

Stephen F. Austin State University

SFA ScholarWorks

Electronic Theses and Dissertations

Fall 8-5-2023

Bioinformatic Analysis of the Interaction Of EVC And EVC2 Proteins

MD ATIKUL ISLAM MAMUN

Stephen F Austin State University, mamunm@jacks.sfasu.edu

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

 Part of the [Chemistry Commons](#)

[Tell us](#) how this article helped you.

Repository Citation

MAMUN, MD ATIKUL ISLAM, "Bioinformatic Analysis of the Interaction Of EVC And EVC2 Proteins" (2023). *Electronic Theses and Dissertations*. 555.
<https://scholarworks.sfasu.edu/etds/555>

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.

Bioinformatic Analysis of the Interaction Of EVC And EVC2 Proteins

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/).

Stephen F. Austin State University

SFA ScholarWorks

Librarian and Staff Publications

Ralph W. Steen Library

8-5-2023

Bioinformatic Analysis of the Interaction Of EVC And EVC2 Proteins

MD ATIKUL ISLAM MAMUN

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



Part of the [Chemistry Commons](#)

[Tell us](#) how this article helped you.

BIOINFORMATIC ANALYSIS OF THE INTERACTION OF
EVC AND EVC2 PROTEINS

By

MD ATIKUL ISLAM MAMUN, Bachelor of Science

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 in Natural and Applied Science in Chemistry

STEPHEN F. AUSTIN STATE UNIVERSITY

August 2023

BIOINFORMATIC ANALYSIS OF THE INTERACTION OF
EVC AND EVC2 PROTEINS

By

MD ATIKUL ISLAM MAMUN, Bachelor of Science

APPROVED:

Odutayo Odunuga, Ph.D., Thesis Director

Darrell Fry, Ph.D., Committee member

Bidisha Sengupta, Ph.D., Committee member

Dipak Singh, Ph.D., Committee member

Sheryll Jerez, Ph.D.

Interim Dean of Research and Graduate Studies

© Md Atikul Islam Mamun
All Rights Reserved
2023

ABSTRACT

Protein to protein interactions are significant to various biological processes, functioning, and the regulation of cells and organisms. Proteins perform essential cellular tasks and often work together in complexes, enabling specific functions like enzymatic activity, signal transduction, DNA replication, and cell division. These interactions are important in signaling pathways, allowing proteins to interact sequentially, leading to specific cellular responses. Some proteins rely on chaperones to fold correctly, and protein to protein interactions enhance stability, preventing degradation.

Researchers at Stanford University School of Medicine discovered that EvC syndrome is related to a malfunction in the Hedgehog (Hh) signaling pathway. They found that Hh agonists stimulate the interaction between the ciliary protein EVC2 and Smoothened (Smo) at a specific ciliary compartment called the EVC zone. This complex is essential for proper Hh signal transmission and plays an important role in the molecular basis of EvC syndrome and Weyers Acrofacial Dysostosis. EVC and EVC2 are found together in a complex, and it is assumed that they interact with each other to influence their activities.

To identify protein to protein interactions, a combination of computational tools and software were used. Alphafold, a learning-based system, predicts the 3D structure of proteins, providing valuable insights into their conformation and

potential interaction sites. ClusPro and ZDock, protein docking software, were used to predict the binding mode between two proteins and generate potential complexes. Prodigy, a web server that calculates the binding affinity between two proteins, was used to assess the strength of the predicted interactions. Furthermore, Pymol, a molecular visualization tool, aided in visualizing and analyzing the 3D structures of proteins and their interactions, facilitating the exploration of potential binding sites and intermolecular interactions.

In the fragment region (451-750), we observed probable interactions between mEVC and mEVC2 involving 11 amino acids: LEU-540, LYS-541, PRO-544, GLU-545, SER-548, LEU-549, PRO-550, VAL-551, ALA-552, GLU-553, and THR-556. These contacts included van der Waals forces (hydrophobic) and electrostatic interactions, with some of them forming hydrogen bonds. In most of the complexes, EVC2 did not make direct contact with the P-loop and Leucine zipper region of EVC.

The study provided valuable insights into the amino acids responsible for the interaction between mEVC and mEVC2 proteins, highlighting the importance of specific residues and the effects of mutations on their interaction dynamics. However, obtaining consistent interaction interfaces across various models was challenging, possibly due to limitations in generating repeated or sufficient models using computational tools and software.

ACKNOWLEDGEMENTS

First and foremost, I thank God for providing me with all the opportunities needed for me to succeed in life. Without His blessings, I would not have come this far. Next, I thank my family back home in Bangladesh for their constant love and support throughout my academic journey.

I am immensely thankful to my thesis advisor and interim chair of Department of Chemistry and Biochemistry, Professor Dr. Odutayo Odunuga, for his support and encouragement during my graduate career. A research position in his laboratory introduced me to this important topic and paved the way for my education in the field of biochemistry and molecular biology. I would also like to thank my thesis committee, who were flexible and were willing to accept changes to my thesis proposal. I am immensely grateful to Dr. Bidisha Sengupta, who has been a constant source of support and guidance since the moment I joined SFA. Her unwavering presence has made a significant difference in my journey, and I cannot thank her enough for it. In addition, I would like to acknowledge the faculty and staff in the department for providing a warm learning environment. Lastly, I would like to acknowledge The Robert A. Welch Foundation for supporting my research.

