Isolated proteins using three different lysis buffers

Performed cross-linking of proteins with different conc. of formaldehyde (0.0-2.0%)

Performed Co-IP on the selected cross-linked samples

Eluted samples were electrophoresed on 4-20% SDS-PAGE gradient gels

Western blot analysis was performed

Cross-linked proteins were detected above 250kDa bands

RIPA

Lysis buffer

2X Laemmli w/o BME

10 mins formaldehyde exposure

20 mins formaldehyde exposure

Used anti-GFP monoclonal agarose beads

Stained gels in Coomassie and destained to obtain unique bands

Bands were cut out and sent for Mass spectrometry analysis

Data received were compared to identify unique genes/proteins

Unique proteins/gene identified

Performed Initial Characterization on Surf-4
3. RESULTS

3.1 Selection of Isolation Buffer

From the literature, we know that collagen IV accounts for 50% of the total BM mass [4]. Hence, we decided to concentrate the entire project on a single protein, collagen IV. To isolate it, it was first necessary to determine the best isolation buffer. We performed a literature search on buffers for protein isolation and found three commonly used buffers, namely RIPA, Lysis, and 2X Laemmli [72–74]. After proteins were extracted, Western blot analysis was performed using an anti-GFP antibody (Figure 7). As collagen IV (Vkg) is a high molecular weight protein, we expected a band at ~250kDa, while GFP being a small size protein a band is expected at ~25kDa. When using RIPA or lysis buffers, we did not observe protein bands detection at 250kDa and 25kDa, respectively; while clear bands were detected with 2X Laemmli buffer samples (Figure 7, Lysis buffer data not shown). Hence, further experiments were performed using 2X Laemmli buffer.
Figure 7. Comparison of proteins isolated using RIPA and Laemmli buffers. Vkg and GFP represent the samples isolated from Viking-GFP and Actin-GFP transgenic larvae, respectively. Collagen IV is detected at ~250kDa (green arrows), while GFP is detected at ~25kDa (red arrows) in a Western blot using an anti-GFP rabbit polyclonal antibody.
3.2 Standardizing the Formaldehyde Cross-linking Conditions

The main purpose of this study was to identify various interactive partners of collagen IV during the third larval instar stage. One of the best ways to perform interaction studies \textit{in vivo} is to arrest the interacting partners using formaldehyde as a cross-linker. However, the use of formaldehyde as a cross-linker needs to be evaluated to obtain an optimal cross-linking complex. Hence, three parameters were considered 1) The reaction temperature, 2) formaldehyde concentration, and 3) reaction time. We started with different concentrations of formaldehyde ranging from 0.2\% to 2.0\%, at an incremental rate of 0.2\%. The reaction was set up at room temperature (25°C) for 10 and 20 minutes, respectively, with vigorous vortexing. After formaldehyde incubation, proteins were isolated as mentioned earlier in Section 2.9.1.3. Before performing SDS-PAGE, the cross-linked samples were heated with 2X Laemmli buffer containing BME (\(\beta\)-mercaptoethanol) at 65°C for 5 minutes instead of 95°C for 10 minutes. At 95°C with 10 minutes of heat exposure proteins tend to lose their quaternary structure which might also break the formaldehyde bridge between the cross-linked proteins. SDS-PAGE and Western blot analysis showed the presence of high molecular weight bands (above 250kDa) in Vkg-GFP samples and at \(~\)25kDa for Actin-GFP samples. We compared the Western blot results of Vkg-GFP and Actin-GFP which led us to choose the lowest concentration of formaldehyde (0.4\%) as a standard cross-linker with an exposure time of 10 mins for further experiments (Figure 8).
Figure 8. Cross-linking using different concentrations of formaldehyde. Western blot analysis results using an anti-GFP rabbit polyclonal antibody. (a) *Actin-GFP* samples
with 10 minutes of increasing concentrations of formaldehyde exposure; (b) Vkg-GFP samples with 10 minutes of increasing concentrations of formaldehyde exposure; (c) Actin-GFP samples with 20 minutes of increasing concentrations of formaldehyde exposure; and (d) Vkg-GFP samples with 20 minutes of increasing concentrations of formaldehyde exposure. “L” represents ladder, X-axis represents formaldehyde concentrations, and Y-axis represents molecular weight in kDa. The lanes corresponding to the selected concentration is enclosed by a red box.
3.3 Co-immunoprecipitation of Cross-linked Protein Complex

