

Stephen F. Austin State University

**SFA ScholarWorks**

---

Electronic Theses and Dissertations

---

5-2021

## Simulation of the Interaction Between Striated Muscle UNC-45 and Transcription Factor GATA-4

Drake Alexander Duncan

*Stephen F. Austin University*, drakeaduncan@gmail.com

Follow this and additional works at: <https://scholarworks.sfasu.edu/etds>



Part of the [Biochemistry Commons](#), [Bioinformatics Commons](#), [Chemistry Commons](#), and the [Molecular Biology Commons](#)

Tell us how this article helped you.

---

### Repository Citation

Duncan, Drake Alexander, "Simulation of the Interaction Between Striated Muscle UNC-45 and Transcription Factor GATA-4" (2021). *Electronic Theses and Dissertations*. 371.

<https://scholarworks.sfasu.edu/etds/371>

This Thesis is brought to you for free and open access by SFA ScholarWorks. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of SFA ScholarWorks. For more information, please contact [cdsscholarworks@sfasu.edu](mailto:cdsscholarworks@sfasu.edu).

---

# Simulation of the Interaction Between Striated Muscle UNC-45 and Transcription Factor GATA-4

## Creative Commons License



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

SIMULATION OF THE INTERACTION BETWEEN STRIATED MUSCLE UNC-45  
AND TRANSCRIPTION FACTOR GATA-4

By

DRAKE DUNCAN, Bachelor of Science, Master of Business Administration

Presented to the Faculty of the Graduate School of

Stephen F. Austin State University

In Partial Fulfillment

Of the Requirements

For the Degree of

Master of Science

STEPHEN F. AUSTIN STATE UNIVERSITY

May 2021

SIMULATION OF THE INTERACTION BETWEEN STRIATED MUSCLE UNC-45  
AND TRANSCRIPTION FACTOR GATA-4

By

DRAKE DUNCAN, Bachelor of Science, Master of Business Administration

APPROVED:

---

Odutayo Odunuga, Ph.D., Thesis Director

---

Darrell Fry, Ph.D., Committee Member

---

Michele Harris, Ph.D., Committee Member

---

Josephine Taylor, Ph.D., Committee Member

---

Pauline M. Sampson, Ph.D.  
Dean of Research and Graduate Studies

## ABSTRACT

Striated Muscle UNC-45, also known as UNC-45b, is an important protein that acts as a chaperone for myosin in cardiac and skeletal muscles, binding to myosin at its C-terminal UCS domain and regulating its assembly into thick filaments and sarcomeric structures. The UCS domain contains a large loop that is believed to be the first point of interaction between myosin and UNC-45b. GATA-4 is an essential transcription factor that facilitates transcription of several genes in cardiac development, particularly alpha-heavy chain myosin in heart tissue. Recently, studies have shown that there is interaction of GATA-4 with UNC-45b and that GATA-4 binds to the UCS domain. The implications of this interaction suggests that UNC-45b may play a role in not only the folding of myosin, but in the transcription of it as well. The aim of this study was to identify potential interaction interfaces between GATA-4 and the loop of UNC-45b and determine if the interactions are specific through computational models. Computational analysis suggests that the UCS loop is the preferential binding site of GATA-4 and that van der Waals packing is the primary method of binding.

## ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr. Odunuga, for his support and encouragement during my graduate career. His offer of a research position in his laboratory introduced me to this topic and paved the way for my education in the field. I would also like to thank my thesis committee, who were flexible and were willing to accept changes to my thesis proposal because of the COVID pandemic.

I would also like to thank my family, who have been a source of constant support and encouragement throughout the past eight years of my college career. Without them, I would not be where I am today.

## TABLE OF CONTENTS

ABSTRACT .....	i
ACKNOWLEDGEMENTS.....	ii
TABLE OF CONTENTS .....	iii
LIST OF FIGURES .....	v
LIST OF TABLES .....	vii
LIST OF ABBREVIATIONS .....	ix
CHAPTER 1 .....	1
Introduction .....	1
1.1 Background.....	1
1.2 UCS Proteins and UNC-45b .....	3
1.3 UNC-45 and Hsp90 Interaction in Myosin Folding .....	7
1.4 GATA-4 .....	9
1.5 UNC-45 and GATA-4 .....	11
1.6 Research Hypothesis .....	12
1.7 Objectives .....	13
CHAPTER 2 .....	14
Materials and Methods.....	14

2.1 UNC-45b and GATA-4 Model Generation.....	14
2.2 Docking of UNC-45b and GATA-4 .....	15
2.3 Obtaining Kd Values and Interaction Interfaces.....	15
2.4 Modeling of Interaction Interfaces .....	16
2.5 Mutagenesis.....	16
CHAPTER 3 .....	19
Results.....	19
3.1 Mouse UNC-45b Models.....	19
3.2 Mouse GATA-4 Model .....	21
3.3 Docking of UNC-45b and GATA-4 .....	23
3.4 Interaction Interface Modeling.....	26
3.5 Mutations .....	38
CHAPTER 4 .....	42
Discussion .....	42
Conclusion .....	48
REFERENCES.....	50
APPENDIX .....	60
A1. List of Programs and Websites.....	60
A2. Repeated Interactions Within Models .....	63
A3. The Terminal Loop of the UCS Domain.....	64
VITA .....	66

## LIST OF FIGURES

Figure 1: The structure of UNC-45 from <i>C. elegans</i> in complex with Hsp90.....	5
Figure 2: Visualization of UNC-45 forming a linear protein chain to stabilize and promote folding of myosin thick filaments .....	8
Figure 3: Visualization of the zinc-finger structure of GATA-4 from <i>Homo sapiens</i> .....	10
Figure 4: Flowchart of the experimental procedure. ....	18
Figure 5: Open Loop model.....	20
Figure 6: Closed Loop model.....	20
Figure 7: Model 2 of GATA-4.....	22
Figure 8: Model 14 of Open Loop and FL GATA-4.....	24
Figure 9: Model 25 of Closed Loop and FL GATA-4.....	25
Figure 10: Interaction of PRO 607 of UNC-45b and ALA 262 of GATA-4 in Model 15 of Open Loop.....	29
Figure 11: Interaction of PRO 607 of UNC-45b and ALA 262 of GATA-4 in Model 8 of Closed Loop .....	30

Figure 12: Interaction of GLN 604 of UNC45b and ARG 317 of GATA-4 in Model 19 of Open Loop.....	31
Figure 13: Interaction of GLN 604 of UNC45b and ARG 317 of GATA-4 in Model 3 of Closed Loop .....	32
Figure 14: Interaction of GLN 604 of UNC-45b to ARG 317 of GATA-4 in Model 23 of Open Loop.....	33
Figure 15: Interactions of GLU 590 of UNC-45b with LEU 260 and ILE 314 of GATA-4 in Model 22 of Open Loop .....	34
Figure 16: Interactions of GLU 590 of UNC-45b with LEU 260 and ILE 314 in GATA-4 in Model 23 of Closed Loop .....	35
Figure 17: Interactions of VAL 592 of UNC-45b with THR 279 and GLN 315 of GATA-4 in Model 16 of Open Loop.....	36
Figure 18: Interactions of VAL 592 of UNC-45b with ARG 259, THR 279, GLN 315, and THR 317 of GATA-4 in Model 23 of Closed Loop.....	37

## LIST OF TABLES

Table 1: RMSD calculation of the Closed Loop model .....	21
Table 2: RMSD calculation of the Open Loop model.....	21
Table 3: RMSD Calculation for FL GATA-4.....	22
Table 4: List of the shared interactions on the loop of UNC-45b between Models of Open Loop with FL GATA-4 and Closed Loop with FL GATA-4.....	27
Table 5: Amino acids of UNC-45b selected for visualization in ChimeraX and their interactions with FL GATA-4 .....	28
Table 6: List of mutations conducted .....	38
Table 7: Mutations conducted and their change in interacting residues shared between Open Loop/FL GATA-4 and Closed Loop/FL GATA-4 .....	39
Table 8: Repeat of mutations for GLN 604 and GLU 590 and the change in interacting residues.....	40
Table 9: The lowest Kd value for each type of mutation and if the model interacted with the mutation .....	40

Table 10: Lowest Kd values for non-mutated models and if they interacted with the non-mutated residues .....	41
Table 11: Comparison of the lowest Kd values and the average of the five lowest Kd values of mutated and non-mutated UNC-45b models .....	41
Table 12: List of Repeating Interactions on the Terminal Loop of the UCS Domain.....	64

## LIST OF ABBREVIATIONS

Alpha-MHC	Alpha-myosin heavy-chain
ARM	Armadillo repeat
ATP	Adenosine Triphosphate
GMQE	Global model quality estimation
Hop	Hsp70/90 organizing protein
Hsp	Heat shock protein
Kd	Dissociation constant
PDB ID	Protein data bank identification
PDB	Protein data bank
QMEAN	Quality model energy analysis
RMSD	Root-mean-square deviation
TPR	Tetratricopeptide repeat
UCS	UNC-45/CRO1/She4p
$\Delta G$	Gibbs free energy

## CHAPTER 1

### Introduction

#### 1.1 Background

Transcription factors and chaperones are two incredibly important protein groups in living organisms. Transcription factors are proteins that bind to specific DNA sequences and increase the rate of gene expression, allowing for specific, essential proteins to be transcribed when the cell needs them, such as during cell growth and development or during an immune response (Reviewed in Lambert et al. 2018). Chaperones are proteins that will bind to and help unfolded proteins fold into their correct forms, preventing misfolding and making the protein functional (Reviewed in Buchner 2019). Chaperones that bind to and fold transcription factors are of extreme importance, as they would exert control over the transcription factor and affect gene transcription. Such interactions would give insight to cellular development and would have implications for medical treatment.

One such chaperone protein is UNC-45b. UNC-45b is an essential chaperone protein in cellular development and is found in skeletal and cardiac

muscle (Price et al. 2002). It is necessary for the organization of myosin into thick filaments and the formation of contractile rings for cytokinesis (Barral et al. 1998; 2002). Mutations in UNC-45b can lead to serious defects, such as loss of cardiac function (Jansen et al. 1996) and failure during cardiogenesis (Chen et al. 2012).

The transcription factor it binds to is GATA-4. GATA-4 is a zinc-finger transcription factor protein that is necessary for cardiogenesis (Kuo et al. 1997, Molkentin et al. 1997) and septation of heart muscle (Zhou et al. 2017). It has been shown that high levels of GATA-4 leads to regeneration of cardiac tissue in neonatal mice after cryogenic injury (Mohammadi et al. 2017). The protein promotes transcription of several genes, including the gene for alpha-heavy chain myosin (Molkentin et al. 1994).

Research done by Chen et al. (2012) showed that there is a binding interaction between UNC-45b and GATA-4. This interaction is believed to be a chaperoning relationship between UNC-45b and GATA-4. Recently, Anderson (2019, Odunuga et al. 2020) showed evidence with pull-down assays that GATA-4 binds with the UCS domain of UNC-45b, and that the UCS domain is the only domain needed for binding to occur.

Modeling programs and software play an important role in the experimental process in biochemistry. The use of visualization software allows for proteins to be displayed and to view interactions in detail. Computational

systems can also run simulations and create models of potential experimental results. The information gained from such systems can help provide targets for further experimentation in the lab, helping to prevent trial and error and cutting down on operating costs (Cramer 2004). The focus of this research was to use modeling software to simulate and confirm that the UCS domain is where GATA-4 binds with high specificity and to determine the interaction interfaces between the two proteins. The list of programs used and their descriptions can be found in appendix A1 on Page 60.

## 1.2 UCS Proteins and UNC-45b

The UCS is a family of proteins (UNC-45/CRO1/She4p) that all function as chaperones for myosin, assisting in folding and assembly into thick filaments (Barral et al. 1998, 2002; Price et al. 2002; Lee et al. 2014; reviewed in Ni and Odunuga, 2015). UCS proteins in animals are specifically called UNC-45. Two variants of *unc-45* genes exist, *unc-45A* and *unc-45B* in humans and *unc-45a* and *unc-45b* in mice (Price et al. 2002). In mice, UNC-45a protein is expressed in organs such as the kidneys, liver, and lungs and is often referred to as general cell UNC-45 (GC UNC-45). UNC-45b is found only in skeletal and cardiac muscles; it is often referred to as striated muscle UNC-45 (SM UNC-45) (Price et al. 2002). The UNC-45b protein is essential to cellular development, as it is

necessary for myosin to form contractile rings during cytokinesis and to assemble into thick filaments (Barral et al. 1998; 2002). It is also essential in myosin and actin related processes in eukaryotic cells (Shi and Blobel 2010; Hutagalung et al. 2002).

UNC-45 was first identified when mutations causing a structural defect in myosin filaments in muscle cells of *C. elegans* were found (Epstein and Thomson 1974). Mutations in UCS proteins lead to serious defects in cytokinesis and myofibril organization in *C. elegans* (Epstein and Thomson 1974), endocytosis in *S. cerevisiae* (Jansen et al. 1996), paralysis and loss of cardiac function in zebrafish (Wohlgemuth et al. 2007), and failure during cardiogenesis in mice (Chen et al. 2012). It also leads to failure in sarcomeric organization (Donkervoort et al. 2020). The structure of UNC-45 is shown in Figure 1.