TABLE OF CONTENTS

ABSTRACT	II
ACKNOWLEDGEMENTS	IV
TABLE OF CONTENTS	V
LIST OF FIGURES	IX
LIST OF TABLES	XII
LIST OF ABBREVIATIONS	XIV
CHAPTER ONE: LITERATURE REVIEW	1
1.1 ELLIS VAN CREVELD (EVC) SYNDROME	1
1.2 EVC	2
1.2.1 Overview of EVC Gene and Protein	2
1.2.2 Leucine Zipper Motif	3
1.2.3 P-loop Motif	5
1.3 EVC2	6
1.4 EVC, EVC2, AND THE HEDGEHOG SIGNALING PATHWAY	8
1.4.1 The Hedgehog (Hh) Signaling Pathway	8

1.4.2	Involvement of the EVC Protein in the Hedgehog Signaling.....	10
1.4.3	Regulation of the Hedgehog Signaling Pathway.....	11
1.4.4	Relation of EVC, Hedgehog Signaling Pathway and Cancer.....	14
1.5	PROTEIN-PROTEIN INTERACTIONS	14
1.6	EXPERIMENTAL METHODS FOR DETERMINING PROTEIN STRUCTURES.....	15
1.7	PROTEIN STRUCTURE PREDICTION, COMPUTATIONAL MODELING	15
1.7.1	Importance of Accurate Protein Structure Prediction in Gaining Insights into Protein Functions.....	16
1.7.2	Challenges in Protein Structure Prediction	17
1.7.3	Modeling of Protein Structure Prediction	17
1.7.4	Improving Accuracy in Protein Structure Prediction.....	18
1.8	ALPHAFOLD, A MODELING WEB SERVER	19
1.9	PROTEIN-PROTEIN DOCKING.....	22
1.9.1	ClusPro, a Web Server for Protein-Protein Docking	25
1.9.2	ZDock, a Protein Docking Program.....	26
CHAPTER TWO: MATERIALS AND METHODS		28
2.1	OBTAINING PRIMARY SEQUENCES	28
2.2	ALIGNMENT.....	28
2.3	FINDING COILED-COIL REGION.....	29
2.4	MODEL GENERATION.....	29

2.5	SUPERIMPOSING OF PROTEINS.....	29
2.6	DOCKING OF MEVC AND MEVC2	30
2.7	OBTAINING K_D VALUES AND INTERACTION INTERFACES	30
2.8	MODELING OF INTERACTION INTERFACES.....	31
2.9	MUTAGENESIS	31
2.10	DOCKING OF MEVC AND MEVC2 BY ZDOCK	32
CHAPTER THREE: RESULTS		34
3.1	COILED-COIL REGION OF MEVC AND HEVC	34
3.2	ALIGNMENT.....	37
3.3	MODELING OF PROTEINS	38
1.1	DOCKING BY CLUSPRO.....	40
3.3.1	Interaction between fragment mEVC and full mEVC2	40
3.3.2	Ful length mEVC and fragment (247 to 661) mEVC2.....	41
3.3.3	Interaction between both mEVC and mEVC2 fragment protein.....	42
3.3.4	Interaction between full length mEVC and full length mEVC2	44
3.4	MUTATION	58
3.5	DOCKING BY ZDOCK.....	59
3.5.1	Interaction between the Fragment mEVC and the Full mEVC.....	59
3.5.2	Interactions between the Full mEVC and the Fragment mEVC2	60

3.5.3 Interaction between the Full length mEVC and the Full Length mEVC2.....	61
3.5.4 Contacts between the Fragment mEVC and the Fragment mEVC2.	62
CHAPTER FOUR: DISCUSSIONS	65
CHAPTER FIVE: CONCLUSIONS.....	68
APPENDIX LIST OF PROGRAMS AND WEBSITES.....	78
VITA.....	80

LIST OF FIGURES

Figure 1 Bilateral polydactyly with short fingers in patient	1
Figure 2 Distribution of the mutations described in EVC and EVC2	2
Figure 3 Crystal structure of basic leucine zipper transcription factor RepA1 bound to DNA	4
Figure 4 P-loop multiple sequence alignment.....	6
Figure 5 Summary of expected EVC2 translated protein products. Colored boxes specify putative domains.....	7
Figure 6 A simplified model for Hh signaling in Mammalian cells. SMO is the key signal transducer of the Hh pathway	10
Figure 7 Primary cilium, EVC and EVC2/LIMBIN proteins, and the mechanism of EVC-EVC2/LIMBIN in regulating Hedgehog signaling.....	12
Figure 8 Flowchart of the experimental procedure. Shows the steps taken and what website/program was utilized.	33
Figure 9 Protein-protein interaction of EVC and EVC2	34
Figure 10 Predicted coiled-coil region of mouse EVC Protein (NP_0672672):.....	36