As we were interested in the protein or factors associated with collagen IV, our next step was to isolate only the collagen IV cross-linked complex and rule out the other proteins from the sample. We accomplished this by utilizing the co-immunoprecipitation technique. The process uses a bait protein linked to agarose beads to pull out the target protein from the sample. From Figure 8 it is evident that there is a lot of non-specific binding that may result in a false positive. As we were interested only in the collagen IV associated complex and the protein from Viking-GFP and from the Actin-GFP control was GFP tagged, an anti-GFP bait antibody linked to an agarose bead was judged to work best for our experiments. We used MBL’s anti-GFP mAb agarose beads as a bait protein to extract the collagen IV cross-linked protein complex from the sample. The extracted protein samples (Actin-GFP and Vkg-GFP) were resolved on SDS-PAGE gels and Western blot analysis was performed as previously described. The Western blot analysis revealed five distinct bands in the Vkg-GFP sample and a single band in the Actin-GFP sample (Figure 9). The detection of five bands in the Vkg-GFP sample could be due to the heating step at the SDS-PAGE analysis stage where proteins might have lost the formaldehyde spacer arm separating the complex. We also ran the supernatant from the IP washes. Detection of no high molecular weight band (250kDa) in the Vkg-GFP supernatant confirmed the successful pull out of the collagen IV associated protein complex.
Figure 9. Co-immunoprecipitation results of the Actin-GFP and Viking-GFP cross-linked samples. Western blot analysis on Co-IP samples using an anti-GFP rabbit polyclonal as primary antibody. Lane 2 and 7 shows 0.4% formaldehyde cross-linked samples of Actin-GFP and Vkg-GFP, respectively. Lanes 3 & 4 and 8 & 9 show cross-linked co-immunoprecipitated samples of Actin-GFP and Vkg-GFP. Lanes 5 and 10 show the supernatant flow-through from the co-immunoprecipitated samples. Absence of 250kDa and 25kDa bands in the co-immunoprecipitated supernatant samples confirms the successful isolation of cross-linked complex. “L” represents ladder, X-axis represents formaldehyde concentrations, and Y-axis represents molecular weight in kDa.
3.4 Mass Spectrometric Analysis of Complexes Containing Collagen IV

Western blot analysis of the co-immunoprecipitated samples provided us with the confirmation of isolation of collagen IV and its associated complex. However, we still did not know which proteins or factors were present in this complex. To identify these collagen IV associated proteins and factors we had mass spectrometry analysis (LC-MS/MS) performed on the co-immunoprecipitated cross-linked samples. The collagen IV complex was precipitated using anti-GFP agarose beads, separated by SDS-PAGE, and visualized using Coomassie stain (Figure 10). The stained bands were excised carefully and sent to the collaborators for mass spectrometry analysis. The results from experimental (Viking-GFP) and control (Actin-GFP) samples were then compared and those proteins or factors that were unique to Vkg were selected. Literature reviews were performed on these unique proteins for their possible role in the development of Drosophila (Table 1). We again performed mass spectrometry analysis on the fresh Co-IP coomassie stained samples which resulted in several new candidates. Comparison of the unique Viking candidates between the two mass spectrometry results provided us with six candidates that were common (Table 2). Being able to find these six candidates in both rounds of mass spectrometry analysis we were confident of their interaction with collagen IV and decided to perform initial characterization on one of the candidate.
**Figure 10. Coomassie stain on co-immunoprecipitated samples.** Co-immunoprecipitated samples electrophoresed on 7% SDS-PAGE gel. “L” represents ladder, “G” represents Co-IP Actin-GFP sample, and “V” represents Co-IP Viking-GFP sample. Individual bands were excised carefully and sent for mass spectrometry analysis.
Table 1. Comparison of the mass spectrometry result between *Actin-GFP* and *Viking-GFP*. Only proteins unique to Vkg-GFP are listed.