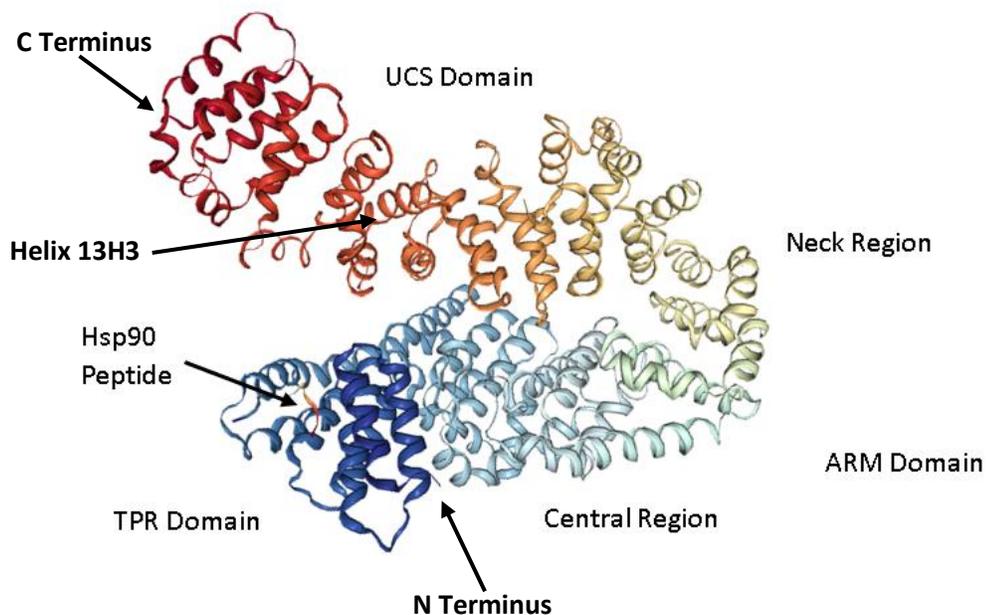


Figure 1: The structure of UNC-45 from *C. elegans* in complex with Hsp90; PBD ID: 4I2Z (Gazda et al. 2013)

UNC-45 proteins contain three domains: an N-terminal tetratricopeptide repeat domain (TPR), a central armadillo repeat (ARM), and a C-terminal UCS domain (Gazda et al. 2013) (Figure 1). The TPR domain binds preferentially to heat shock protein (Hsp) 90 but also to Hsp70 molecular chaperones (Barral et al. 2002; Srikakulam et al. 2008). The ARM domain contains a central and neck region (Gazda et al. 2013). The UCS domain is the binding site for myosin (Barral et al. 1998; 2002).

The TPR domain binds to Hsp90 and Hsp70 through the EEVD peptide sequence of the heat shock protein. This relationship has been well

characterized in the interaction of Hsp70 and Hsp90 with Hop (Hsp70/90 organizing protein) (Scheufler et al. 2000, Odunuga et al. 2003). The recruitment of Hsp proteins to the TPR domain helps support UNC-45b during myosin folding (Gazda et al. 2013).

The UCS domain is the known location where chaperone activity with myosin occurs, but the exact binding and folding mechanisms are unknown (Gazda et al. 2013, Bujalowski et al. 2014). The UCS domain has two areas of importance: a groove containing the  $\alpha$ -helix 13H3 and a large loop near the middle of the domain. The helix 13H3 is highly conserved and is believed to be the area of binding for myosin during folding (Gazda et al. 2013). The larger loop is a highly conserved motif among the UCS family that lines the crest of the groove (Gazda et al. 2013) and is believed to be a putative binding loop that helps in myosin binding (Gazda et al. 2013, Gaziova et al. 2020). Additionally, the grooves present in the UCS domain interact in myosin folding and mutations in the grooves alter the formation of sarcomeres (Gaziova et al. 2020). Mutations outside of the binding area can also alter the structure significantly, such as the naturally occurring mutation from arginine 805 to tryptophan (R805W) that is known to cause cataracts in humans (Gaziova et al. 2020).

### 1.3 UNC-45 and Hsp90 Interaction in Myosin Folding

While myosin binds to the UCS domain of UNC-45, the TPR-mediated interaction with Hsp90 is highly specific and is important in the folding of myosin. It has been shown that mutant UNC-45 that lacks a TPR domain will not interact with Hsp90 but will still interact with Hsp70, suggesting that the relationship between UNC-45 and Hsp90 is specifically mediated by the TPR domain (Barral et al. 2002).

Hsp90 can act as a chaperone for myosin on its own and when in complex with UNC-45. UNC-45 recruits Hsp90 to bind to its TPR domain to assist in myosin folding and maintaining the integrity of myosin assemblies (reviewed in Ni and Odunuga, 2015). Hsp90 also forms intermediate multimeric structures during myosin folding (Srikakulam and Winkelman, 2004).

The target of the interaction between SM UNC-45 and Hsp90 was further demonstrated by Liu et al. (2008) and Srikakulam et al. (2008). They showed that the complex that formed between SM UNC-45 and Hsp90 bound to unfolded myosin and promoted its folding. As a result, both SM UNC-45 and Hsp90 interact directly with myosin and serve as chaperones for myosin thick filament assembly (Du et al. 2008; Gaiser et al. 2011).

Furthermore, it has been shown that SM UNC-45 and Hsp-90 act to maintain integrity of myosin thick filaments and other myosin assemblies (Etard et al. 2008; Gazda et al. 2013). Existence of SM UNC-45 in multimeric form

(Figure 2) also supports its role in the organization of sarcomeric proteins. The protein was shown to form linear chains that can offer multiple binding sites for cooperative chaperones and client proteins (Gazda et al. 2013). These linear chains of UNC-45 may serve to stabilize thick filaments, inhibit the myosin ATP hydrolyzing activity, and prevent disorganization of the sarcomere during assembly (Nicholls et al. 2014). The UCS domain of SM UNC-45 alone also serves to inhibit the myosin powerstroke (Bujalowski et al. 2017).

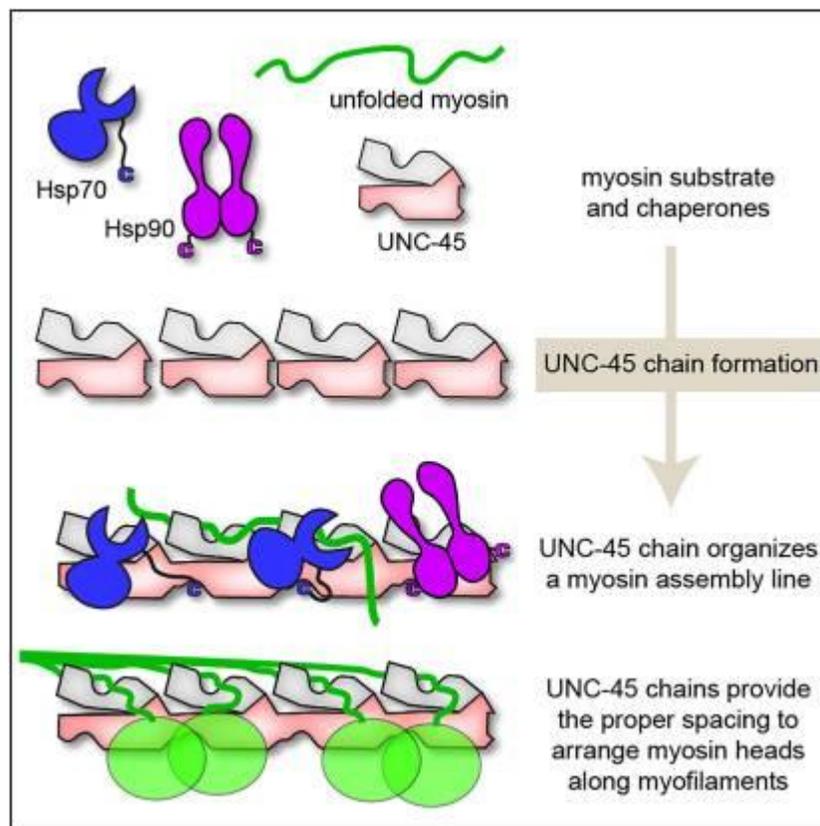
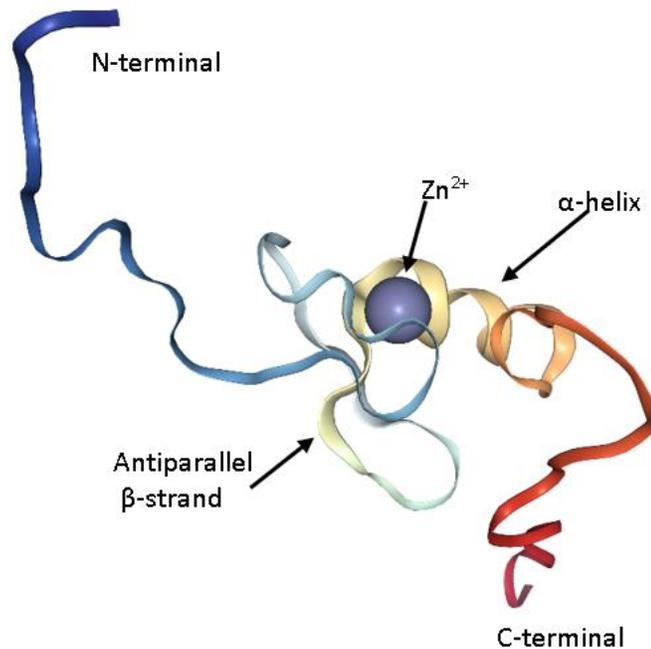


Figure 2: Visualization of UNC-45 forming a linear protein chain to stabilize and promote folding of myosin thick filaments (Gazda et al. 2013)

#### 1.4 GATA-4

GATA-4 is a cysteine zinc finger protein that acts as an essential transcription factor. It is important for transcription of several genes, including the alpha-myosin heavy-chain (alpha-MHC) gene in cardiac tissue (Molkentin et al. 1994). GATA-4 is part of the GATA family of transcription factors containing zinc fingers (Figure 3, Page 10). A zinc finger consists of four cysteine residues that bind to a zinc atom in a tetrahedral formation, two anti-parallel beta sheets, and a single  $\alpha$ -helix (Välimäki and Ruskoaho 2020). GATA proteins bind to the specific DNA sequence WGATAR, where W is an A or T and R is an A or G. The GATA binding motif is consistent in all forms of the GATA family and serves as the family name (Evans et al. 1988).

GATA-4 has been shown to be essential for cell differentiation and gene expression during cardiac development (Kuo et al. 1997, Molkentin et al. 1997, Dobrzycki et al. 2020). GATA-4 acts within a subset of progenitor cells known as the “second heart field,” the region of cells that is important for cardiac septation. The protein is responsible for two pathways that lead to hedgehog signaling and cell-cycle transition that result in septation of cardiac muscle (Zhou et al. 2017, Liu et al. 2019).



*Figure 3: Visualization of the zinc-finger structure of GATA-4 from Homo sapiens; PDB ID: 2M9W*

GATA-4 has also been shown to be important in regeneration of cardiac tissue. In neonatal mice, high levels of GATA-4 led to significant regeneration of tissue after cryogenic injury (Mohammadi et al. 2017). Studies have shown that mutations in GATA-4, particularly a 354 A>C mutation, have a significant association with congenital heart disease (Zhang et al. 2017). Mutations in GATA-4 have been associated with defects such as the misalignment of the aorta and pulmonary artery over the right ventricle, atrial septal defects, ventral septal defects, and others (Liu et al. 2019).

In addition to being a transcription factor in cardiac tissue, GATA-4 also plays a role in immune response. It is a transcription factor for tumor suppressors in several cancers, such as pancreatic cancer (Gong et al. 2018) and colorectal and gastric cancer (Akiyama et al. 2003). It is also involved in development of the gut and small intestine (Laverriere et al. 1994) and the liver (Geraud et al. 2017).

### 1.5 UNC-45 and GATA-4

The interaction between UNC-45 and GATA-4 was first discovered in an experiment performed by Chen et al. (2012). In the experiment, loss-of-function mutations of UNC-45b in mice were shown to be lethal during embryonic development, causing a failure in cardiac development and decreased cardiac contraction. This highlighted how important UNC-45 is in forming and accumulating myosin heavy chains and thick filaments (Chen et al. 2012).

However, Chen et al. (2012) also demonstrated that the transcription activity of GATA-4 was increased when expressed together with UNC-45b. This observation showed that UNC-45b had functions that were not just limited to myosin regulation at the protein level, but potentially regulation of its gene transcription via interaction with its transcription factor, GATA-4 (Chen et al. 2012). The study suggested that UNC-45b and GATA-4 have a chaperoning-client interaction and that UNC-45b may have an impact on the function of GATA-4.

Using ClusPro protein docking and pull-down assays, Anderson (2019, Odunuga et al. 2020) was able to determine that the UCS was the binding domain in UNC-45b for GATA-4. This finding was consistent with the chaperoning/client relationship suggested by Chen, et al. (2012). Additionally, Anderson was able to determine that the TPR domain is not needed for GATA-4 to bind to UNC-45b. The study done by Anderson (2019, Odunuga et al. 2020) supported the idea that UNC-45b and GATA-4 have a chaperone-client relationship and further suggested that UNC-45b exerts control over myosin expression at the genetic level. However, specific areas of interaction and interaction residues were not identified.