Figure 11 Predicted coiled-coil region of mouse EVC2 Protein (limbin isoform X1 [Mus musculus] NCBI Reference Sequence: XP_036021349.1).....	37
Figure 12 3D model of full-length mouse EVC Protein (NP_0672672_75e86).....	38
Figure 13 3D model of the fragment of mouse EVC protein	38
Figure 14: 3D model of mouse EVC2 protein	39
Figure 15 3D model of the fragment of mouse EVC2 Protein.....	39
Figure 16 Interaction between fragment mEVC and full length mEVC2	41
Figure 17 Interaction full length Mouse EVC and fragment mouse EVC2	41
Figure 18 Interaction between the fragment mouse EVC protein and fragment mouse mEVC Protein	42
Figure 19 Interaction between full length mEVC and full-length mEVC2 proteins	44
Figure 20 Contacts of LEU-540 of mEVC to ASN-262 of mEVC2 in Model 000	47
Figure 21 Contacts of LYS-541 of mEVC to ASN 262, GLN 354, GLU 358 of mEVC2 in Model 000.	48
Figure 22 Contacts of PRO-544 of mEVC to ASN 262, ALA-262 of mEVC2 in Model 000	49

Figure 23 Contacts of GLU-545 of mEVC to PRO 255, SER 256, PRO 257, SER 260, ASN 262 of mEVC2 in Model 000	50
Figure 24 Contacts of SER-548 of mEVC to ALA-261, GLN-597 of mEVC2 in Model 000	51
Figure 25 Contacts of LEU-549 of mEVC to ALA 261, SER 265, LYS 594 of mEVC2 in Model 000	52
Figure 26 Contacts of PRO-550 of mEVC to ALA 261, SER 265, LEU 268, LEU 590, LYS 594 of mEVC2 in Model 000	53
Figure 27 contacts of VAL-551 of mEVC to ALA 261, ASN 262, SER 265 of mEVC2 in Model 000	54
Figure 28 Contacts of ALA-552 of mEVC to SER 265, SER 269, ASP 272, ARG 586 of mEVC2 in Model 000	55
Figure 29 Contacts of GLU-553 of mEVC to ARG 586, LEU 590, LYS 594 of mEVC2 in model 000	56
Figure 30 Contacts of THR-556 of mEVC to ARG 586 of mEVC2	57

LIST OF TABLES

Table 1 Alignment of mouse EVC, EVC2 and human EVC and EVC2.....	37
Table 2 List of the contacts between full length mEVC and fragment (residue 247 to 661) mEVC2 interface.....	42
Table 3 List of the contacts between the fragment of mEVC (451 to 750) protein and fragment of mEVC2 (247 to 661) protein interface	43
Table 4 List of the contacts between full length mEVC and mEVC2	45
Table 5 Compare of K _d value and ΔG for top five docked models	45
Table 6 List of the bond Length and probable bond type in between the interface of fragment mEVC and mEVC2 proteins	46
Table 7 Interaction between full length muted (GLU 545 to ALA) mouse EVC and full-length mouse EVC2 docked by ClusPro	58
Table 8 K _d value and ΔG when muted mouse EVC docked with full length mEVC2 by clusPro.....	59
Table 9 Compare of K _d value and ΔG when fragment mEVC docked with full length mEVC2 by ZDock.....	59
Table 10 List of contacts between full length mEVC and fragment mEVC2	60
Table 11 Compare of K _d value and ΔG when full length mEVC docked with fragment mEVC2 by ZDock	60

Table 12 Compare of K_d value and ΔG of first 10 complex when full length mEVC docked with full length mEVC2 by ZDock.....	61
Table 13 Compare of K_d value and ΔG when fragment mEVC docked with fragment mEVC2 by ZDock	62
Table 14 List of contacts between fragment of mouse EVC and fragment of mouse EVC2 (ZDocked)	62

LIST OF ABBREVIATIONS

EvC	Ellis–van Creveld
hEVC	Human homolog of <i>EVC</i>
mEVC	Mouse homolog of <i>EVC</i>
SMo	Smoothened
Hh	Hedgehog
K _d	Dissociation Constant
ΔG	Gibbs Free Energy
PDB	Protein Data Bank
ATP	Adenosine Diphosphate
GTP	Guanosine Triphosphate
GDP	Guanosine Diphosphate
IDDT	The Local Distance Difference Test
pLDDT	The Predicted Local Distance Difference Test
mmCIF	The Macromolecular Crystallographic Information File
RMSD	Root-Mean-Square Deviation
FFT	Fast Fourier Transform

CHAPTER ONE: LITERATURE REVIEW

1.1 Ellis van Creveld (EvC) Syndrome

Ellis–van Creveld (EvC) syndrome was first described by Richard W.B. Ellis of Edinburgh and Simon van Creveld of Amsterdam in 1940¹. This disorder is an unusual autosomal recessive genetic irregularity affecting tissues, organs, and skeletal structure. It is characterized by additional fingers in hands or toes in feet (polydactyly), short figure and shorter arms, legs, or ribcage (uneven dwarfism) and dysplasia. Dental aberrations include short upper lip, multiple frenula, a fusion of teeth and lack incisors².



Figure 1 Bilateral polydactyly with short fingers in patient (Shi et al. 2016). (Right): Oral abnormalities in a patient with Ellis–van Creveld syndrome. Congenital absence of the incisors, conical shaped teeth, and multiple frenula³.