<table>
<thead>
<tr>
<th>Annotation Symbol</th>
<th>Gene Name</th>
<th>Gene Role</th>
<th>Functional Category</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG1528</td>
<td>gammaCop</td>
<td>Responsible for secreting luminal components and assembles luminal chitinous during embryonic stage.</td>
<td>Development</td>
<td>[76, 77]</td>
</tr>
<tr>
<td>CG15792</td>
<td>zip</td>
<td>Associated with ECM through PS2 integrin. Acts as an actin cross-linker and maintain the structural integrity of sarcomeric muscle cytoskeleton. It is also involved in the D-V compartmentalization.</td>
<td>Development</td>
<td>[78, 79]</td>
</tr>
<tr>
<td>CG5210</td>
<td>CG5210</td>
<td>No data found.</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG1483</td>
<td>Map205</td>
<td>Interacts with polo, a kinase involved in regulating cell cycle division.</td>
<td>Cancer</td>
<td>Cell [80]</td>
</tr>
<tr>
<td>CG8193</td>
<td>PPO2</td>
<td>Found in the crystal cells in the hemolymph and responsible for protecting the injury site by producing melanin.</td>
<td>Protection</td>
<td>[81]</td>
</tr>
<tr>
<td>Gene</td>
<td>Symbol</td>
<td>Description</td>
<td>Category</td>
<td>Reference(s)</td>
</tr>
<tr>
<td>---------</td>
<td>--------</td>
<td>------------------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>CG6705</td>
<td>tsl</td>
<td>A member of the membrane attack complex/perforin-like protein superfamily (MACPF), involved in pore-forming and immune defense roles and in influencing the developmental timing by localizing in the prothoracic gland.</td>
<td>Protection and Development</td>
<td>[82]</td>
</tr>
<tr>
<td>CG4314</td>
<td>st</td>
<td>Responsible for the biosynthesis of xanthommatin.</td>
<td>Eye pigmentation</td>
<td>[83]</td>
</tr>
<tr>
<td>CG5020</td>
<td>CLIP-190</td>
<td>Regulates microtubule dynamics and links microtubule plus-ends with other cellular structures.</td>
<td>Development</td>
<td>[84]</td>
</tr>
<tr>
<td>CG3910</td>
<td>mtTFB2</td>
<td>Has a role in cell proliferation and differentiation.</td>
<td>Development</td>
<td>[85]</td>
</tr>
<tr>
<td>CG31623</td>
<td>dtr</td>
<td>No data found.</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG17291</td>
<td>Pp2A-29B</td>
<td>Regulates starvation-induced autophagy. It also acts as a brain tumor-suppressor that controls the self-renewal and differentiation of neural stem cells.</td>
<td>Cell Death and cancer</td>
<td>[86, 87]</td>
</tr>
<tr>
<td>CG1848</td>
<td>LIMK1</td>
<td>Involved in Rho signaling during metamorphosis.</td>
<td>Development</td>
<td>[88]</td>
</tr>
<tr>
<td>CG42338</td>
<td>Ten-a</td>
<td>Involved in the selection of specific synaptic partners in the olfactory circuit of Drosophila.</td>
<td>Others</td>
<td>[89]</td>
</tr>
</tbody>
</table>
Alpha-Actinin is an actin filament cross-linking protein found in muscle and non-muscle cells. Muscle alpha-actinin is found to be associated with the Z-disc involved in cross-linking of actinin filaments to the adjacent sarcomeres. Non-muscle alpha-actinin is found to be present in stress-fibers, lamellipodia, cell-cell and cell-matrix adhesion sites.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein</th>
<th>Description</th>
<th>Section</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG4376</td>
<td>Actn</td>
<td></td>
<td>Development</td>
<td>[90]</td>
</tr>
<tr>
<td>CG17420</td>
<td>RpL15</td>
<td>Defect in the gene is associated with minute syndrome.</td>
<td>Disease</td>
<td>[91]</td>
</tr>
<tr>
<td>CG3395</td>
<td>RpS9</td>
<td>Associated with transcription of genes.</td>
<td>Others</td>
<td>[92]</td>
</tr>
<tr>
<td>CG10944</td>
<td>RpS6</td>
<td>Associated with the maintenance of the hematopoietic organ where mutation results in cell enlargement and over-proliferation of the hemocytes.</td>
<td>Cancer</td>
<td>[93, 94]</td>
</tr>
<tr>
<td>CG3203</td>
<td>RpL17</td>
<td>Mutation results in minute syndrome.</td>
<td>Disease</td>
<td>[91]</td>
</tr>
<tr>
<td>CG8922</td>
<td>RpS5a</td>
<td>Mutation results in minute syndrome.</td>
<td>Disease</td>
<td>[91]</td>
</tr>
<tr>
<td>CG4533</td>
<td>l(2)efl</td>
<td>Has role in clearing of poly-glutamine proteins.</td>
<td>Disease</td>
<td>[95]</td>
</tr>
<tr>
<td><strong>CG4800</strong></td>
<td><em>Tctp</em></td>
<td>Associates with ATM kinase and functions in growth regulation and cancer.</td>
<td><strong>Cancer</strong></td>
<td>[96, 97]</td>
</tr>
<tr>
<td><strong>CG34410</strong></td>
<td><em>Rab26</em></td>
<td>Interacts with different effector proteins and facilitates the transport of vesicles to the membrane.</td>
<td><strong>Others</strong></td>
<td>[98]</td>
</tr>
<tr>
<td><strong>CG7576</strong></td>
<td><em>Rab3</em></td>
<td>Localized at the presynaptic boutons and facilitates synaptic vesicle cycle.</td>
<td><strong>Others</strong></td>
<td>[99]</td>
</tr>
<tr>
<td><strong>CG6601</strong></td>
<td><em>Rab6</em></td>
<td>Regulates the N-cadherin trafficking in association with the Rich protein.</td>
<td><strong>Others</strong></td>
<td>[100]</td>
</tr>
<tr>
<td><strong>CG1088</strong></td>
<td><em>Vha26</em></td>
<td>Responsible for transporting ions in exchange of ATP. Maintain acidic environment of the lysosomes and pericellular spaces.</td>
<td><strong>Others</strong></td>
<td>[101, 102]</td>
</tr>
<tr>
<td><strong>CG34090</strong></td>
<td><em>mt:Cyt-b</em></td>
<td>No data found.</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>CG6202</strong></td>
<td><em>Surf-4</em></td>
<td>Being a cargo receptor protein, surf-4 interacts with ERGIC-53 and proteins of the p25 family. Responsible for proper recruitment of COPI in the early secretory pathway.</td>
<td><strong>Development</strong></td>
<td>[54–59]</td>
</tr>
<tr>
<td>Gene</td>
<td>Description</td>
<td>Others</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>CG1404</td>
<td><em>ran</em> Involved in regulating several cellular functions of cell cycle which includes nucleocytoplasmic transport, nuclear membrane assembly, and spindle assembly</td>
<td>Others</td>
<td>[103]</td>
<td></td>
</tr>
<tr>
<td>CG6510</td>
<td><em>RpL18A</em> Mutation results in minute syndrome.</td>
<td>Disease</td>
<td>[91]</td>
<td></td>
</tr>
<tr>
<td>CG31034</td>
<td><em>Jon99Cii</em> Found abundantly in larval gut.</td>
<td>Others</td>
<td>[104]</td>
<td></td>
</tr>
<tr>
<td>CG7360</td>
<td><em>Nup58</em> No data found.</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CG8987</td>
<td><em>tam</em> Mutation results in developmental defects of adult fly visual system.</td>
<td>Development</td>
<td>[105] Eye</td>
<td></td>
</tr>
<tr>
<td>CG12530</td>
<td><em>Cdc42</em> Regulator of axon outgrowth, branching, and guidance.</td>
<td>Others</td>
<td>[106, 107]</td>
<td></td>
</tr>
<tr>
<td>CG33180</td>
<td><em>Ranbp16</em> No data found.</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CG3167</td>
<td><em>aub</em> Involved in guiding the siRNA for performing degradation activity.</td>
<td>Programmed cell death</td>
<td>[109]</td>
<td></td>
</tr>
<tr>
<td>Gene ID</td>
<td>Symbol</td>
<td>Description</td>
<td>Category</td>
<td>Reference</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td>CG33114</td>
<td>45E</td>
<td>No data found.</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CG31196</td>
<td>14-3-3Epsilon</td>
<td>Involved in Ras1 signaling.</td>
<td>Cancer</td>
<td>[110]</td>
</tr>
<tr>
<td>CG30263</td>
<td>stum</td>
<td>Has role in proprioception feedback resulting in proper locomotion of</td>
<td>Others</td>
<td>[111]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Drosophila.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CG8416</td>
<td>Rho1</td>
<td>Required for neuroblast proliferation.</td>
<td>Others</td>
<td>[112]</td>
</tr>
</tbody>
</table>
Table 2. Proteins found common in both rounds of mass spectrometry analysis and unique to Viking-GFP sample.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Actn</strong></td>
<td>alpha-Actinin is an actin filament cross-linking protein found in muscle and non-muscle cells. In Drosophila, three alpha-actinin isoforms – non-muscle, larval muscle-specific, and adult muscle-specific – have been identified. Muscle alpha-actinin is found to be associated with the Z-disc involved in cross-linking of actinin filaments to adjacent sarcomeres. While non-muscle alpha-actinin is found to be present in stress-fibers, lamellipodia, cell-cell, and cell-matrix adhesion sites.</td>
<td>[90]</td>
</tr>
<tr>
<td><strong>RpL18A</strong></td>
<td>60S ribosomal protein L18a is also known as minute gene. Though studies could not find any direct correlation of it causing the syndrome as it lies in a gap in deletion coverage of the genome, but with point or transposon insertion mutations it can cause minute phenotype.</td>
<td>[91]</td>
</tr>
<tr>
<td><strong>RpS9</strong></td>
<td>40S Ribosomal protein S9 is a type of non-ribosomal protein associated with the transcription of genes. Studies carried out in Drosophila showed its relation with the transcription factors and found to localize at the transcription site.</td>
<td>[92]</td>
</tr>
<tr>
<td><strong>RpL17</strong></td>
<td>The mutation in the 60S Ribosomal protein L17 gene was found to cause minute syndrome in <em>Drosophila</em>, which results in prolonged development, short and thin bristles, and poor fertility and viability.</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Surf-4</strong></td>
<td>Co-localization studies showed that surf-4 localizes to ER, ERGIC (Endoplasmic Reticulum-Golgi Intermediate Compartment), and Golgi apparatus depending on the signals present at its N- and C-terminal. Being a cargo receptor protein, surf-4 interacts with ERGIC-53 and proteins of the p25 family. Double knockdown of surf-4 and ERGIC-53 showed that the complex is responsible for proper recruitment of COPI in the early secretory pathway. Studies have also shown that it is responsible for transport of soluble proteins.</td>
<td></td>
</tr>
<tr>
<td><strong>ran</strong></td>
<td>Studies showed that Ran is involved in regulating several cellular functions of the cell cycle, which includes nucleocytoplasmic transport, nuclear membrane assembly, and spindle assembly. One of the studies performed on Drosophila’s neuroblast had shown that Ran associates with Canoe, a protein that regulates spindle orientation and cell polarity. Canoe/Ran-GTP complex then interact with Pins (Partner of Inscuteable) and then form complex with Mud (Mushroom body defect) resulting in the activation of Pins/Mud/dynein spindle orientation pathway.</td>
<td></td>
</tr>
</tbody>
</table>
3.5 Surf-4 and its Localization in Drosophila Third Instar Larvae