### 1.6 Research Hypothesis

Anderson (2019, Odunuga et al. 2020) found evidence that suggested that the binding site for GATA-4 to UNC-45b is the UCS domain, and Chen, et al. (2012) conducted experiments that showed that GATA-4 transcriptional activity increased several fold when coexpressed with UNC-45b. Together, these data suggest that the UCS domain, the same domain that is involved in myosin folding, has a chaperone relationship with GATA-4, the transcription factor that promotes the transcription of the alpha-heavy chain myosin gene in heart tissue (Molkentin et al. 1994).

Using the protein docking software program ClusPro and protein binding energy prediction software PRODIGY, it should be possible to identify where GATA-4 binds to on the UCS domain. The focus of this research was to use modeling software to simulate and confirm the UCS domain of UNC-45b as the module where GATA-4 binds with high specificity and to determine and characterize the interaction interfaces between the two proteins. Based on the above, the hypothesis of the study was that the large loop of the UCS domain of SM UNC-45 specifically binds to GATA-4. This hypothesis was tested based on the following objectives.

### 1.7 Objectives

1. Model full-length UNC-45b (FL UNC-45b) and validate the structure.
2. Model full-length GATA-4 (FL GATA-4) and validate the structure.
3. Dock the two protein models to identify domains of interaction and interaction interfaces.
4. Characterize the interaction interface through mutagenesis of amino acids.

## CHAPTER 2

### Materials and Methods

A flow chart graphically showing the experimental procedure can be found in Figure 4 (Page 18). A list of all programs and websites used, as well as a brief description of their functions and where they can be obtained, can be found in the List of Programs and Websites section in Appendix A1 on Page 60.

#### 2.1 UNC-45b and GATA-4 Model Generation

Sequences for UNC-45b and GATA-4 of *Mus musculus* were obtained through UniProtKB. These sequences were uploaded to SWISS-MODEL to generate the models to be used in this study. The models selected to be used for docking were chosen based on a combination of QMEAN and GMQE scores and their template models.

PyMol was used to visualize the models and conduct a RMSD analysis to validate the model structures. The align command was used to superimpose the model with its template and return a RMSD value. The model of GATA-4 had its

chain renamed to “B” in PyMol for the purpose of avoiding a conflict in analysis of interacting residues. UNC-45b models kept their original chain name of “A.”

## 2.2 Docking of UNC-45b and GATA-4

Docking of UNC-45b and GATA-4 was done in the ClusPro webserver. The PDB files of the receptor (UNC-45b) and the ligand (GATA-4) were uploaded as a job to ClusPro. The jobs were run on the CPU server. No chains were specified for interaction as each model contained only one chain, and no advanced options were selected for the job.

## 2.3 Obtaining Kd Values and Interaction Interfaces

Models generated through ClusPro docking were downloaded and placed into a zip file. The zip file was uploaded to the PRODIGY webserver to obtain Kd values of the models. Interactor 1 was listed as “A” for UNC-45b, and Interactor 2 was listed as “B” for GATA-4. The temperature was left at 25.0 °C. The job was submitted, and PRODIGY returned a table listing various values, including Kd and  $\Delta G$  values, for each model. The table was copied into an excel file for analysis.

Interaction interfaces were obtained by uploading individual models to the PRODIGY server. Interactor 1 was listed as “A” for UNC-45b, Interactor 2 listed as “B” for GATA-4, and temperature was left as 25.0 °C. PRODIGY returned a text file listing all interaction residues between UNC-45b and GATA-4 in that model. Each text file was copied into Excel.

For ease of viewing and analysis, the Ablebits addon for Excel was downloaded in order to consolidate the different spreadsheets of interaction interfaces into a single sheet comparing each model.

#### 2.4 Modeling of Interaction Interfaces

Modeling of interaction interfaces was done in ChimeraX, a visualization software. Models of UNC-45b docked with GATA-4 that contained interactions of interest were opened in ChimeraX. The interacting amino acids were edited to display in ball-and-stick form and were labeled. The distance function was used to measure distances between atoms of interacting amino acids to determine the type of bonding interaction. The distances were labeled, and a picture of the interaction was taken.

#### 2.5 Mutagenesis

Mutagenesis was accomplished through the use of the Mutagenesis Wizard in PyMol. Mutations were only done on UNC-45b and only one mutation

was done per model. The model of UNC-45b was opened in PyMol, the residue to be mutated was selected, and the desired mutation was chosen in the Mutagenesis Wizard and applied to the model. The model was renamed and saved.

The mutated models were docked with GATA-4 in ClusPro. The resulting models were analyzed in PRODIGY to obtain their Kd values and interaction interfaces. Interactions with the mutated residue were recorded and compared to interactions in models of non-mutated docking.

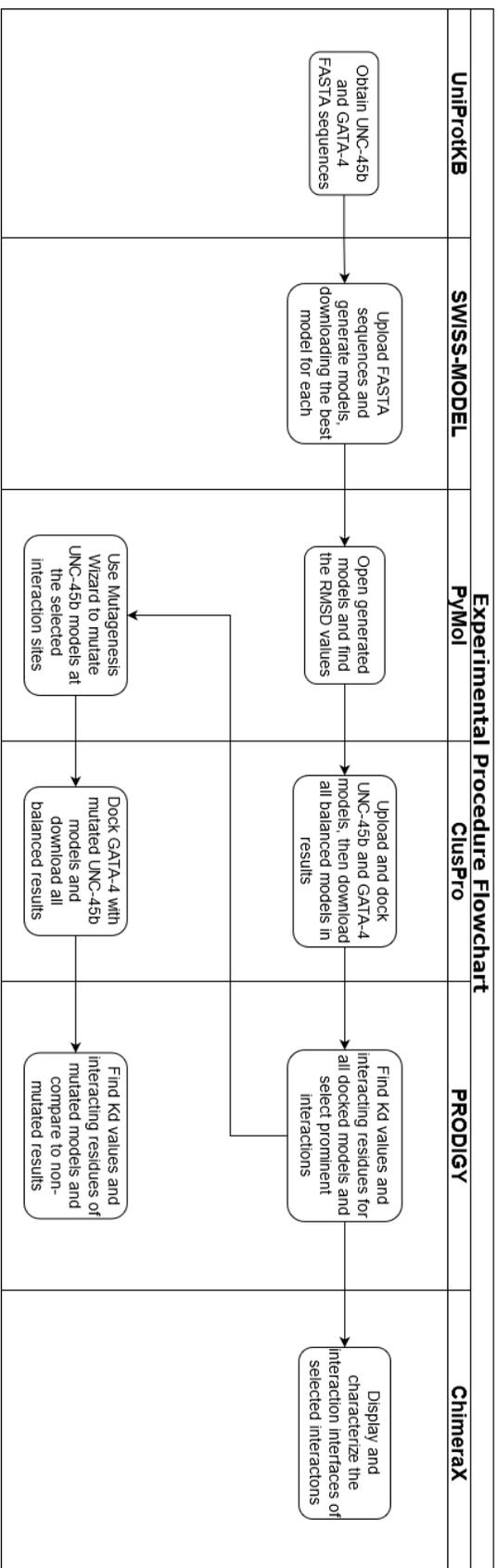


Figure 4: Flowchart of the experimental procedure. Shows the steps taken and what website/program was utilized.

## CHAPTER 3

### Results

#### 3.1 Mouse UNC-45b Models

The protein sequence for *Mus musculus* UNC-45b was obtained from the UniProtKB website. The sequence was uploaded to SWISS-MODEL to generate models. Nine models were generated. Model 1 and Model 2 were selected, with Model 1 using the PDB ID 4I2Z (UNC-45b in *C. elegans*) and Model 2 using 3NOW (UNC-45b in *D. melanogaster*) as templates. Model 1 was renamed as Open Loop (Figure 4) and Model 2 was renamed as Closed Loop (Figure 5), due to the orientation of the large loop. The Closed Loop model had better QMEAN and GMQE scores, but the Open Loop model used a template that is considered to be a more accurate representation of UNC-45b because the loop is extended and the TPR domain is present.

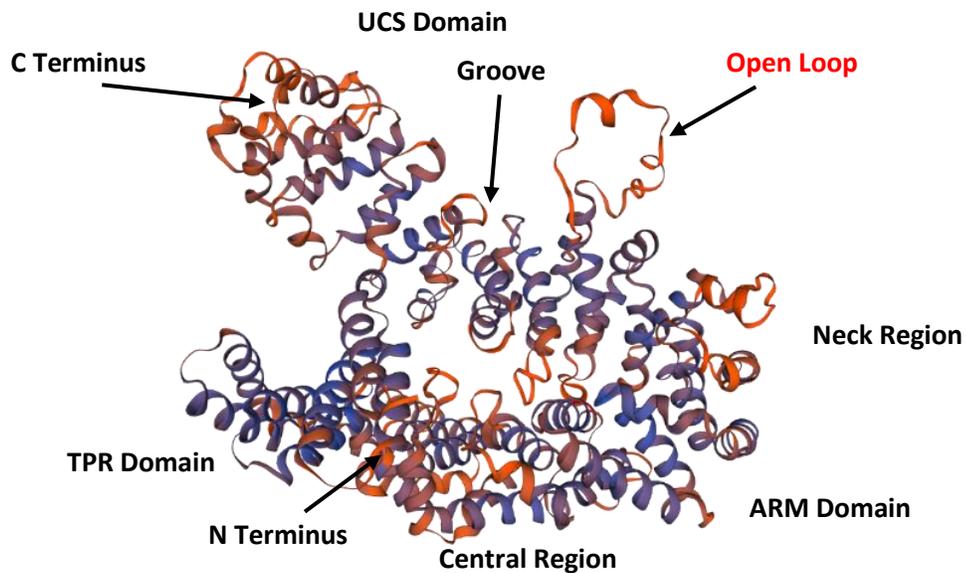


Figure 5: Open Loop model using Template 4I2Z from *C. elegans*, viewed in SWISS-MODEL

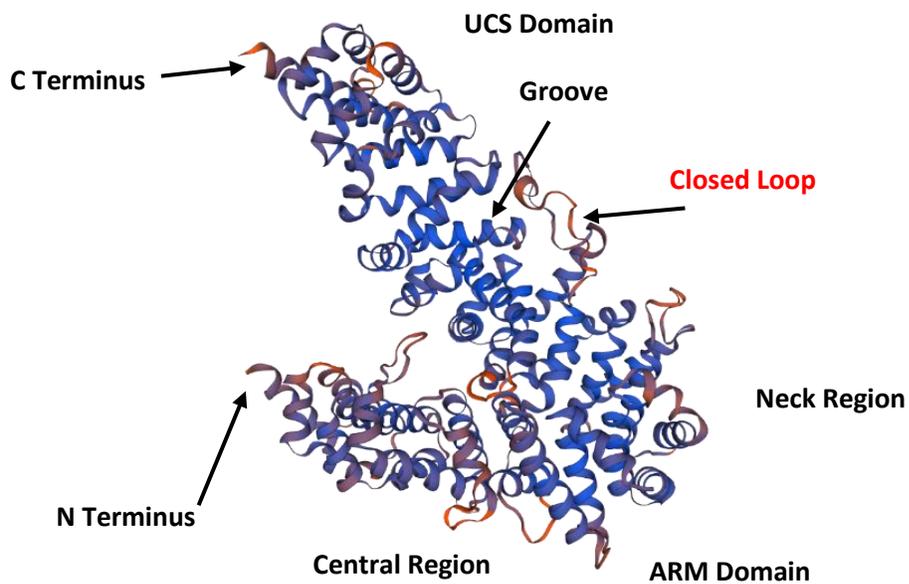


Figure 6: Closed Loop model using Template 3NOW from *D. melanogaster*, viewed in SWISS-MODEL

The two model structures were confirmed by superimposing each model with its template in PyMol and calculating RMSD values. This was accomplished using the align command.

*Table 2: RMSD calculation of the Open Loop model*

Confirmation of Open Loop and 4I2Z		
Cycle	Atoms Excluded	RMSD
0	0	1.862
1	313	0.92
2	354	0.63
3	231	0.55
4	133	0.52
5	59	0.511

*Table 1: RMSD calculation of the Closed Loop model*

Confirmation of Closed Loop and 3NOW		
Cycle	Atoms Excluded	RMSD
0	0	1.272
1	259	0.44
2	204	0.18
3	251	0.14
4	193	0.12
5	119	0.115

The RMSD for both models, before considering atoms that were outliers, was less than two angstroms. This shows that the models were very similar in structure to the template. After removing outlier atoms, such as different sidechains between the model and the template, the RMSD dropped further, validating the models as acceptable models.

### 3.2 Mouse GATA-4 Model

The protein sequence for *Mus musculus* GATA-4 was obtained from the UniProtKB website. The sequence was uploaded to SWISS-MODEL to generate models. 7 models were generated. Model 2 was selected due to it containing both zinc-finger motifs, containing two zinc ions, and having the best QMEAN

and GMQE scores of all models generated. It used PDB ID 4HC9 (GATA-3 in *Homo sapiens*) as a template. This model was renamed as FL GATA-4.