The typical population frequency of Ellis–van Creveld (EvC) syndrome is 7 per 1,000,000 globally. EvC disorder is more frequent in the Amish population

(Lancaster County, Pennsylvania, US) with radically increased incidence of 5 per 1000, so the frequency of transporters in this population may be high as 13%⁴. Around 60% of patients affected by the syndrome have genetic cardiac defects⁵. EvC may be identified prenatally by ultrasound checkup. The clinical diagnosis is based on analysis of the manifestations explained above. Though, the ultimate diagnosis is molecular analysis by directly sequencing for diagnosis mutations in the *EVC* and *EVC2* genes⁵. Mutations in these two genes have been revealed to be liable for EvC disorder⁶.



Figure 2 Distribution of the mutations described in *EVC* and *EVC2*.³

1.2 EVC

1.2.1 Overview of *EVC* Gene and Protein

Polymeropoulos et al. (1996) plotted the EvC disorder to chromosome 4p16. Ruiz-Perez et al. (2000) labelled the *EVC* gene locus by positional cloning, and three years later, this group reported a second gene locus, *EVC2*, immediately adjacent to *EVC* in a head-to-head design⁷. The genomic size of *EVC* in mice is 117 kilobase pairs, and the mRNA codes for a 1005 amino acid is 112 kDa protein consisting of a leucine zipper, two putative nuclear localization signals, and a

putative transmembrane domain⁸. The human homolog of *EVC* (*hEVC*) is a 992-amino acid, 112 kDa protein⁹. A *EVC* protein was identified at the base of the cellular organelles, cilia, in cell culture⁶. By *in situ* hybridizing using a whole mouse, *EVC* expression was detected in growing bones such as limbs, vertebrae, and ribs¹⁰. Studies on the mice deficient of *EVC* indicated that the protein is also present in the orofacial region in addition to other bone structures.

1.2.2 Leucine Zipper Motif

The leucine zipper motif was first revealed as a dimerization domain in the yeast transcriptional factor GCN4 and in the oncogenic proteins Fos and Jun¹¹. The leucine zipper pattern consists of a repeating of 4-5 leucine's spaced seven residues apart (figure 4). The result of genetic and functional analysis of a leucine zipper-like located at the N-terminus of RepA revealed that leucine zipper motif modulates the equilibrium between monomeric and dimeric forms of the RepA protein. In addition, seven residues (a, b, c, d, e, f, g) of leucine zipper (figure 5) have diverse functional characters such as determining the specificity of heterodimerization, contributing moderately towards dimer stability, and determining the degree of oligomerization between leucine zipper coils¹¹. Dimerization assay by inserting a leucine zipper mutation at position d of the putative α -helix suggested that leucine residues in position d are participating in the dimerization of RepA, a protein required for transcriptional autoregulation¹¹.

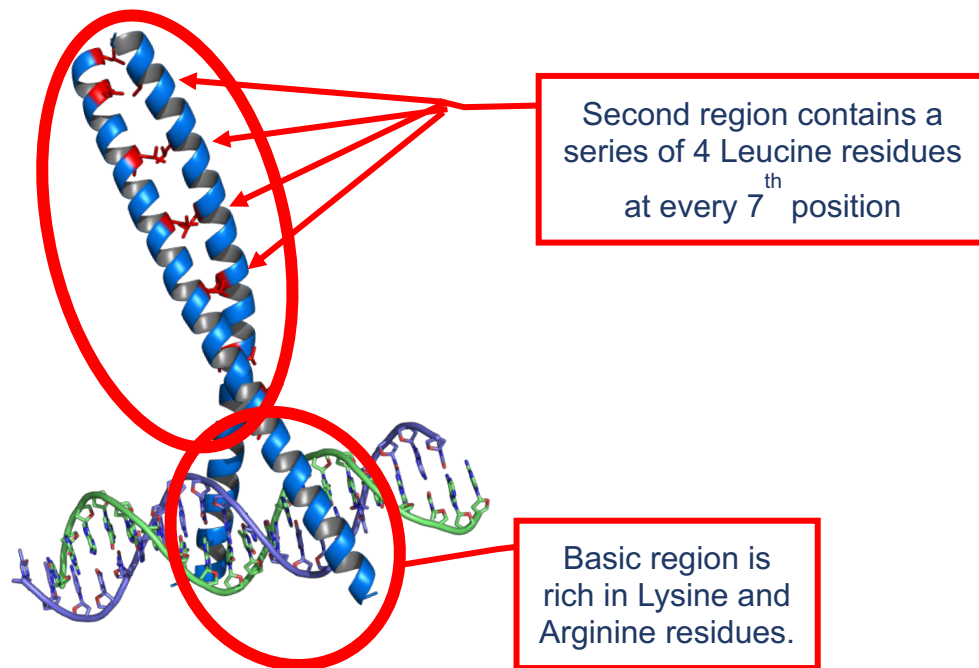


Figure 3 Crystal structure of basic leucine zipper transcription factor RepA1 bound to DNA (method: x-ray diffraction, Source: www.cleanpng.com)

Leucine zipper motif is mostly found in transcription factors that control DNA transcription¹². Another study also described leucine zipper as being responsible for the dimerization and regulation of the GTPase activity of human guanylate binding proteins (hGBP-1 and 2). This analysis also revealed that a monomeric form of hGBP-1 and 2 generates only GDP, while dimeric form yields both GDP and GMP during guanylate binding and hydrolysis¹³.