From all the unique proteins that were found associated with the collagen IV we decided to test the expression of surf-4 protein in vivo. Literature reviews indicated that surf-4 acts as a membrane receptor for the transport of soluble proteins between Endoplasmic Reticulum and Golgi apparatus [54–59]. As from the literature review on the biogenesis of collagen IV we found that the protein during its initial biosynthesis stage is expressed in the soluble form [50]. Therefore, we propose that surf-4 may be involved in the early secretory pathway and might play an important role in the transport of collagen IV. In order to understand the role of the surf-4 protein and its possible involvement in the early secretory pathway, we performed protein trap expression assay, over-expression, and knockdown studies. To better understand the function of Surf-4 we utilized a transgenic third instar larvae expressing surf-4-GFP protein to locate the protein expression. We fixed various larval tissues from the surf-4 protein trap and analyzed the samples using fluorescence microscopy. Results from this experiment are provided in Figure 11, which displays the expression pattern of surf-4 protein in different Drosophila third instar larval tissues. It is evident from this figure that the protein is expressed ubiquitously throughout the organism. We also looked at the expression pattern for GFP-tagged collagen IV. From the figure it is evident that collagen IV is localized only at the basal side of the tissues.
Figure 11. Expression pattern of Surf-4 and Collagen IV in different tissues of *Drosophila melanogaster*. A–D. Expression of GFP-tagged Surf-4 in various structures of third instar larvae. E and F. Expression of GFP-tagged Collagen IV in various structures of third instar larvae. Green channel represents expression of GFP while blue channel represents nucleus stained with DAPI.
3.6 *Surf-4* RNAi Knockdown Study