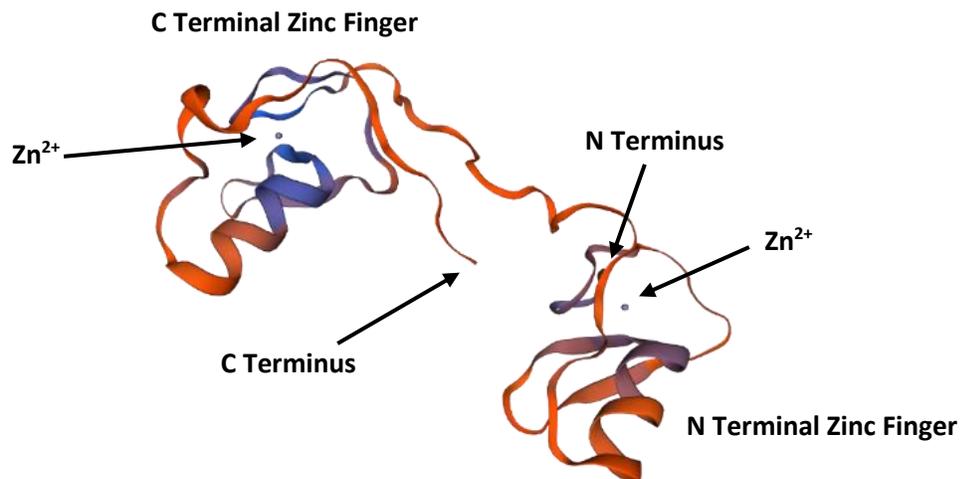


Figure 7: Model 2 of GATA-4 based on Template 4HC9 from *Homo sapiens*; viewed in SWISS-MODEL

The structure of FL GATA-4 was confirmed by superimposing the model with its template in PyMol and calculating the RMSD value with the align command.

Table 3: RMSD Calculation for FL GATA-4

Confirmation of FL GATA-4 and 4HC9			
Cycle	Atoms Excluded	RMSD	
0	0	0.218	
1	10	0.11	
2	39	0.09	
3	20	0.08	
4	12	0.08	
5	8	0.079	

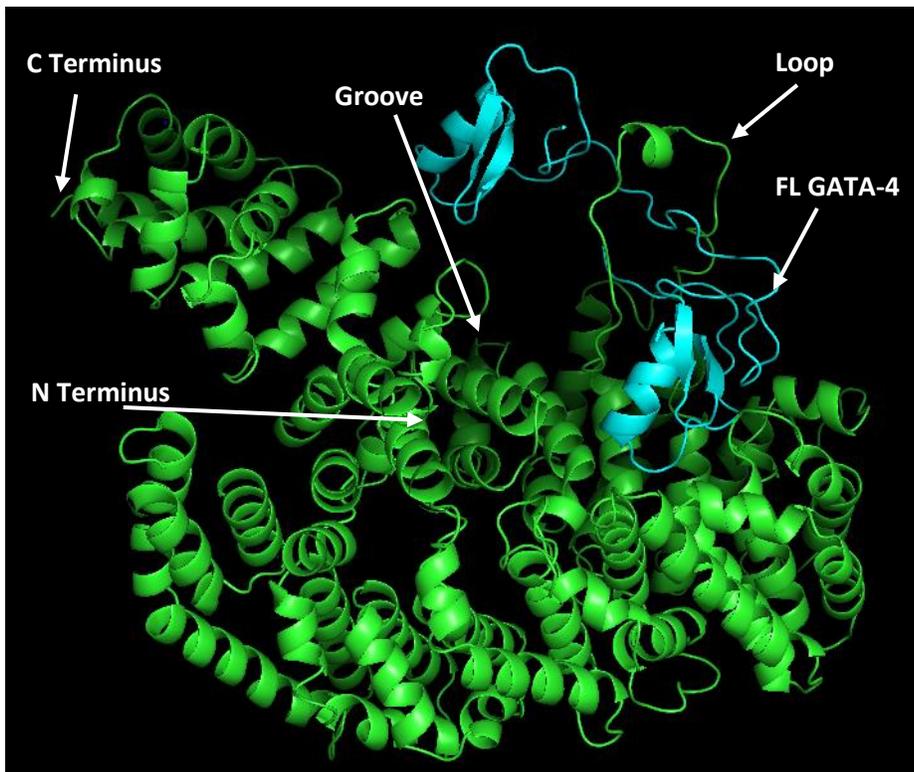
The RMSD values in Table 3 for FL GATA-4 were excellent, showing that it matches the template extremely well. Before outliers were excluded, the RMSD value was 0.218 angstroms, an acceptable number. But exclusion of outliers led to extremely low RMSD values, further showing how well the model matches to its template.

### 3.3 Docking of UNC-45b and GATA-4

Because ClusPro does not rename chains that are used in construction of its docking models, the FL GATA-4 was edited in PyMol to reidentify the sequence chain to “B”. UNC-45b models kept their chain ID of “A”. This was necessary for saving time when doing Kd calculations and finding interaction residues.

The PDB files of the receptor (UNC-45b) and ligand (GATA-4) were uploaded to ClusPro. The CPU server was used. Because the models only contained one chain each, no chains were specified. No advanced options were used. Two separate jobs were run on ClusPro for each combination of receptor and ligand – Open Loop/FL GATA-4 and Closed Loop/FL GATA-4. Each job generated 30 models. The balanced models were selected to be analyzed, as they took a balanced approach to docking instead of favoring electrostatic or hydrophobic interactions. All balanced models were downloaded and put into zip files to be analyzed in the PRODIGY webserver.

The PRODIGY protein-protein function was used. The zip file containing docked models from a job was uploaded. Interacting ID chains were specified, A for UNC-45b, and B for GATA-4. The temperature was left at 25 degrees Celsius. The job was submitted, and results recorded. The model that had the lowest Kd value was selected as being the best representation of the protein docking for that job.



*Figure 8: Model 14 of Open Loop and FL GATA-4 resulted in the lowest Kd value of 2.1E-16 at 25 °C*

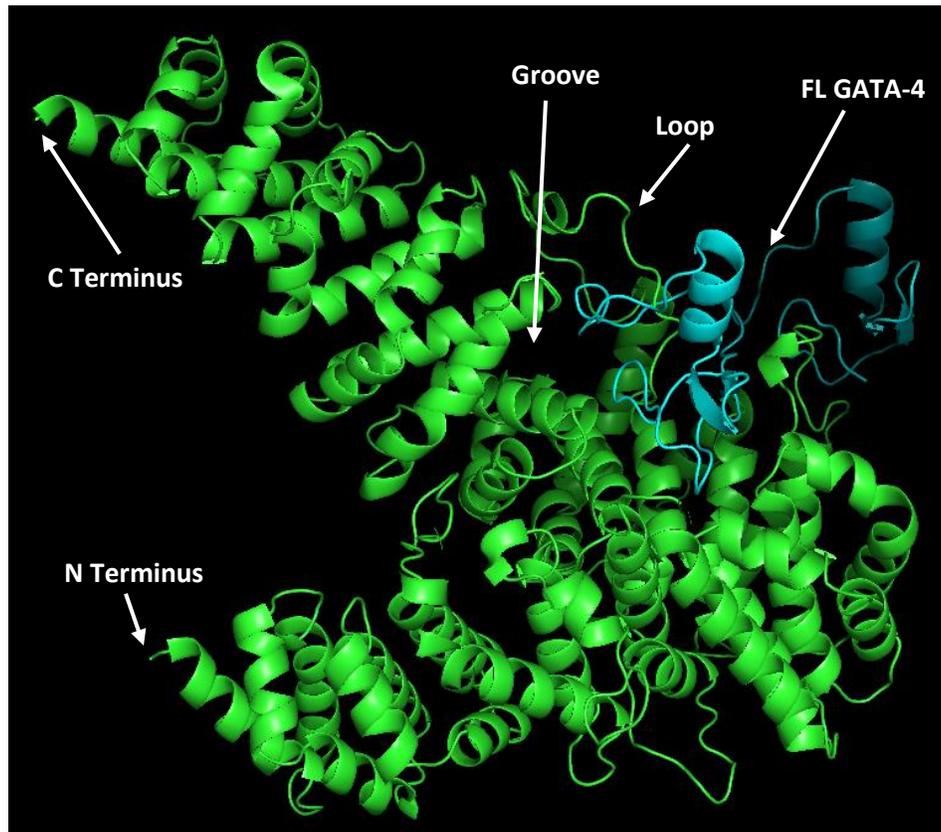


Figure 9: Model 25 of Closed Loop and FL GATA-4 resulted in the lowest  $K_d$  value of  $8.1E-15$  at  $25^\circ\text{C}$

Both Figure 7 and Figure 8 had GATA-4 interacting with the UCS domain of UNC-45b, and both were interacting with the large loop in the central area of the UCS domain.

### 3.4 Interaction Interface Modeling

When a single model is run through the PRODIGY webserver, a list of interaction residues is generated. All interacting residues for all models were compiled into Excel. To narrow down potential specific interactions, a list of all interactions that occurred in more than one model was generated. For the purposes of this experiment, this list of interactions was focused on the large loop of the UCS domain, as it is believed to be a putative binding loop.

The set of models for Open Loop UNC-45b with Full Length GATA-4 had 226 interactions with the loop that occurred in at least two models. The set with Closed Loop UNC-45b and Full Length GATA-4 had 97. To narrow the list down further, interactions that occurred at least twice in each interaction scenario were compiled into Table 4.

Table 4: List of the shared interactions on the loop of UNC-45b between Models of Open Loop with FL GATA-4 and Closed Loop with FL GATA-4

UNC-45b	Interacting Acid of GATA-4	Open Loop/FL GATA-4 models where interaction is present	Closed Loop/GATA-4 models where interaction is present
VAL 588	LEU 260	2, 15, 19, 22, 26	8, 25
VAL 588	ARG 264	1, 22	2, 23
VAL 588	THR 277	15, 19, 22	2, 23
LYS 589	LEU 260	1, 2, 22, 28	8, 25
LYS 589	GLY 313	19, 22	2, 23
LYS 589	ILE 314	2, 15, 19, 22, 28	23, 25
LYS 589	THR 316	2, 11, 15, 21, 22, 26, 28	8, 25
GLU 590	LEU 260	1, 2, 15, 19, 22	8, 23
GLU 590	ILE 314	1, 2, 19, 21, 22	23, 25
VAL 591	GLN 315	2, 16, 19, 21, 22	8, 25
VAL 591	THR 316	2, 19, 21, 22, 26, 28	8, 25
VAL 592	ARG 259	0, 25, 28	8, 23
VAL 592	LEU 260	1, 16	8, 23
VAL 592	THR 279	2, 16, 22	8, 23
VAL 592	ILE 314	1, 2, 16	8, 23
VAL 592	GLN 315	1, 16, 21	8, 23
VAL 592	THR 316	1, 16, 21, 27	8, 23
VAL 592	ARG 317	2, 15, 20, 21, 22, 25	8, 23, 25
PRO 593	ARG 259	11, 28	8, 23
PRO 593	ARG 317	2, 20, 21, 22, 26	8, 23
GLU 594	ARG 317	2, 16, 21, 22	5, 8, 11, 23
GLN 597	ARG 259	11, 16	11, 18
LEU 598	HIS 234	15, 28	18, 23
GLN 604	ARG 317	19, 20, 26	3, 17
PRO 607	ALA 262	15, 28	8, 23

Four amino acids of FL UNC-45b and their interactions with FL GATA-4 were selected from Table 4 for visualization and further analysis through mutagenesis. Those amino acids and their contacts with FL GATA-4 are shown in Table 5.

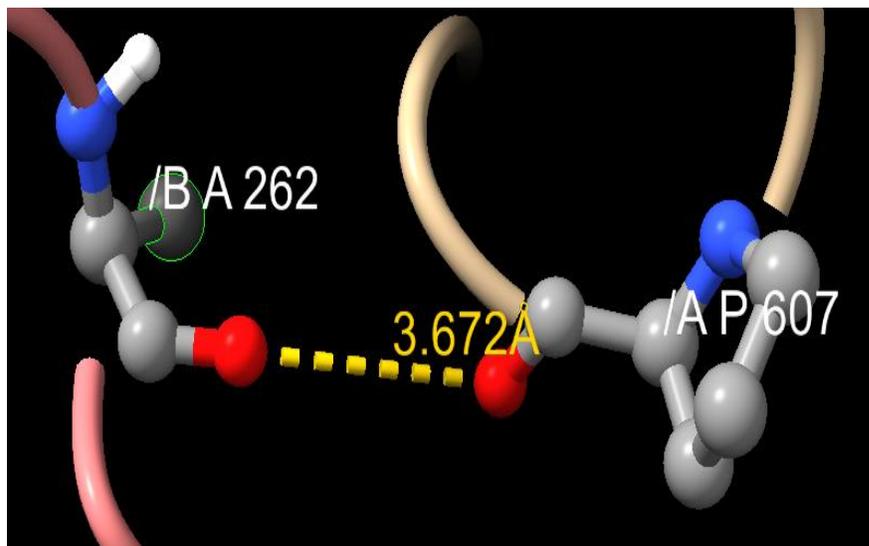
*Table 5: Amino acids of UNC-45b selected for visualization in ChimeraX and their interactions with FL GATA-4*

<b>UNC-45b</b>	<b>Interacting Acid of GATA-4</b>	<b>Open Loop/FL GATA-4 models where interaction is present</b>	<b>Closed Loop/GATA-4 models where interaction is present</b>
<b>GLU 590</b>	LEU 260	1, 2, 15, 19, 22	8, 23
<b>GLU 590</b>	ILE 314	1, 2, 19, 21, 22	23, 25
<b>VAL 592</b>	ARG 259	0, 25, 28	8, 23
<b>VAL 592</b>	LEU 260	1, 16	8, 23
<b>VAL 592</b>	THR 279	2, 16, 22	8, 23
<b>VAL 592</b>	ILE 314	1, 2, 16	8, 23
<b>VAL 592</b>	GLN 315	1, 16, 21	8, 23
<b>VAL 592</b>	THR 316	1, 16, 21, 27	8, 23
<b>VAL 592</b>	ARG 317	2, 15, 20, 21, 22, 25	8, 23, 25
<b>GLN 604</b>	ARG 317	19, 20, 26	3, 17
<b>PRO 607</b>	ALA 262	15, 28	8, 23

Of the interactions in Table 5, the interaction of GLN 604 in UNC-45b and ARG 317 in GATA-4 was of special consideration, as it had been identified as part of a semi-conserved sequence on the loop believed to play a role in the binding of myosin (Gaziova et al. 2020). The interaction of PRO 607 in UNC-45b

with ALA 262 in GATA-4 was also considered as it was close in proximity to the conserved sequence in question.