A typical EVC splicing transcript has been described that may lead to irregular EVC mRNA function and result in clinical phenotypes⁶. Furthermore, previous studies have predicted that these splicing errors would occur just prior to a leucine zipper domain which leads to another theory that the putative leucine zipper domain may be crucial for protein function⁸. In humans, a EVC amino acid sequence analysis study indicated there is a less conserved 34-amino-acid region (residue number 489–523) that is adjacent to a putative *EVC* ATP/GTP binding site (565–572)⁶. However, this putative EVC protein domain and its biological purpose are still unknown.

1.2.3 P-loop Motif

P-loop (phosphate-binding loop) motif is usually found in ATP- and GTP-binding proteins¹⁴. The P-loop clearly interacts with the phosphates of ATP¹⁵. The primary P-loop structure includes a glycine-rich sequence followed by a conserved lysine (K) and a serine (S) or threonine (T). The lysine (K) residue in P-loop motif-A may play a vital role in nucleotide-binding¹⁴. It is understood that the P-loop motif-A is placed close to the N-terminus of adenylate kinases which catalyze the conversion of ATP and AMP into two molecules of ADP. The Walker-B motif structure also contacts the nucleotide, and its acidic residues are vital for ATPase motion. It is described that the Ras proteins, which are possibly engaged in growth-promoting signal transduction pathways in the cell, have the capability to bind GTP

and GDP and retain a weak GTPase activity, which is impacted by the P-loop mutations¹⁴.

Odunuga et al., (2011) discovered the existence of the P-loop motif in EVC sequences (figure 4) of primate origin by using stringent bioinformatic parameters. The P-loop motif seemed less conserved in EVC sequences of other species. For example, mice under less stringent. The existence of P-loop signals that EVC possesses putative ATPase/GTPase activity that might be required for its function.

Chimpanzee	MTGLPPEECDYLRQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Bonobo	MTGLPPEECDYLRQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Gibbon	MTGLPPEECDYLRQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Baboon	MTGLPPEECDYLRQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Human	MTGLPPEECDYLRQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Gorilla	MTGLPPEECDYLRQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Marmoset	VTGLPPEECDYLRQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Monkey	MTGLPPEECDYLRQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Cat	MTGLSQADCEYLQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Dog	VTGLSQAECEYLQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Cow	VTGLSQAECEYLQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Mouse	VTSLPVAECETLRQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Rat	VSSLPVAECETLRQEVQENAAQGLGKSDRFRQQQWKVQELLEQQQVWMEECALSSVLQTHLRDHKSTIRGVLRGLGGLTEESTRCV	LQGHDLRLRSALRRLALRGNALATLTQ
Consensus	A****GKS	L*****L*****L***.*L
	P-loop	Leucine zipper

Figure 4 P-loop multiple sequence alignment (Odunuga et al.,2011)

1.3 EVC2

The genomic size of mEVC2 is 166 kilobase pairs and the mRNA encodes a 1228- amino acid protein. *EVC2* is encoded in the vicinity to *EVC* and they are split by 2,624 base pairs in human and 1,647 base pairs in mice¹⁰.

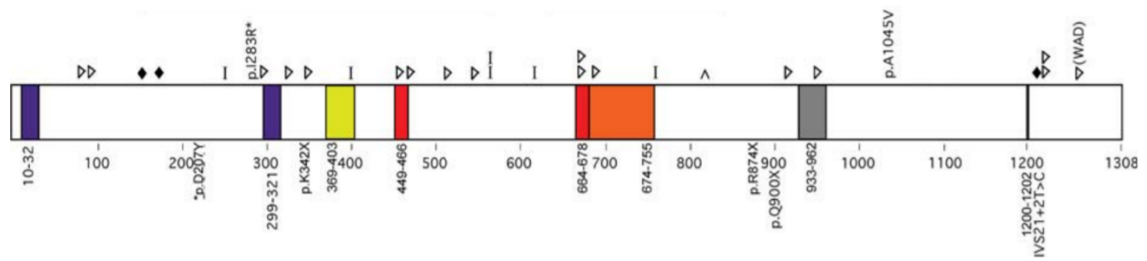


Figure 5 Summary of expected *EVC2* translated protein products. Colored boxes specify putative domains.

It is proposed that *EVC* and *EVC2* act co-dependently because of undifferentiated phenotypes related to mutations in the two genes. To illustrate the functional collaboration between *EVC* and *EVC2* proteins, deletion constructs for the genes were created and used in a directed yeast-two-hybrid assay.

Based on their head-to-head configuration on the chromosome and proximity of their transcriptional start sites, it is recommended that coordinated expressions of *EVC* and *EVC2* proteins is important for maintaining their stoichiometric amounts in the cell⁷. By immunostaining analysis, *EVC* and *EVC2* proteins were detected on the primary cilia in various cell types. Full length mouse *EVC2* protein was found both in the cytoplasmic and in the nuclear fractions. However, full-length mouse *EVC* protein was only found in the cytoplasmic fractions by western blot analyses of mouse embryonic fibroblasts derived from null mice. To characterize the interaction between *EVC* and *EVC2* proteins, deletion constructs of the genes were generated and used in a directed yeast-two-

hybrid assay⁹. Considerable binding was noted between the mouse EVC2 fragment having residue number 250-671 and full-length mouse EVC.