The *Surf-4* GFP protein trap expression study provided evidence of the protein being ubiquitously expressed throughout the larval tissues. Hence, we decided to perform knockdown study to see if there is any robust effect on the phenotype produced. We used two *Surf-4* RNAi lines obtained from Vienna Drosophila Research Center (VDRC - GD5883) and Bloomington Drosophila Stock Center (BDSC - 57471). We performed crosses between *Surf-4* RNAi lines and different RNAi drivers. The RNAi drivers contain Dicer-2 gene (a type of RNase III enzyme) which enhances the RNAi effect. The crosses were performed between *Surf-4* RNAi virgin females and males of different RNAi drivers. The phenotypes produced using 57471 *Surf-4* RNAi line show defects in the wings (Figure 12C, D, G, H, K, and L and Figure 13G) and thorax bristles (Figure 13C and D). The phenotypic defects indicate that *surf-4* might play some role during the development of *Drosophila* (Table 3, Figures 12 and 13). However, we could not observe any defects in the phenotypes produce in the GD5883 *Surf-4* RNAi line (Table 4). One of the possible reasons for this could be that the *Surf-4* RNAi line (GD5883) is weak to provide any robust phenotypic effect. In addition, we were still not sure if the phenotypic defects observed from the *Surf-4* RNAi line (57471) were due to the knockdown of *surf-4*. Therefore, performing RT-PCR on one of the *Surf-4* RNAi crosses would provide us with the indication of downregulation of *Surf-4* gene expression. In addition, the phenotypes from the knockdown of *surf-4* were indicative of apoptosis, which could be mediated by the JNK pathway [113]. Hence, to explore the possibility that the downregulation of *surf-4* resulted in induction of the JNK pathway we decided to assay for JNK pathway activation using a well-established LacZ reporter.
3.6.1  **Lac Z Staining**

We crossed *Surf-4* RNAi virgin females with the males of *Ptc-Gal4, UAS-GFP; Puc-lacZ/TM6Tb* genotype. The Puc-lacZ gene present in the driver helps to report the up-regulation of the JNK pathway, which will be indicated by a blue precipitate when stained for β-galactosidase reporter gene activity. However, when the cross was performed at 25°C we observed that most of the progenies growth was arrested at the first instar stage. We suspect that over-expression of *Surf-4* RNAi might have blocked the overall production of surf-4 mRNA which would have resulted in the cell death at the early growth stage. We also performed the cross at 18°C to observe the effect, as the Gal4 expression is reduced at lower temperature leading to less severe phenotype. However, we observed the same result as observed at 25°C. These observations are suggestive of a possible role for surf-4 in the early larval development stage.
Table 3. Generated phenotype description when *Surf-4* (BDSC - 57471) is knocked down using different RNAi drivers.

<table>
<thead>
<tr>
<th>RNAi Drivers</th>
<th>Where Expressed</th>
<th>Progeny Phenotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>25706 ♀</td>
<td>Wings</td>
<td>Wings found unfolded in ♀ progenies.</td>
</tr>
<tr>
<td>25706 w1118 P{GawB}BxMS1096; P{UAS-Dcr-2.D}2</td>
<td>Wings</td>
<td>No defects observed in progenies.</td>
</tr>
<tr>
<td>25708 ♀</td>
<td>Wings</td>
<td>Found oval-shaped eye in two ♀ progenies only.</td>
</tr>
<tr>
<td>25750 ♀</td>
<td>Eye</td>
<td>No defects observed in progenies.</td>
</tr>
<tr>
<td>25752 ♀</td>
<td>Wings</td>
<td>Wing defects were observed in progenies.</td>
</tr>
<tr>
<td>25753 ♀</td>
<td>Wings</td>
<td>No defects observed in progenies.</td>
</tr>
<tr>
<td>25757 ♀</td>
<td>Wings</td>
<td>No defects observed in ♀ progenies.</td>
</tr>
<tr>
<td>25758 ♀</td>
<td>Thorax</td>
<td>Progenies showed less bristles on thorax.</td>
</tr>
</tbody>
</table>

*Surf-4* RNAi line: y1 sc* v1; P[TRiP.HMC04782]attP40
$P\{\text{UAS-Dcr-2.D}\}1, \text{w1118}; P\{\text{en2.4-GAL4}\}e16E, P\{\text{UAS-2xEGFP}\}AH2$