Prodigy does not determine the type of interactions that occurred. The interfaces of the interactions between PRO 607 and ALA 262 (Figures 9 and 10) and the interactions of GLN 604 and ARG 317 (Figures 11 and 12) from both Open Loop and Closed Loop models were modeled in ChimeraX in order to visualize the binding and determine what types of bond were involved. These interactions were compared to see if there were similar binding interactions between the models.



*Figure 10: Interaction of PRO 607 of UNC-45b and ALA 262 of GATA-4 in Model 15 of Open Loop/FL GATA-4. Shows a potential van der Waals interaction.<sup>1</sup>*

---

<sup>1</sup> The broken yellow lines represent distances between atoms. They are not necessarily bonds but do represent potential bonds and interactions.

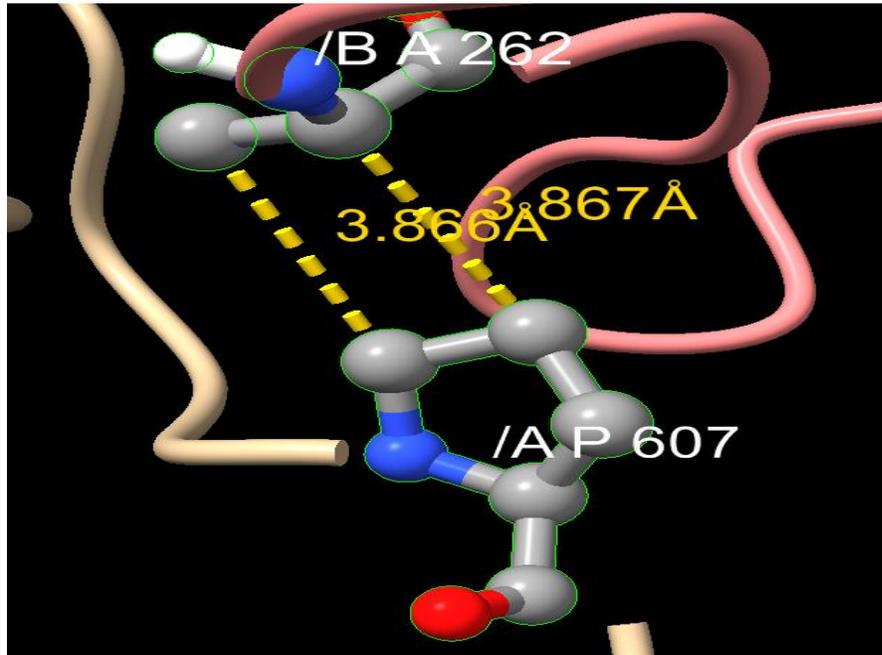
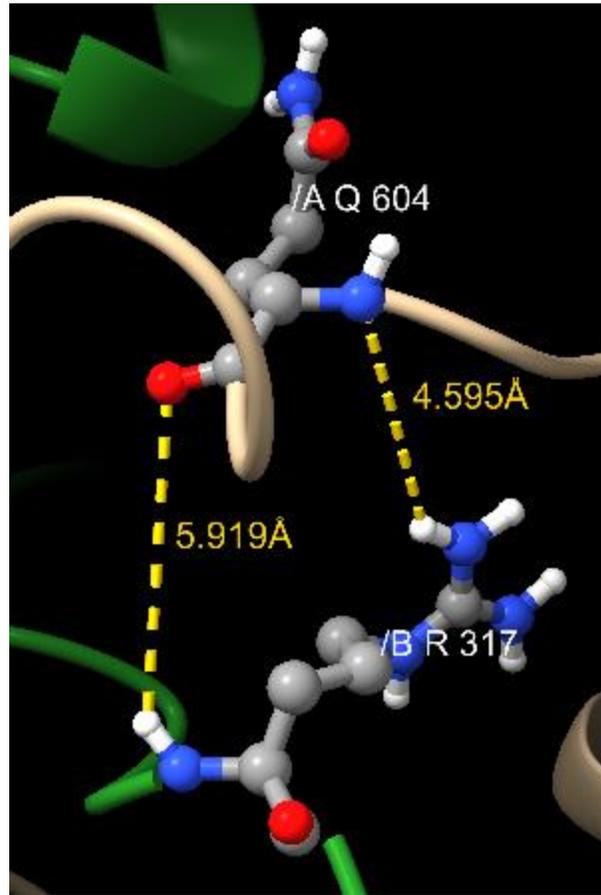


Figure 11: Interaction of PRO 607 of UNC-45b and ALA 262 of GATA-4 in Model 8 of Closed Loop/FL GATA-4. Shows potential van der Waals interactions.<sup>2</sup>

The yellow lines are the closest distances between atoms of each residue. Due to the distance and lack of hydrogens, the interactions in both Figure 9 and Figure 10 are the result of van der Waals forces. The orientation of the residues in the interaction are not similar between the two figures.

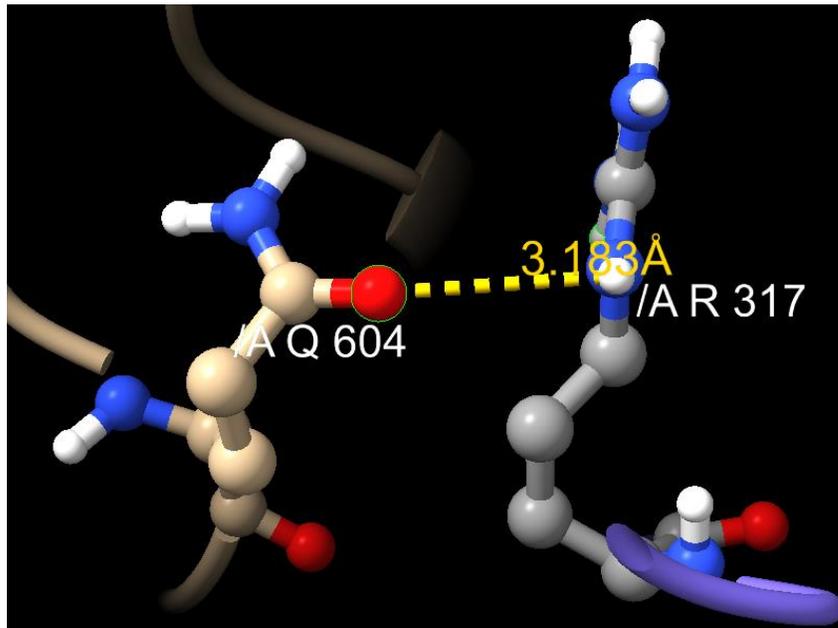
<sup>2</sup> The broken yellow lines represent distances between atoms. They are not necessarily bonds but do represent potential bonds and interactions.



*Figure 12: Interaction of GLN 604 of UNC45b and ARG 317 of GATA-4 in Model 19 of Open Loop/FL GATA-4. Shows potential van der Waals interactions.<sup>3</sup>*

---

<sup>3</sup> The broken yellow lines represent distances between atoms. They are not necessarily bonds but do represent potential bonds and interactions.

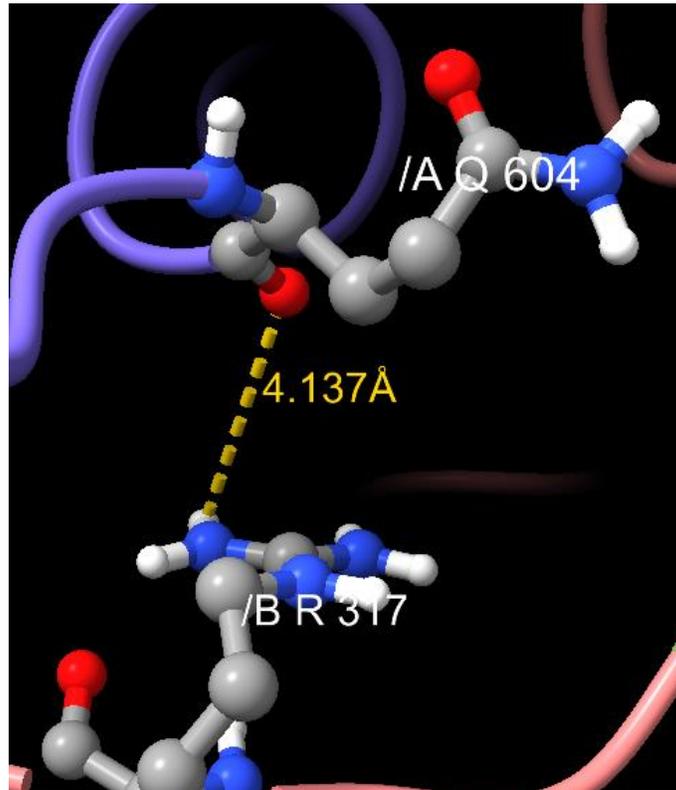


*Figure 13: Interaction of GLN 604 of UNC45b and ARG 317 of GATA-4 in Model 3 of Closed Loop/FL GATA-4. Shows a moderate hydrogen bond.<sup>4</sup>*

In the Figure 11, the interaction between GLN 604 and ARG 317 is extremely distant, so these are van der Waals forces. In the Figure 12, the sidechain of GLN 604 is oriented more to interact with ARG 317. ChimeraX does not recognize this as a hydrogen bond as the default setting only classifies bonds within three Angstroms as hydrogen bonds, but the distance is close enough to be considered a moderate hydrogen bond.

---

<sup>4</sup> The broken yellow lines represent distances between atoms. They are not necessarily bonds but do represent potential bonds and interactions.



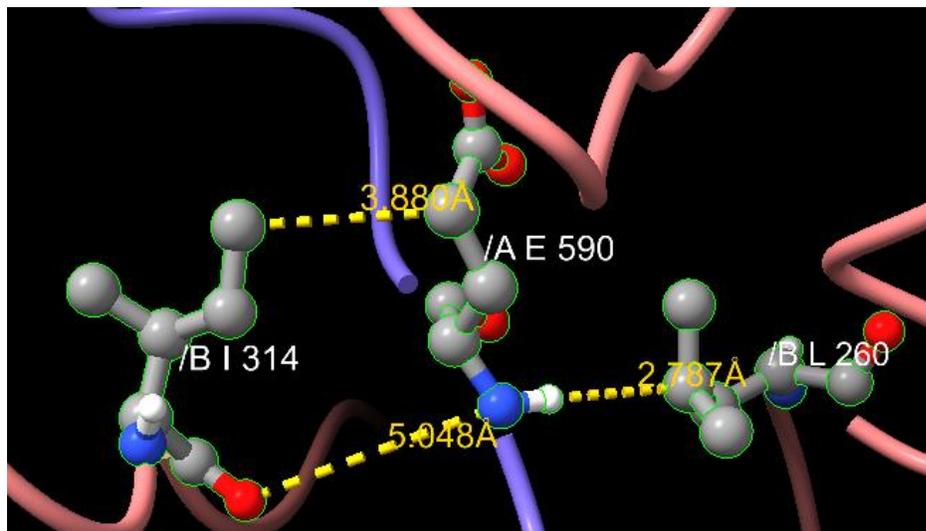
*Figure 14: Interaction of GLN 604 of UNC-45b to ARG 317 of GATA-4 in Model 23 of Open Loop/FL GATA-4. Shows a potential weak hydrogen bond.<sup>5</sup>*

However, in Figure 13, another Open Loop model that showed a GLN 604 – ARG 317 interaction, a different conformation of the arginine sidechain allowed for a possible weak hydrogen bond to form between the residues.

---

<sup>5</sup> The broken yellow lines represent distances between atoms. They are not necessarily bonds but do represent potential bonds and interactions.

Other interactions that were modeled were the interactions with GLU 590 of UNC-45b, as interactions with LEU 260 and ILE 314 were shared in both Open Loop (Figure 14) and Closed Loop (Figure 15) models, and GLU 590 was one of the residues with the highest count of total interactions. The other interaction modeled was interactions with VAL 592 of UNC-45b (Figure 16 and Figure 17). VAL592 had the highest number of shared interactions between Open Loop and Closed Loop, interacting with ARG 259, LEU 260, THR 279, ILE 314, GLN 315, THR 316, and ARG 317 of GATA-4 in at least two models.