Labelled EVC fragments (residues 463-1005) and EVC2 fragment (residues 250-671) co-precipitated from HEK 293 cells as proven by western blotting analyses. Hence, it has been proved that both proteins physically interact in the cell. Based on their head-to-head configuration on the chromosome and vicinity of their transcriptional start sites, it is suggested that simultaneous expression of EVC and EVC2 proteins is required for retaining proper stoichiometry⁹. By immunostaining analysis, EVC and EVC2 proteins were identified on the primary cilia in distinct cell types and EVC2 protein was essential for the localization of EVC protein at the base of primary cilia. Full-length mouse EVC2 protein was identified both in the cytoplasmic and in the nuclear portions, but full-length mouse EVC protein was only identified in cytoplasmic portions by western blots of mouse embryonic fibroblasts.

1.4 EVC, EVC2, and the Hedgehog Signaling Pathway

1.4.1 The Hedgehog (Hh) Signaling Pathway

The Hedgehog (Hh) signaling pathway plays a significant role in maintaining cell growth and differentiation in normal human embryonic growth. Key actors with the Hh pathway in mammals involve three ligands (Sonic hedgehog—Shh, Indian

hedgehog—Ihh and Desert hedgehog—Dhh), two receptors, Patched (Ptch) and smoothened (Smo), and three transcription factors, Gli1, Gli2, Gli3¹⁵.

Without Hh ligand, the Hh signaling pathway is turned off by Ptch, which prevents the triggering of Smo. The Ptch restrains Smo by controlling it in a vesicle inside the cell. This inhibits downstream Hh signaling processes. Hh signaling is initiated when Hh ligands bind to Ptch on the cell surface. Elimination of Ptch allows Smo to turn from intracellular partition to the cell surface where it elicits Hh signaling pathway. Activation of Smo flows downstream leading to the Gli-mediated initiation of transcription of the target genes¹⁵.

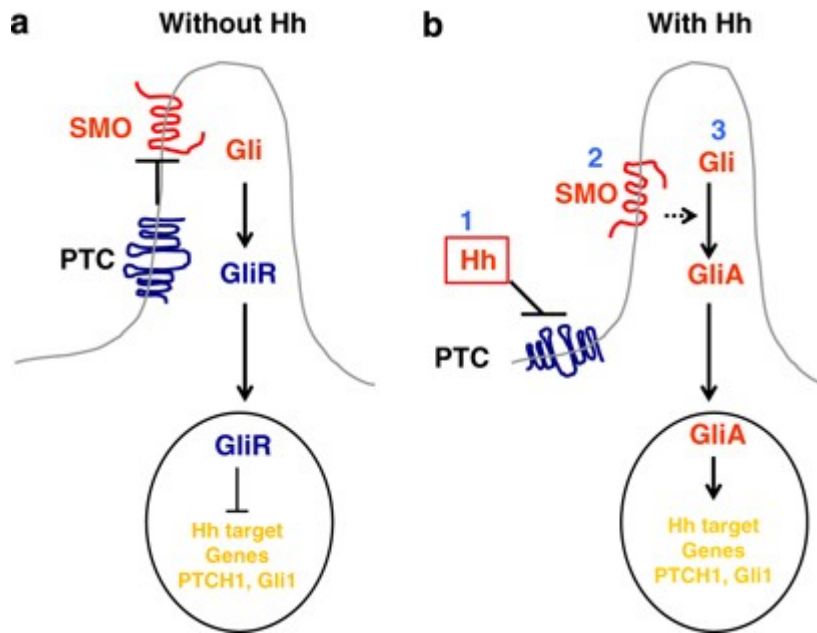


Figure 6 A simplified model for Hh signaling in Mammalian cells. SMO is the key signal transducer of the Hh pathway. (a) In the absence of the Hh ligands, Hh receptor PTC inhibits SMO signaling via an unknown mechanism. Gli molecules are processed into repressor forms, which turn off the Hh-signaling pathway. (b) In the presence of Hh, PTC is unable to inhibit SMO. SMO undergoes conformational changes and is localized to cilium. Gli molecules are now processed to active forms (GliA), which will activate the Hh target genes¹⁶

1.4.2 Involvement of the EVC Protein in the Hedgehog Signaling Pathway

In the Hh signaling pathway EVC is a positive regulator. It has been proven that the primary cilia are specialized for Hh signal transduction. Sufu and Gli, which are significant elements of the pathway, are confined to the primary cilia, while Smo translocates to the cilia upon pathway stimulation¹⁷. Studying cilia-dependent

processes in mesenchyme-derived tissues of the limb verified that cilia are crucial for Ihh signaling¹⁷. The EVC protein also concentrates close to the base of primary cilia.

Moreover, Smo blocks phosphorylation and cleavage of full-length Gli3 to the transcriptional repressor Gli3R. By western blot analysis, Gli3R appears in the same amount in both EVC-depleted mice and normal mice¹⁸. EVC reacts downstream of Smo to accelerate transcription of the Indian Hh-regulated genes¹⁷.