$y1 \text{ sc}\ast v1; P\{\text{TRiP.HMC04782}\}\text{attP40}$

$P\{\text{UAS-Dcr-2.D}\}1, \text{w1118}; P\{\text{bs-GAL4.Term}\}G1/CyO$

$y1 \text{ sc}\ast v1; P\{\text{TRiP.HMC04782}\}\text{attP40}$

$P\{\text{UAS-Dcr-2.D}\}1, \text{w1118}; P\{\text{GawB}\}bbgC96$

$y1 \text{ sc}\ast v1; P\{\text{TRiP.HMC04782}\}\text{attP40}$
Figure 12. Defective phenotypes when Surf-4 is knocked down. A. RNAi wild type driver: \(P\{UAS-Dcr-2.D\}1, w1118; P\{en2.4-GAL4\}e16E, P\{UAS-2xEGFP\}AH2\) (25752); B, F, and J. Surf-4 RNAi wild type: \(y1 sc^* vl; P\{TRiP.HMC04782\}attP40\) (57471); C and D: Progenies of crosses (A and B) showing defects in wings; E. RNAi Wild type driver: \(P\{UAS-Dcr-2.D\}1, w1118; P\{bs-GAL4.Term\}G1/CyO\) (25753); G and H. Progenies of crosses (E and F) showing defects in wings; I. RNAi wild type driver: \(P\{UAS-Dcr-2.D\}1, w1118; P\{GawB\}bbgC96\) (25757); K and L. Progenies of crosses (I and J) showing defects in wings. Arrows show the defect locations.
Figure 13. Defective phenotypes when Surf-4 is knocked down. A. RNAi wild type driver: P(UAS-Dcr-2.D)1, w1118; P(GawB)pnrMD237/TM3, Ser1 (25758); B and F. Surf-4 RNAi wild type: y1 sc* v1; P(TRiP.HMC04782)attP40 (57471); C and D. Progenies of crosses (A and B) showing less bristles on the thorax, indicated by arrows. E. RNAi wild type driver: w1118 P(GawB)BxMS1096; P(UAS-Dcr-2.D)2 (25706); G. Female progenies of crosses (E and F) showing defects in wing unfolding.
Table 4. Phenotype description when *Surf-4* (VDRC – GD5883) is knocked down using different RNAi drivers.

<table>
<thead>
<tr>
<th>RNAi Drivers</th>
<th>Where Expressed</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>25706♂</td>
<td>Wings</td>
<td>No defects found.</td>
</tr>
<tr>
<td>P{GawB}BxMS1096; P{UAS-Dcr-2.D}2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25708♂</td>
<td>Wings</td>
<td>No defects found.</td>
</tr>
<tr>
<td>P{UAS-Dcr-2.D}1, w1118; P{Act5C-GAL4}25FO1/CyO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25750♂</td>
<td>Eye</td>
<td>No defects found.</td>
</tr>
<tr>
<td>P{GawB}elavC155 w1118; P{UAS-Dcr-2.D}2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25752♂</td>
<td>Wings</td>
<td>No defects found.</td>
</tr>
<tr>
<td>P{UAS-Dcr-2.D}1, w1118; P{en2.4-GAL4}e16E, P{UAS-2xEGFP}AH2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25753♂</td>
<td>Wings</td>
<td>No defects found.</td>
</tr>
<tr>
<td>P{UAS-Dcr-2.D}1, w1118; P{bs-GAL4.Term}G1/CyO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25757♂</td>
<td>Wings</td>
<td>No defects found.</td>
</tr>
<tr>
<td>P{UAS-Dcr-2.D}1, w1118; P{GawB}bbgC96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25758♂</td>
<td>Thorax</td>
<td>No defects found.</td>
</tr>
<tr>
<td>P{UAS-Dcr-2.D}1, w1118; P{GawB}pnrMD237/TM3, Ser1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Surf-4* RNAi line: w1118 P{GD2999}v5883
3.7 **Surf-4 Over-expression Study**

Knockdown studies provided us with an indication of the possible effect of surf-4 in development. So we hypothesized that over-expression of surf-4 would result in an increase in its transport activity leading to aberrant phenotypes as well. We used UAS-Gal4 system to perform the over-expression of surf-4 protein using different UAS-Gal4 drivers. However, we did not find any specific difference in the phenotypes produced with different drivers (Table 5). The possible explanation for this observation could be that Surf-4 over-expression line used was too weakly over-expressed to provide a robust phenotypic effect. This could be tested by performing RT-PCR on these flies in the future.
Table 5. Phenotype description when *Surf-4* (19202) is overexpressed using Gal4 drivers.