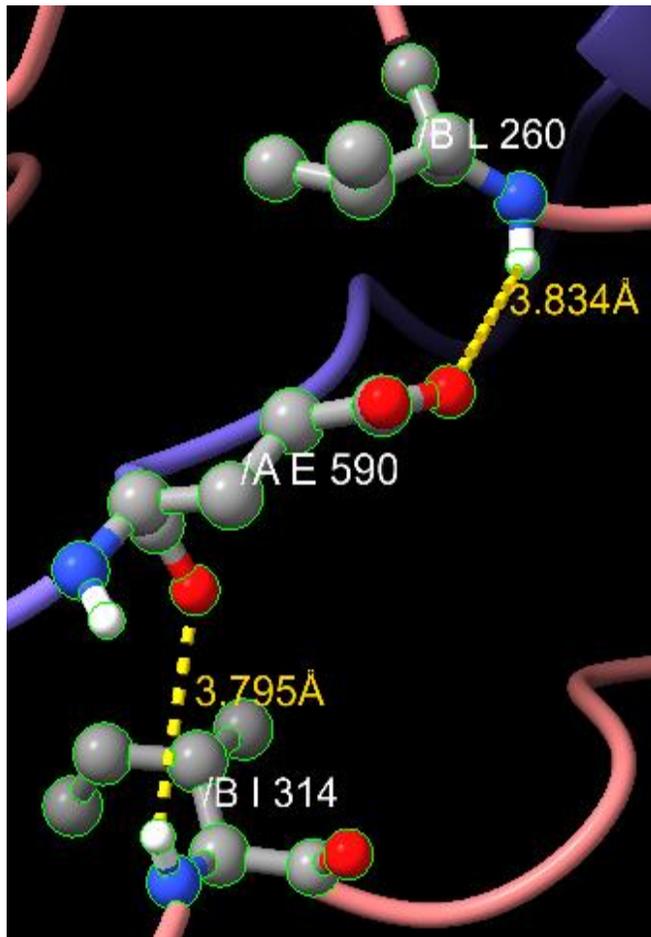


*Figure 15: Interactions of GLU 590 of UNC-45b with LEU 260 and ILE 314 of GATA-4 in Model 22 of Open Loop/FL GATA-4. Shows van der Waals interactions.<sup>6</sup>*

There were no hydrogen bonds present in Figure 14. Interactions were due to van der Waals forces.

---

<sup>6</sup> The broken yellow lines represent distances between atoms. They are not necessarily bonds but do represent potential bonds and interactions.

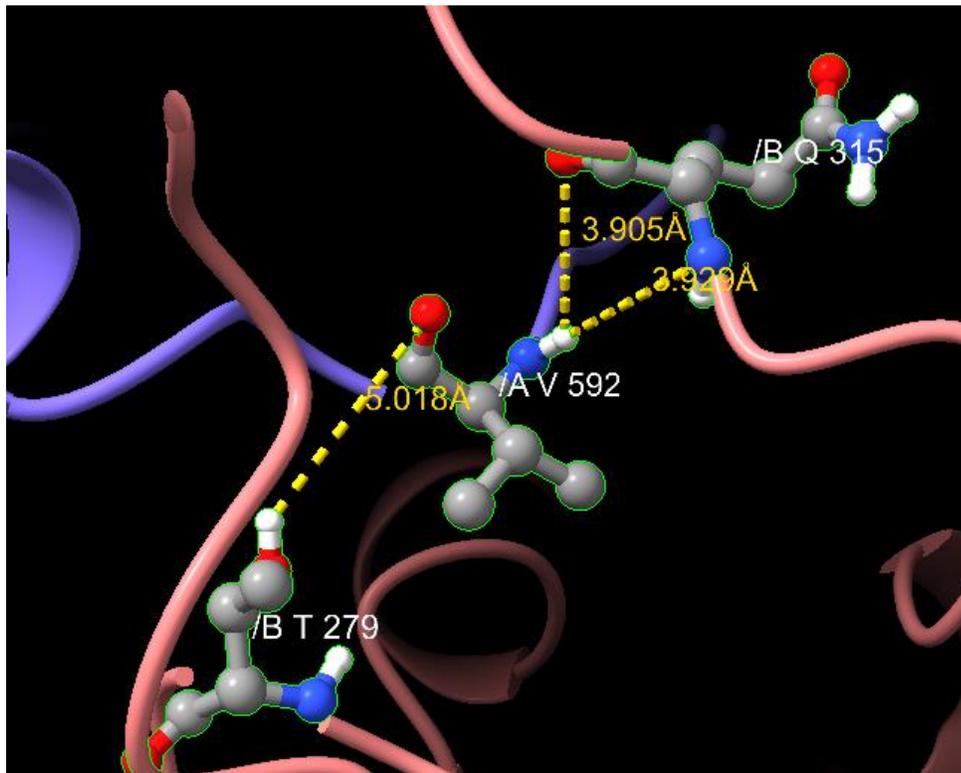


*Figure 16: Interactions of GLU 590 of UNC-45b with LEU 260 and ILE 314 in GATA-4 in Model 23 of Closed Loop/FL GATA-4. Shows two moderate hydrogen bonds.<sup>7</sup>*

In Figure 15, GLU 590 showed two moderate hydrogen bonds with LEU 260 and ILE 314.

---

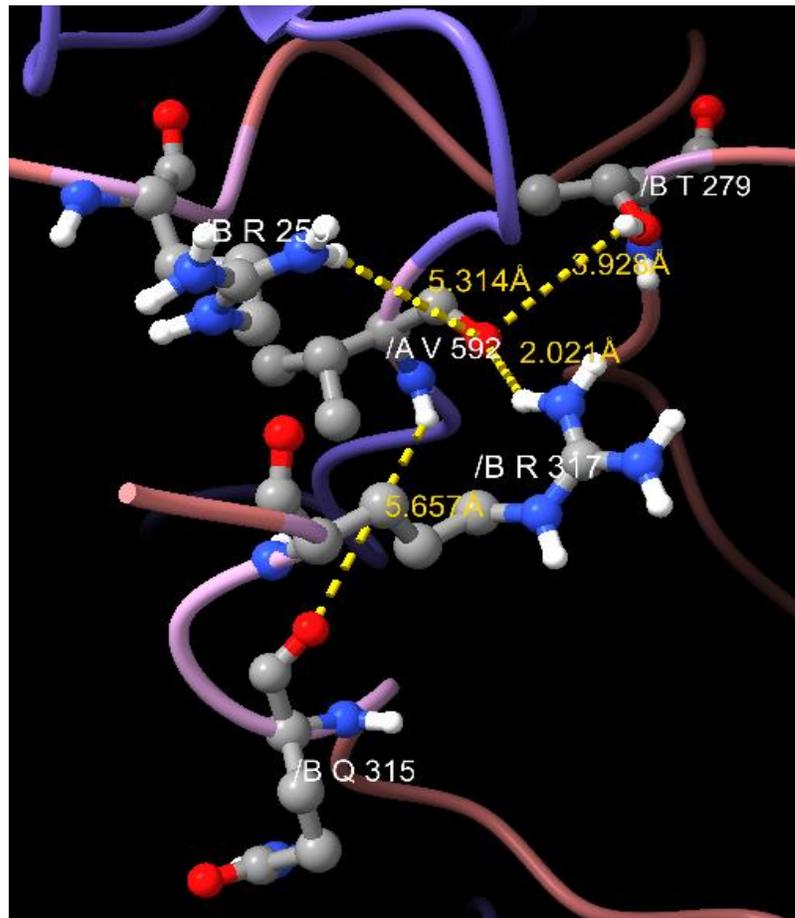
<sup>7</sup> The broken yellow lines represent distances between atoms. They are not necessarily bonds but do represent potential bonds and interactions.



*Figure 17: Interactions of VAL 592 of UNC-45b with THR 279 and GLN 315 of GATA-4 in Model 16 of Open Loop/FL GATA-4. Shows two potential, moderate hydrogen bonds and a potential, weak hydrogen bond.<sup>8</sup>*

Figure 16 shows two possible moderate hydrogen bonds between VAL 592 and GLN 315, and a possible very weak hydrogen bond with THR 279. LEU 260, ILE 314, and THR 316 also interact with Val 592, but are not shown. They interact with the sidechain of VAL 592 in the form of hydrophobic interactions.

<sup>8</sup> The broken yellow lines represent distances between atoms. They are not necessarily bonds but do represent potential bonds and interactions.



*Figure 18: Interactions of VAL 592 of UNC-45b with ARG 259, THR 279, GLN 315, and THR 317 of GATA-4 in Model 23 of Closed Loop/FL GATA-4. Shows a strong hydrogen bond with ARG 317, a moderate hydrogen bond with THR 279, and van der Waals interactions with ARG 259 and VAL 592.<sup>9</sup>*

Figure 17 shows a hydrogen bond with ARG 317 and a moderate hydrogen bond with THR 279. There are long distances between VAL 592 and ARG 259 and GLN 315, but they could be van der Waals interactions. LEU 260,

<sup>9</sup> The broken yellow lines represent distances between atoms. They are not necessarily bonds but do represent potential bonds and interactions.

ILE 314, THR 316, and ARG 317 also all have interactions with VAL 592, however they were all more than 6 angstroms away. These were likely van der Waals or hydrophobic interactions.

### 3.5 Mutations

Mutations of the modeled amino acids were done in PyMol with the Wizard mutagenesis tool. Mutations were conducted on the UNC-45b models. The mutations were designed to be similar to the original amino acid in order to test for specificity in the interactions.

*Table 6: List of mutations conducted*

UNC-45b Original Acid	Result of Mutagenesis
PRO 607	ALA 607
GLN 604	ASN 604
GLU 590	ASP 590
VAL 592	ALA 592

The mutations listed in Table 6 were done on both Open Loop and Closed Loop models of UNC-45b. Mutating the already docked models would not

change the interaction residue results or Kd values obtained through PRODIGY. The models were docked in ClusPro with Full Length GATA-4. Models were then run in the PRODIGY webserver. Kd values and interacting residues with the mutated amino acid were recorded.

The PRODIGY results looked at interacting amino acids with the mutated residue. Similar to the modeling results, interactions that were shared with at least two models in both the Open Loop mutated model with GATA-4 and two models of Closed Loop mutated model with GATA-4 were recorded in Table 7.

*Table 7: Mutations conducted and their change in interacting residues shared between Open Loop/FL GATA-4 and Closed Loop/FL GATA-4*

Mutation	Shared Interactions before Mutation	Shared Interactions After Mutation
PRO 607 to ALA 607	ALA 262	ALA 262
GLN 604 to ASN 604	ARG 317	None
GLU 590 to ASP 590	LEU 260, ILE 314	LEU 260, THR 277, ILE 314, GLN 315
VAL 592 to ALA 592	ARG 259, LEU 260, THR 279, ILE 314, GLN 315, THR 316, THR 317	None

Because the mutations for GLU 590 to ASP and GLN 604 to ASN had little change to their sidechain structure and inconclusive interaction differences,

models were mutated again. Both GLU 590 and GLN 604 were mutated to ALA, as shown in Table 8.

*Table 8: Repeat of mutations for GLN 604 and GLU 590 and the change in interacting residues shared between Open Loop/FL GATA-4 and Closed Loop/FL GATA-4*

Mutation	Shared Interactions before Mutation	Shared Interactions After Mutation
GLN 604 to ALA 604	ARG 317	None
GLU 590 to ALA 590	LEU 260, ILE 314	LEU 260

The Kd values of the mutated models and if they interacted with the mutated residue were recorded in Table 9. Table 9 was compared to Table 10, which contains the lowest Kd values of non-mutated models and if they interacted with the non-mutated residue.

*Table 9: The lowest Kd value for each type of mutation done, and if the lowest Kd value model had an interaction with the mutated residue*

UNC-45b Acid	Open Loop/FL GATA-4		Closed Loop/FL GATA-4	
	Lowest Kd Value	Interact with mutated residue?	Lowest Kd Value	Interact with mutated residue?
PRO 607 to ALA	1.3E-16	Yes	5.5E-15	No
GLN 604 to ASN	6E-17	Yes	5.3E-15	No
GLN 604 to ALA	6.9E-17	Yes	7.1E-15	No
VAL 592 to ALA	2.6E-15	Yes	4.3E-13	Yes
GLU 590 to ASP	4E-14	Yes	3.9E-15	Yes
GLU 590 to ALA	2.1E-15	Yes	4.8E-14	Yes

Table 10: Lowest Kd values for non-mutated models and if they interacted with the non-mutated residues

Non-Mutated Models		Interact With?			
Model System	Lowest Kd (M) at 25.0 °C	PRO 607	GLN 604	GLU 590	VAL 592
Open Loop/FL GATA-4	2.10E-16	Yes	Yes	Yes	Yes
Closed Loop/FL GATA-4	8.10E-15	No	No	Yes	Yes

To further analyze the effect of Kd, the lowest Kd values and the average of the five lowest Kd values for each non-mutated and mutated system were compiled into Table 11 for comparison.