1.4.3 Regulation of the Hedgehog Signaling Pathway

For the Hh signaling pathways, EVC2 also reacts as a positive regulator. EVC2 expression in Hh reporter cells (LIGHT2) indicated weakened response to purmorphamine, a Smo agonist⁹. Similarly, Ptch1 expression in response to purmorphamine in EVC2 mutant mouse embryonic fibroblasts was also lowered. The conclusion of these analyses was that EVC2 is a positive regulator of the Hh signaling pathway⁹.

Molecular and developmental studies suggested that EVC and EVC2 protein's function codependently in cilia mediated cardiac morphogenesis. Although the mechanism by which EVC and EVC2 proteins cause identical phenotypes with characteristic cardiac defects remains unknown, there are observations on colocalization of both proteins in the developing heart supporting

the notion that EVC and EVC2 proteins play a role in atrioventricular structure development¹⁹.

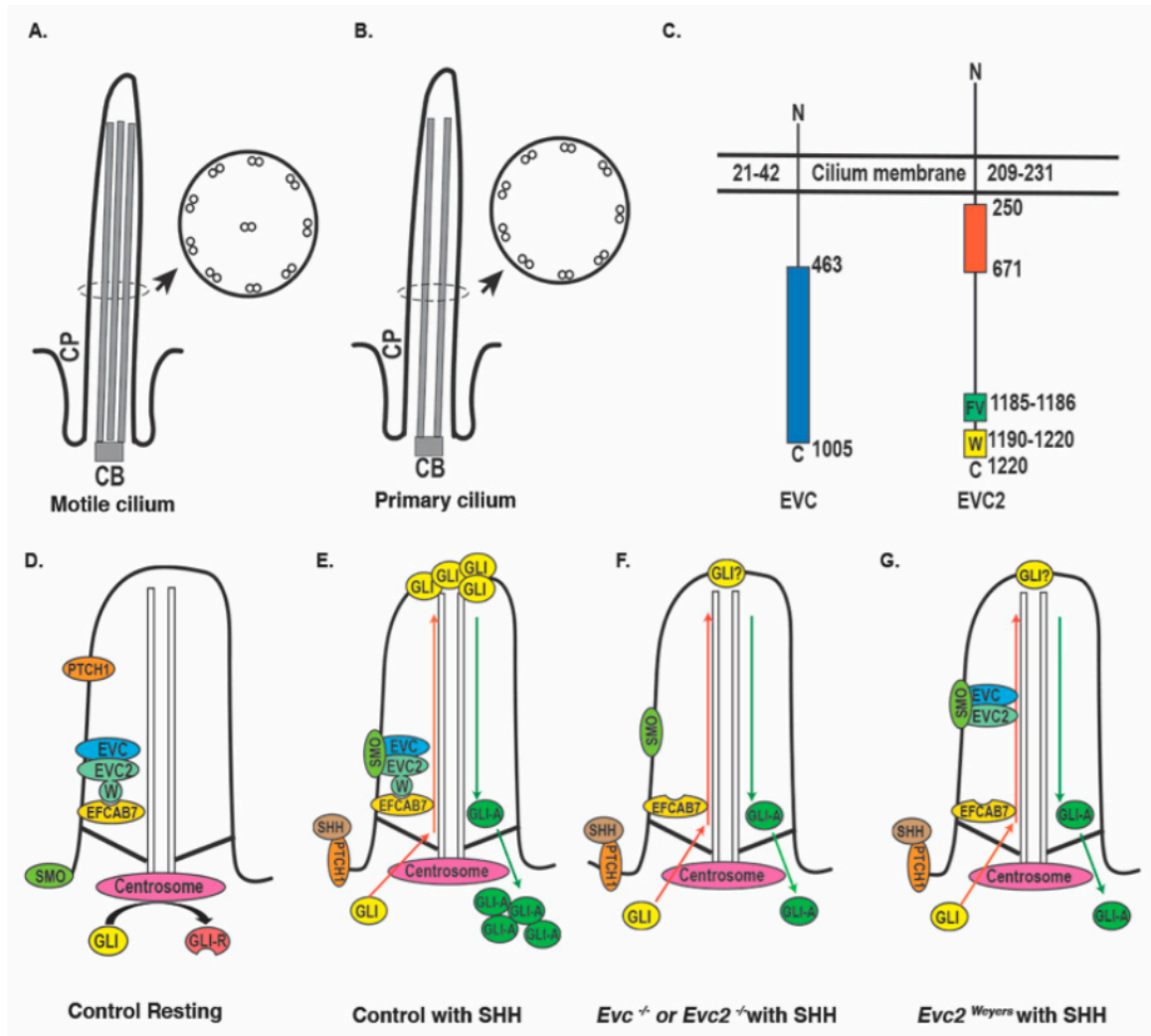


Figure 7 Primary cilium, EVC and EVC2/LIMBIN proteins, and the mechanism of EVC-EVC2/LIMBIN in regulating Hedgehog signaling within the primary cilium. Illustrations of the motile cilium (A) and the primary cilium (B) are shown. CB, cilium base; CP, cilium pocket. (C). Structures of the EVC and the EVC2/LIMBIN are shown. The EVC and EVC2/LIMBIN are N terminal anchored proteins. Blue box in the EVC and orange box in the EVC2/LIMBIN indicate domains for interaction