<table>
<thead>
<tr>
<th>UAS-Gal4 Driver Lines</th>
<th>Where Expressed</th>
<th>Phenotype Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cg-gal4, UAS-GFP/CyO</em></td>
<td>Hemocytes</td>
<td>No phenotypic change observed.</td>
</tr>
<tr>
<td><em>w; Ey-gal4/CyO</em></td>
<td>Eyes</td>
<td>No phenotypic change observed.</td>
</tr>
<tr>
<td><em>GMR-gal4</em></td>
<td>Eyes</td>
<td>No phenotypic change observed.</td>
</tr>
<tr>
<td><em>w; LSP2-gal4</em></td>
<td>Fat Body</td>
<td>No phenotypic change observed.</td>
</tr>
<tr>
<td><em>Pnr-gal4/Tm6Tb</em></td>
<td>Thorax</td>
<td>No phenotypic change observed.</td>
</tr>
<tr>
<td><em>Ptc-gal4, UAS-GFP</em></td>
<td>Wings and other tissues</td>
<td>No phenotypic change observed.</td>
</tr>
<tr>
<td><em>Sd-gal4</em></td>
<td>Wings and other tissues</td>
<td>No phenotypic change observed.</td>
</tr>
<tr>
<td><em>yw; ; Tub-gal4/Tm3Sb</em></td>
<td>Ubiquitous</td>
<td>No phenotypic change observed.</td>
</tr>
<tr>
<td><em>w; ; Ubx-gal4/Tb</em></td>
<td>Haltere and other tissues</td>
<td>No phenotypic change observed.</td>
</tr>
<tr>
<td><em>Vg-gal4</em></td>
<td>Wings</td>
<td>No phenotypic change observed.</td>
</tr>
</tbody>
</table>

*Surf-4* over-expression line: *w1118; P{XP}Surf4d04274/TM6B, Tb1*
4. DISCUSSION

Basement membrane (BM) plays a very important role during the development of an organism. Continuous remodeling and degradation of the BM occurs during the development of an organism, which involves interaction with various components like growth factors, enzymes, and proteins [14]. Invasion of cancer to other parts of the body involves breakage of the BM at various places, which makes it important to understand the BM and its interacting components. This thesis focuses on the isolation of these interacting components with one of the major BM protein, collagen IV, during the third instar larval stage of Drosophila. Because not many studies have been performed on the isolation of collagen IV in Drosophila, we began our study with the selection of a suitable isolation buffer. We used third instar larvae expressing Actin-GFP and Viking-GFP for this study. One of the reasons we chose Actin-GFP larvae as our control was because it enabled us to optimize GFP protein detection during Western blot analysis. In addition, as we were interested in finding the interactive partners of collagen IV, use of Actin-GFP as a control offered a way to rule out any interactive partners that non-specifically associated with GFP under our experimental conditions. We chose three different isolation buffers, namely, RIPA, Lysis, and 2X Laemml to isolate collagen IV. After several repeated attempts and troubleshooting with the sample size, buffers, working conditions, and antibody concentrations, we were able to detect a collagen IV band at approximately 250kDa in the samples isolated using 2X Laemml buffer. We repeatedly performed the isolation steps and optimized the protocol for isolation of collagen IV from the third instar larvae (Figure 7).
Once we confirmed the isolation of protein, our next step was to isolate the protein and its associated interactive partners. To perform protein interaction studies we used formaldehyde as a cross-linker. Formaldehyde has a spacer arm of 2.3–2.7 Å that helps to cross-link only those interactive partners that are in close vicinity of collagen IV. Three important criteria such as exposure times, temperature, and formaldehyde concentrations are required to be optimized when using formaldehyde as a cross-linker. We used different concentrations of formaldehyde (0.0-2.0% at an increment of 0.2%) with 10 and 20 minutes exposure at room temperature to determine the best cross-linking conditions. As we were working with the third instar larval stage we also performed the mentioned conditions on live larvae and dissected larvae. Figure 8 provides the results of the live larvae samples incubated with different formaldehyde concentrations and exposure time of 10 and 20 minutes, respectively at room temperature. As the concentration and exposure time of formaldehyde increased, it resulted in creating non-specific cross-links, which is clearly evident by the presence of less bands (Figure 8C and D). As we were interested in the interactions that are in close proximity, using higher concentration and longer exposure time could provide us with a false positive result. In addition, similar experiment on the dissected larvae samples did not yield good results (data not shown). One of the possible reasons could be the release of protease enzymes during the larval dissection that might have degraded the proteins. Therefore, we decided to use 0.4% of formaldehyde concentration with exposure time of 10 minutes on live larval samples as optimal condition for our further experiments.

Once we optimized the condition for cross-linking our next step was to pull out only the collagen IV and its associated complex and rule out other protein complexes that
might have formed. Hence, we decided to use co-immunoprecipitation technique on the cross-linked samples to pull out our protein of interest. The technique uses an agarose bead linked with a bait protein antibody and when incubated with the sample is capable of pulling out only the protein of interest. As our target protein was tagged with GFP, we used anti-GFP protein linked agarose beads as a bait to pull out collagen IV associated complex. Figure 9 provides the co-immunoprecipitation results on the cross-linked samples. Lanes 5 and 10 contain the samples of flow-through after incubation with the agarose beads. The absence of high molecular weight band (~250kDa) in Lane 10 and 25kDa band in Lane 5 indicated the successful pull out of collagen IV and its associated complex and GFP protein in the control sample, respectively. The presence of multiple bands in Lanes 8 and 9 is probably due to the addition of β-mercaptoethanol and boiling at 65°C before running the samples in the SDS-PAGE, which might have resulted in breakage of cross-links. Another possibility could be the GFP positive bands on the western blots are the degradation products from Viking-GFP.