Table 11: Comparison of the lowest Kd values and the average of the five lowest Kd values of mutated and non-mutated UNC-45b models

Lowest Kd Value							
Model System	Non-Mutated	PRO 607 to ALA	GLN 604 to ASN	GLN 604 to ALA	VAL 592 to ALA	GLU 590 to ASP	GLU 590 to ALA
Open Loop	2.10E-16	1.3E-16	6E-17	6.9E-17	2.6E-15	4E-14	2.1E-15
Closed Loop	8.10E-15	5.5E-15	5.3E-15	7.1E-15	4.3E-13	3.9E-15	4.8E-14
Average Kd Value of Five Lowest Values							
Model System	Non-Mutated	PRO 607 to ALA	GLN 604 to ASN	GLN 604 to ALA	VAL 592 to ALA	GLU 590 to ASP	GLU 590 to ALA
Open Loop	1.78E-13	1.01E-13	1.28E-13	1.42E-13	8.25E-14	7.1E-14	1.4E-13
Closed Loop	1.23E-12	9.85E-13	1.04E-12	8.99E-13	2.07E-12	1.17E-12	6.54E-13

## CHAPTER 4

### Discussion

Since previous work (Anderson 2019, Odunuga et al. 2020) found evidence that the UCS domain of UNC-45b is the preferred binding domain for GATA-4, simulation was used to probe the interactions of Full-Length GATA-4 with residues on the large central loop of the UCS domain of UNC-45b. Of particular importance were the GLN 604 and PRO 607 residues in UNC-45b, as GLN 604 is part of a sequence in the loop that is relatively conserved and believed to play a role in binding of myosin to the loop and PRO 607 is near that sequence.

All of the models generated suggest that GATA-4 binds preferentially to the UCS domain, as fifty-five of the sixty generated models showed binding to the UCS domain. All of the Open Loop models showed binding to UCS domain, and twenty-five Closed Loop models showed binding to the UCS domain. Binding to the UCS central loop also appeared to be preferential, as thirty-four models showed an interaction with the loop, and all models with the lowest Kd value had binding at the loop. Eighteen Open Loop and sixteen Closed Loop models

showed binding to the loop. Furthermore, it seemed that the loop served to stabilize the flexible region that bridges the two zinc fingers of GATA-4, as that is where many of the interactions occurred.

Analysis of shared interactions showed that the majority of repeating interactions between the Open Loop and Closed Loop models were present on the loop. Additionally, interaction interfaces of Open Loop models showed very little interaction with the 13H3  $\alpha$ -helix that is believed to bind to myosin within the UNC-45b groove. In the Closed Loop models, there was almost no interaction with the UNC-45b groove at all, likely due to the orientation of the loop over the groove blocking access to potential binding.

Binding interactions were probed by choosing amino acid interactions that were common when Open Loop and Closed Loop UNC-45b were docked with FL GATA-4. In addition to GLN 604 and PRO 607, GLU 590 was chosen due to having a very high number of interactions with the Open Loop/FL GATA-4 models and the Closed Loop/FL GATA-4 models, although only one interaction was shared between them more than twice. VAL 592 was also chosen, as it had the highest number of shared interactions that happened in each system at least two times.

Analysis of interfaces showed that most of the amino acid interactions identified in PRODIGY were van der Waals interactions, as they had bond

distances too large to be accounted for by covalent, ionic, or hydrogen bonds. Some interactions did have hydrogen bonding, but they were not consistent across the different models. There could be hydrogen bonds and ionic bonds that are mediated by H<sub>2</sub>O, but they are not seen due to ClusPro and ChimeraX not taking water into account in binding. In addition, many of the interactions appeared to be hydrophobic in nature, as they were van der Waals interactions that took place between non-polar groups and the interactions took place in an area that would be exposed to the aqueous environment around the UCS loop.

Mutagenesis of residues showed mixed results. Mutation of GLN 604 to ASN 604 did result in ARG 317 interactions being eliminated, but only two of the original models of Closed Loop had this interaction. Mutating GLN 604 to ALA 604 showed very similar results. These results could be due to variance in the ClusPro docking system, meaning the shared interaction could be by chance. PRO 607 did not show any change in binding, and ALA 262 was still the only GATA-4 residue to bind more than once in both models. It is likely that the mutation from proline to alanine was not sufficient to induce a change.

Mutation of GLU 590 to ASP 590 led to more interactions occurring at least twice between Open Loop/FL GATA-4 and Closed Loop/FL GATA-4, including the ILE 314 and LEU 260 interactions that were present in the original models. It is possible that this could also be due to random chance in the models

generated in ClusPro, as only two interactions of each were present in the Closed Loop/FL GATA-4 models. It is also possible that the change is due to the rearrangement of the interaction interface, leading to more interactions. The number of interactions in the Open Loop/FL GATA-4 models was roughly the same, further showing little change in binding activity. It is likely that mutating from GLU to ASP was not enough of a change to cause a detrimental effect to binding.

Mutation of GLU 590 to ALA allowed for a more severe sidechain mutation. The ILE 314 interaction was not shared across the models, but the LEU 260 interaction was still present. Additionally, the number of interactions in the Open Loop models was roughly the same, while the number of interactions in the Closed Loop Models decreased. Closed Loop models had interactions with GLU 590 in nine models before mutation, and only five models after mutation. It is possible that the change in amino acid caused interactions to occur much less frequently in the Closed Loop models, where the loop is laying against the UCS domain. This could be a critical factor in stability of the interaction between GATA-4 and UNC-45b since the loop is very flexible. The decrease in interactions is likely due to a decrease in van der Waals packing in the Closed Loop orientation.

Mutation of VAL 592 to ALA caused the most drastic change. There were no repeated interactions that occurred after the mutation. Furthermore, the Closed Loop models had no repeating interactions at all, while the Open Loop models did. The number of Closed Loop models that had interactions with VAL 592 was five, while nine had interactions before mutation. However, it could also be due to variance in ClusPro as well, as many of the repeated VAL 592 interactions were due to two models in the Closed Loop group.

Analysis of individual Kd values yielded a mixed outcome. Like the non-mutated models, all of the models with the lowest Kd bound with the UCS domain. The Open Loop models all bound to PRO 607, GLN 604, VAL 592, GLU 590, and their mutations. The changes in Kd were minor, with the lowest Kd of the mutant PRO 607 to ALA being  $1.3E-16$ , compared to the  $2.1E-16$  for the non-mutated model. The minimum Kd was lowered in the GLN 604 to ASN mutation, down to  $6E-17$ . However, the minimum Kd rose in the VAL 592 to ALA and GLU 590 to ASP mutations to  $2.6E-15$  and  $4E-14$ , respectively. This gives the appearance that the mutations of VAL 592 and GLU 590 destabilized binding, particularly for the hundred-fold difference with the GLU 590 mutation. However, the averages of the five lowest Kd values showed that the Kd values were consistent across non-mutated and mutated forms of Open Loop UNC-45b.

Unlike the Open Loop variants, Closed Loop models did not interact with PRO 607, GLN 604, or their mutants in their lowest Kd's. The lowest Kd value for the PRO 607 to ALA mutation was 5.5E-15, similar to the non-mutated Kd of 8.1E-15. The mutation of GLN 604 to ASN had a Kd of 5.3E-15, and the Kd for the mutation of GLU 590 to ASP was 3.9E-15. The only Kd that was significantly different was the VAL 592 to ALA mutation, with a Kd of 4.3E-13. This further supports the idea that VAL 592 could be important in the stability of the binding structure. However, like the Open Loop models, the averages of the five lowest Kd's showed consistent Kd values across all forms of Closed Loop UNC-45b.

It is important to note that the averaging of the five lowest Kd values may be an incorrect approach. It is possible that the models generated could be showing transient binding as GATA-4 moves to the optimal location with the lowest Kd, and that the lowest Kd models are the accurate description of the binding. However, it is difficult to conclude that from the still images the models provide. Further experimentation in a lab would need to be done to conclude one way or the other.

## Conclusion

Docking in ClusPro was used to confirm Anderson's (2019, Odunuga et al. 2020) findings that the UCS domain of UNC-45b is the preferential binding site of GATA-4, with the majority of interactions taking place at that domain. Precursory analysis also seemed to support the idea that the central loop of the UCS domain plays a key role in binding and stabilizing GATA-4. Many models showed the loop binding with the "bridge" between the two zinc fingers of GATA-4, hinting at a chaperone-client relationship and supporting Chen's (2012) findings that the coexpression with UNC-45b increases the transcriptional activity of GATA-4.

However, simulation was not sufficient to find consistent interaction interfaces across models that hinted at specificity. This could be due to failure of ClusPro to generate repeated models or a sufficient number of models to see real patterns emerge. However, through mutation of residues, it was demonstrated that GLN 604 and PRO 607 of UNC-45b may not be as important in binding as believed, even though they are possible putative binders of myosin.

Mutation of VAL 592 showed the greatest results. It was demonstrated that VAL 592 may be involved in the stabilization of the UNC-45b/GATA-4 complex, as a mutation in it resulted in an increase in the lowest Kd of both the closed and open loop forms of UNC-45b, a drastic decrease in interactions in the

closed loop form, and complete elimination of the shared interactions that existed before mutation. This result shows a potential important interaction between the UCS loop and GATA-4, supporting the hypothesis that GATA-4 specifically binds to the loop.

The VAL 592 result shows that ClusPro and PRODIGY were able to identify a potential interaction to be tested in the lab. The VAL 592 interaction can be tested by mutating the residue and measuring any changes in chaperoning activity. A similar result in a laboratory setting would help validate the ClusPro and PRODIGY systems.

There are other steps that can be taken. The experiment should be repeated for other residues of UNC-45b, such as the other residues of the loop or the residues on  $\alpha$ -helix 13H3 and the groove itself. Additionally, crystallization of GATA-4 should be done in order to obtain a full model of GATA-4. Because GATA proteins have long strands outside of their zinc fingers that extend into and move in space, only the zinc finger complexes have ever been modeled to be used as a template. GATA-4 itself has only had a single zinc finger modeled. The full structure of GATA-4 would be useful for repeating of the experiment, as the full structure may significantly alter the binding.

## REFERENCES

- Akiyama, Y.; Watkins, N.; Suzuki, H.; Jair, K.; Engeland, M. GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer. *Mol Cell Biol.* **2003**, 23 (23): 8429-39.
- Anderson, M. *Generation of Full-Length Wild-Type GATA4 Protein and Characterization of its Binding to UNC-45 Domains*. Master's thesis, Stephen F. Austin State University, Nacogdoches, TX, **2019**.
- Barral, J. M.; Bauer, C. C.; Ortiz, I.; Epstein, H. F. Unc-45 Mutations in *Caenorhabditis elegans* Implicate a CRO1/She4p-like Domain in Myosin Assembly. *J. Cell. Biol.* **1988**, 143 (5), 1215–1225.
- Barral, J. M.; Hutagalung, A. H.; Brinker, A.; Hartl, F. U., Epstein, H. F. Role of the Myosin Assembly Protein UNC-45 as a Molecular Chaperone for Myosin. *Science* **2002**, 295, 669–671.
- Benkert, P., Biasini, M., Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **2011**, 27, 343-350.

- Buchner, J. Molecular chaperones and protein quality control: an introduction to the JBC Reviews thematic series. *J Biol Chem.* **2019**, 294 (6), 2074-75.
- Bujalowski, P. J.; Nicholls, P.; Garza, E.; Oberhauser, A. F. The central domain of UNC-45 chaperone inhibits the myosin power stroke. *FEBS Lett.* **2017**, 8 (1), 41-48.
- Bujalowski, P. J.; Nicholls, P.; Oberhauser, A. F. UNC-45B Chaperone: The Role of its Domains in the Interaction with the Myosin Motor Domain. *Biophys J.* **2014**, 107 (3), 654-661.
- Chen, D.; Li, S.; Singh, R.; Spinette, S.; Sedlmeier, R. Dual function of the UNC-45b chaperone with myosin and GATA4 in cardiac development. *J. Cell. Sci.* **2012**, 125 (16), 3893–3903.
- Chen, Y.; Bates, D. L.; Dey, R.; Chen, L. DNA binding by GATA transcription factor-complex 3. **2012**, doi: 10.2210/pdb4HC9/pdb.
- Clausen, T.; Gazda, L.; Hellerschmied, D. Crystal structure of the myosin chaperone UNC-45 from *C.elegans* in complex with a Hsp90 peptide. **2012**, doi: 10.2210/pdb4I2Z/pdb.
- Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*, 2nd ed., John Wiley & Sons, Ltd: West Sussex, **2004**; pp 1-16.

Dobrzycki, T.; Lalwani, M.; Telfer, C.; Monteiro, R.; Patient, R. The roles and controls of GATA factors in blood and cardiac development. *IUBMB Life*. **2020**, 72 (1), 39-44.

Donkervoort, S.; Kutzner, C. E.; Hu, Y.; Lornage, X.; Rendu, J. Pathogenic Variants in the Myosin Chaperone UNC-45B Cause Progressive Myopathy with Eccentric Cores. *Am J Hum Genet*. **2020**, 107 (6), 1078-95.

Du, S. J.; Li, H.; Bian, Y.; Zhong, Y. Heat-shock protein 901 is required for organized myofibril assembly in skeletal muscles of zebrafish embryos. *Proc. Natl. Acad. Sci. U S A*. **2008**, 105 (2), 554–559.

Epstein, H. F.; Thomson, J. N. Temperature-sensitive mutation affecting myofilament assembly in *Caenorhabditis elegans*. *Nature*. **1974**, 250, 579–580.

Etard, C.; Roostalu, U.; Strähle, U. J. Shuttling of the chaperones Unc45b and Hsp90a between the A band and the Z line of the myofibril. *Cell. Biol*. **2008**, 180 (6), 1163–1175.

Evans, T.; Reitman, M.; Felsenfeld, G. An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes." *Proc. Natl. Acad. Sci*. **1988**, 85, 5876-5980.

- Gaiser, A. M.; Kaiser, C. J.; and Haslbeck, V; Richter, K. Downregulation of the Hsp90 System Causes Defects in Muscle Cells of *Caenorhabditis Elegans*. *PLoS One* **2011**, 6, e25485.
- Gazda, L.; Pokrzywa, W.; Hellerschmied, D.; Löwe, T.; Forné, I. The Myosin Chaperone UNC-45 Is Organized in Tandem Modules to Support Myofilament Formation in *C. elegans*. *Cell*. **2013**, 152 (1–2), 183–195.
- Gaziova, I.; Moncrief, T.; Christian, C. J.; White, M. A.; Benian, G. M. Mutational Analysis of the Structure and Function of the Chaperoning Domain of UNC-45B. *Biophys. J.* **2020**, 119 (4), 780-791.
- Géraud, C.; Koch, P. S.; Zierow, J.; Klapproth, K., Busch, K. GATA4-dependent organ-specific endothelial differentiation controls liver development and embryonic hematopoiesis. *S. J Clin Invest.* **2017**, 127 (3), 1099–1114.
- Gong, Y.; Zhang, L.; Zhang, A.; Chen, X.; Gao, P. GATA4 inhibits cell differentiation and proliferation in pancreatic cancer. *PloS one*. **2018**, 13 (8), e0202449.
- Hutagalung, A. H.; Landsverk, M. L.; Price, M. G.; Epstein, H. F. The UCS family of myosin chaperones. *J. Cell. Sci.* **2002**, 115, 3983–3990.

- Jansen, R. P.; Dowzer, C.; Michaelis, C.; Galova, M.; Nasmyth, K. Mother cell-specific HO expression in budding yeast depends on the unconventional myosin myo4p and other cytoplasmic proteins. *Cell* **1996**, 84, 651–654.
- Kozakov, D.; Beglov, D.; Bohnuud, T.; Mottarella, S.; Xia, B. How good is automated protein docking? *Proteins: Structure, Function, and Bioinformatics* **2013**, 81 (12), 2159-66.
- Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D. The ClusPro web server for protein-protein docking. *Nature Protocols*. **2017**, 12 (2), 255-278.
- Kuo, C.T.; Morrissey, E. E.; Anandappa, R.; Sigrist, K.; Lu, M. M. GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev.* **1997**, 11 (8), 1048-1060.
- Lambert, S. A.; Jolma, A.; Campitelli, L. F.; Das, P. K.; Yin, Y. The Human Transcription Factors. *Cell*, **2018**, 172 (4), 650-665.
- Laverriere, A.; MacNeill, C.; Mueller, C.; Poelmann, R. E.; Burch, J. B. GATA-4/5/6, a Subfamily of Three Transcription Factors Transcribed in Developing Heart and Gut. *J. Bio. Chem.* **1994**, 269 (37), 23177-84.
- Lee, C. F.; Hauenstein, A. V.; Fleming, J. K.; Gasper, W. C.; Engelke, V. UNC-45 from *Drosophila melanogaster*. **2010**, doi: 10.2210/pdb3NOW/pdb.

- Lee, C. F.; Melkani, G. C.; Bernstein, S. I. The UNC-45 Myosin Chaperone: From Worms to Flies to Vertebrates. *Int. Rev. Cell. Mol. Biol.* **2014**, 313, 103–144.
- Liu, J.; Cheng, H.; Xiang, M.; Zhou, L.; Wu, B. Gata4 regulates hedgehog signaling and Gata6 expression for outflow tract development. *PLoS Genet.* **2019**, 15 (5); e1007711.
- Liu, L.; Srikakulam, R.; Winkleman, D. A. Unc45 activates Hsp90-dependent folding of the myosin motor domain. *J. Biol. Chem.* **2008**, 283 (19), 13185–13193.
- Mohammadi, M.; Kattih, B.; Grund, A.; Froese, N.; Korf-Klingebiel, M. The transcription factor GATA4 promotes myocardial regeneration in neonatal mice. *EMBO Mol Med.* **2017**, 9 (2), 265-279.
- Molkentin, J.D.; Kalvakolanu, D.V.; Markham, B.E. Transcription factor GATA-4 regulates cardiac muscle-specific expression of the alpha-myosin heavy-chain gene. *Mol Cell Biol.* **1994**, 14 (7), 4947-4957.
- Molkentin, J. D.; Lin, Q.; Duncan, S. A.; Olson, E. N. Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev.* **1997**, 11 (8), 1061-72.

Ni, W.; Odunuga, O.O. UCS proteins: chaperones for myosin and co-chaperones for Hsp90. *Subcell Biochem.* **2015**, *78*, 133–152.

Nicholls, P.; Bujalowski, P.J.; Epstein, H.F; Boehning, D. F.; Barral, J. M. Chaperone-mediated reversible inhibition of the sarcomeric myosin power stroke. *FEBS Lett.* **2014**, *588*, 3977-3981.

Odunuga, O. O.; Anderson, M.; Duncan, D. GATA-4 interacts with the UCS domain of the Myosin Chaperone, Striated Muscle UNC-45. *FASEB J.* **2020**, *34* (S1), doi: 10.1096/fasebj.2020.34.s1.06027.

Odunuga, O. O.; Hornby, J. A.; Bies, C.; Zimmerman, R.; Pugh, D. J. Tetratricopeptide repeat motif-mediated Hsc70-mSTI1 interaction. Molecular characterization of the critical contacts for successful binding and specificity. *J. Biol.* **2003**, *278*, 6896-6904.

Petterson, E. F.; Goddard, T. D.; Huang, C. C.; Meng, E. C.; Couch, G. S. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **2021**, *30* (1), 70-82.

Porter, K. A.; Xia, B.; Beglov, D.; Bohnuud, T.; Alam, N. ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. *Bioinformatics* **2017**, *33* (20), 3299-3301.

Price, M. G.; Landsverk, M. L.; Barral, J. M.; Epstein, H. F. Two mammalian UNC-45 isoforms are related to distinct cytoskeletal and muscle-specific functions. *J. Cell. Sci.* **2002**, 115, 4013–4023.

Scheufler, C.; Brinker, A.; Bourenkov, G.; Pegoraro, S.; Moroder, L. Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell* **2000**, 101 (2), 199-210.

Shi, H.; Blobel, G. UNC-45/CRO1/She4p (UCS) protein forms elongated dimer and joins two myosin heads near their actin binding region. *Proc. Natl. Acad. Sci. U S A.* **2010**, 107 (50), 21382–21387.

Srikakulam, R.; Liu, L.; Winkelmann, D.A. Unc45b Forms a Cytosolic Complex with Hsp90 and Targets the Unfolded Myosin Motor Domain. *PLoS One* **2008**, 3 (5), e2137.

Srikakulam, R.; Winkelmann, D.A. Chaperone-mediated folding and assembly of myosin in striated muscle. *J. Cell. Sci.* **2004**, 117, 641-652.

The PyMOL Molecular Graphics System, Version 2.4 Schrödinger, LLC.

Vajda, S; Yueh, C.; Beglov, D.; Bohnuud, T.; Mottarella, S. E. New additions to the ClusPro server motivated by CAPRI. *Proteins: Structure, Function, and Bioinformatics* **2017**, 85 (3), 435-444.

Välimäki, M.J., Ruskoaho, H.J. Targeting GATA4 for cardiac repair. *IUMBD Life* **2020**, 72 (1), 68-79.

Vangone, A.; Bonvin, A. M. J. J. Contact-based prediction of binding affinity in protein-protein complexes. *eLife* **2015**, 4, e07454.

Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, 46, W296-W303.

Wohlgemuth, S.L.; Crawford, B.D.; Pilgrim, D.B. The myosin co-chaperone UNC-45 is required for skeletal and cardiac muscle function in zebrafish. *Dev. Biol.* **2007**, 303 (2), 483–492.

Xu, X.; Eletsky, A.; Lee, D.; Kohn, E.; Janjua, H. Solution NMR Structure of Transcription Factor GATA-4 from Homo sapiens, Northeast Structural Genomics Consortium (NESG) Target HR4783B. **2013**, doi: 10.2210/pdb2M9W/pdb.

Xue, L.; Rodrigues, J.; Kastritis, P.; Bonvin, A. M. J. J.; Vangone, A. PRODIGY: a web-server for predicting the binding affinity in protein-protein complexes. *Bioinformatics*, **2016**, doi:10.1093/bioinformatics/btw514.

Zhang, Y.; Ai, F.; Zheng, J; Peng B. Associations of GATA4 genetic mutations with the risk of congenital heart disease. *Medicine (Baltimore)* **2017**, 96 (18), e6857.

Zhou, L.; Liu, J.; Xiang, M.; Olson, P.; Guzzetta, A. Gata4 potentiates second heart field proliferation and Hedgehog signaling for cardiac septation. *Proc Natl Acad Sci USA* **2017**, 114 (8), E1422-E1431.

## APPENDIX

### A1. List of Programs and Websites

AbleBits	A Microsoft Excel addon created by 4Bits Ltd. Contains functions that allow for efficient collation of spreadsheets for easier parsing of multiple data files. Found at <a href="https://www.ablebits.com">https://www.ablebits.com</a>
ChimeraX	A molecular visualization software created by the Resource for Biocomputing, Visualization, and Informatics group at the University of California at San Francisco. Allows for displaying of molecules and their interactions. Found at <a href="https://www.cgl.ucsf.edu/chimerax">https://www.cgl.ucsf.edu/chimerax</a>
ClusPro	A protein docking webserver created and maintained by Vajda Lab and ABC Group at

Boston University and Stony Brook University.

Can dock protein structures or amino acid sequences and build a series of models based on various conditions. Found at

<https://cluspro.bu.edu>

## PRODIGY

A protein binding energy prediction (PRODIGY) webserver created by the Computational Structural Biology group/NMR Research Group at Utrecht University.

Analyzes models and returns various values such as  $K_d$  and  $\Delta G$  as well as interacting amino acids between protein chains. Found at

<https://bianca.science.uu.nl/prodigy>

## PyMol

An open-source molecular visualization software maintained and distributed by Schrödinger. Can open and edit PDB files to rename or mutate models. Found at

<https://pymol.org>

## RCSB PDB

The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) that is maintained by several universities. A database that contains the models of proteins that have been visualized and published in scientific papers. Found at <https://www.rcsb.org>

## SWISS-MODEL

A modeling software created by the Protein Structure Bioinformatics Group at the University of Basel. Uses known template structures from PDB to build new models from an amino acid sequence and calculates QMEAN and GMQE scores to estimate if the models are reasonable. Found at <https://swissmodel.expasy.org>

## UniProtKB

Universal protein resource knowledgebase (UniProtKB), maintained by the UniProt consortium. Contains information on specific proteins. Found at <https://www.uniprot.org>

## A2. Repeated Interactions Within Models

The loop was not the only place where interactions occurred, but it was the focus of this experiment. Additionally, Table 4 does not have all the interactions that took place within the models, the interactions that occurred at least twice on the Loop in each of Open Loop and Closed Loop models. In the Open Loop model, the loop consisted of amino acids 583 – 616. When docked with GATA-4, there were 226 unique interactions that occurred at least twice in Open Loop models. The loop in Closed Loop models consisted of the amino acids 584 – 612, and there were 97 unique interactions with GATA-4 that occurred at least twice. For the sake of simplifying these lists, only interactions that occurred at least twice in both models were considered and are listed in Table 4.

### A3. The Terminal Loop of the UCS Domain

Another area of UNC-45 that was analyzed was the N-terminal area of the UCS domain. At the terminal end, there is a small loop where binding had occurred in docking. Interactions on this loop were counted and compiled into a list. This terminal loop consisted of amino acids 895 – 907 in both Open Loop and Closed Loop models. Open Loop models had 84 unique interactions with GATA-4 that occurred at least twice, and Closed Loop models had 29. Only five of these interactions occurred at least twice in both Open and Closed Loop models and are listed in Table 12.

*Table 12: List of Repeating Interactions on the Terminal Loop of the UCS Domain*

UNC-45b Amino Acid	GATA-4 Amino Acid	Open Loop/FL GATA-4 Models where Interaction is Present	Closed Loop/FL GATA-4 Models where Interaction is Present
THR 896	ALA 262	3, 5	7, 14
VAL 897	ARG 283	9, 10	9, 26
VAL 898	MET 222	6, 18	14, 15
GLU 905	ARG 214	6, 18	3, 15, 17
LYS 906	GLU 215	3, 7	3, 17, 24, 28

Since there were only five separate interactions that were shared across the models, it was decided to focus on the large central loop where there was a higher number of interactions. Interestingly, Open Loop models had a higher

occurrence of binding to the terminal loop, despite the fact that Open Loop models were much more likely to bind to the central loop than Closed Loop models. None of the Open Loop models that showed binding to the terminal loop were bound to the central loop.

## VITA

Drake Duncan grew up in Liberty Hill, Texas where he graduated from Liberty Hill High School in the top ten of his class. In the Fall of 2013, he attended Texas A&M University. He graduated with a Bachelor of Science in Biomedical Sciences in the Spring of 2017. In the Fall of 2017, he attended Texas Tech University and graduated in the Summer of 2018 with a Master of Business Administration degree. In the Spring of 2019, he attended Stephen F. Austin State University. He began his graduate work in the Fall of 2019 under the supervision of Dr. Odutayo Odunuga while working as a Graduate Teaching Assistant. In May 2021, he graduated with a Master of Science in Chemistry.

Permanent Address:                   19501 Sangremon Way  
  Pflugerville, TX 78660

Style manual designation:           The ACS Style Guide

This thesis was typed by Drake Alexander Duncan