Finally, we used mass spectrometry analysis to identify the interacting partners of collagen IV from the cross-link samples. Before the mass spectrometry analysis was performed by our collaborators at the University of Kentucky, the cross-linked Co-IP samples were electrophoresed on 7% SDS-PAGE gel. Upon coomassie staining of the gel similar bands appeared in both control Actin-GFP samples and experimental Viking-GFP samples. The reason for these similar bands could be due to the boiling step and addition of β-mercaptoethanol before the samples were electrophoresed. The presence of β-mercaptoethanol along with boiling would break the covalent bonds between the cross-links as well as the disulfide bonds in the protein structure resulting in similar band
detection. Once the data analysis was received we compared the genes detected in both the samples and ruled out all those genes from the Viking-GFP sample that were also found in the Actin-GFP control sample. The first round of the mass spectrometry analysis provided us with a total of 42 unique candidates while second round provided 51 unique candidates. The detection of more candidates in the second round could be due to the dynamic nature of BM in third instar larval samples. If this was the case then future experiment should be focused on timed-protein isolation. Another possibility could be the variability in protein isolation and cross-linking. While the difference in isolated proteins from the first and second round may be perplexing, it is important to note that we were able to identify six that were common in both rounds. Hence, we decided to perform further characterization studies on one of them. Initial literature reviews on these six candidates revealed that they play important roles during development. Out of these six candidates, we decided to perform initial characterization studies on Surf-4 to understand about its possible role and effect through association with collagen IV.

Literature review showed that Surf-4 plays an important role in the early secretory pathway and works as a cargo receptor protein between ER and Golgi apparatus [54]. It was also found that the protein might be responsible for the transport of soluble proteins [54–59]. In addition, biogenesis of collagen IV provided with the information that the protein during its initial biosynthesis stage is secreted in the soluble form. So we hypothesized that surf-4 might be involved in the transport of collagen IV from the ER to Golgi apparatus. In order to link the role of surf-4 with collagen IV we initially performed characterization studies on surf-4 and its role during the development of *Drosophila*. Protein trap studies using transgenic larvae expressing GFP-tagged surf-4
showed ubiquitous expression of protein within various tissues (Figure 1). This result indicates the possible involvement of surf-4 protein in the development of various tissues. We also performed knockdown of surf-4 protein using different RNAi drivers to observe any defect in the progenies. The progenies showed defects in the wings and thorax (Figures 12 and 13). However, when we over-expressed Surf-4 we were unable to detect any phenotypic change in the progenies. From this result it could be inferred that the Surf-4 over-expression line we used was weak to provide an obvious phenotype. Another possibility might be the existence of a feedback inhibition mechanism that might have blocked the surf-4 from performing its activity. Additionally, the amount of surf-4 protein may be present and in saturating quantities so that any additional protein has no effect. RT-PCR on the progenies generated would be required to rule out the latter possibilities. In addition, we noticed that when Surf-4 RNAi crosses were driven using the Ptc-Gal4 driver, most of the progenies growth were arrested at the early development stage (1\textsuperscript{st} instar) when incubated at 25°C. Similar result was observed when the cross was performed at 18°C, as the lower temperature reduces the Gal4 expression resulting in less severe phenotype. These observations indicate that Surf-4 might be involved during the early growth stage of Drosophila.

4.1 Future Directions

4.1.1 In-vivo Protein Knockdown Studies

Protein trap, and lacZ staining allowed for initial characterization studies for Surf-4. However, we were not able to provide any connection between the surf-4 and collagen IV. Hence, the next step would be to confirm these interactions \textit{in vivo} and \textit{in vitro}. In order to perform this experiment, we would require to perform crosses between
transgenic lines expressing GFP-tagged collagen IV and *Surf-4* RNAi with different RNAi drivers. We hypothesize that if surf-4 is involved in the transport of collagen IV, performing this cross would result in either no or defective progenies. From the literature review of *Surf-4* [103–107] we know that it is involved in the transport of soluble proteins during the early secretory pathway. If *Surf-4* is knocked down there would not be any or only low level of collagen IV transport. As collagen IV is important in maintaining the structural integrity of various tissues, its irregular supply will result in absence or defective BM formation around the cells/tissues, causing early stage lethality.

4.1.2 Characterization Studies for other Mass Spectrometry Candidates

This study provides for the initial characterization of a single candidate found in the mass spectrometry analysis. However, there are other interactive partners of collagen IV that are yet to be explored. Initial literature reviews on these candidates showed that they are involved in performing critical functions during the cell cycle, for example, Ran [103]. One of these candidates is also involved in the transcription of genes, for example, Rps9 [92]. Thus, understanding their interaction mechanism with collagen IV could open-up new insights into the role they play during the development of *Drosophila*.

4.1.3 Repetition of the Mass Spectrometry Analysis

From the first two mass spectrometry analysis we observed that there were detections of several new candidates during each round. If the same pattern was repeated during the third round performing a time-based experiment would provide us with the pattern of the protein interaction at that particular time. Information gained through these timed experiments would indicate the dynamic nature of BM composition.
5. REFERENCES