with each other. Green box indicates the domain for the ciliary localization of the EVC2/LIMBIN and the yellow box indicates the domain for localization of the EVC2/LIMBIN at the EVC zone. Numbers indicate the numbers of amino acids in each protein. (D) EVC-EVC2/LIMBIN complexes are localized at the bottom of cilia by tethering to EFCAB7 through the W domain in EVC2/LIMBIN. In the absence of Hedgehog ligand, PTCH1 resides within the primary cilium, and GLI proteins are processed to the repressor form (GLI-R) at the centrosome. (E) In the presence of Hedgehog ligand, binding of the ligands with PTCH1 leads to exclusion of PTCH1 out of the primary cilium, which allows SMO to enter the primary cilium. Within the primary cilium, SMO interacts with EVC-EVC2/LIMBIN at the bottom of the primary cilium, which allows GLI trafficking into the primary cilium and accumulation at the tip of the primary cilium. After entering the primary cilium, GLI are processed to the activator form (GLI-A). GLI activators exit the primary cilium and translocate into the nucleus to activate Hedgehog responsive genes. (F) EVC or EVC2/LIMBIN loss of function leads to absence of EVC-EVC2/LIMBIN complexes within the primary cilium. When Hedgehog signaling is activated, SMO still moves into the primary cilium, but without EVC-EVC2/LIMBIN complexes, SMO cannot lead to full activation of GLI. (G) In Weyers form of mutant cells, EVC-EVC2/LIMBIN complexes cannot be restricted at the bottom of primary cilium due to no interactions with EFCAB7 caused by loss of the W domain in EVC2/LIMBIN, thus EVC-EVC2/LIMBIN-SMO complex cannot lead to full activation of GLI²⁰.

Dorn et al. has shown that interaction of Smo with the EVC/EVC2 complex is essential for Hh signaling¹⁸. It was observed that the endogenous ciliary proteins EVC/EVC2 complex function downstream of SmoA1 is also necessary downstream of Sufu. Therefore, it is suggested that the EVC/EVC2 complex modulates Hh signaling by cooperating with Smo to stimulate Sufu/Gli3 separation

and Gli3 recruitment to the cilium tip²¹. Moreover, normal endochondral bone growth needs EVC/EVC2-dependent variations in Hh pathway.

1.4.4 Relation of EVC, Hedgehog Signaling Pathway and Cancer

Takahashi et al. has revealed that overexpression of EVC boosted the survival ability of adult T-cell leukemia (ATL)²². These ATL cells were found to have a substantially higher EVC amount than usual cells. Therefore, EVC might be used as a cell marker and a new drug target for human T-cell leukemia²². From clinical and molecular studies, two new heterozygous splice site mutations in intron 3 and intron 10 of genomic EVC were discovered, which triggered full exon 11 skipping that possibly resulted in lack of crucial operational domains in the protein. It is believed that unusual signaling in Hh pathway can cause cancer. This may happen through two different processes such as mutation-driven signaling and irregular signaling in tumor environment (ligand-dependent signaling)²³. Further EVC and EVC2 studies may reveal new targets for anti-cancer treatment.

1.5 Protein-Protein Interactions

Protein to protein interactions (PPIs) play a significant role in the functioning of living organisms. Several cellular processes such as metabolic pathways, transport, and immune response rely greatly on networks of protein interactions²⁴. When PPIs malfunction, they can be associated with diseases, making them a key target in drug discovery. Protein assemblies come in a wide range of sizes and

shapes, ranging from simple dimers to large viral capsids composed of thousands of protein chains. Protein complexes can also be categorized based on their lifespan within cells, ranging from short-lived transient complexes that last only a few seconds to stable complexes with long-lasting nanomolar binding affinities. Proteins are the building blocks of life and are involved in numerous biological processes²⁵. Understanding their structure and function is crucial for deciphering their roles in health and disease. Accurate prediction of protein structures can provide valuable insights that aid in the interpretation of their functions²⁶. This knowledge is particularly significant in drug discovery and development, where protein targets are identified for potential therapeutic interventions.

1.6 Experimental Methods for Determining Protein Structures

Traditional experimental methods like X-ray crystallography and NMR spectroscopy have been instrumental in determining protein structures. However, these techniques have limitations, including high costs, time requirements, and challenges associated with obtaining high-quality crystals. As a result, experimental structure determination methods are often unable to keep pace with the rapid increase in the number of newly discovered protein sequences²⁷.

1.7 Protein Structure Prediction, Computational Modeling

Protein structure prediction refers to the computational modeling of protein structures based on their amino acid sequences. It involves the application of algorithms that use empirical knowledge, statistical methods, and physical

principles to generate three-dimensional models. These models can provide critical insights into protein functions and facilitate the identification of potential therapeutic targets^{27,28}

1.7.1 Importance of Accurate Protein Structure Prediction in Gaining Insights into Protein Functions

Accurate prediction of protein structures enables the identification and annotation of protein functions. Structural information assists in understanding protein-ligand interactions, substrate binding sites, enzymatic activities, and protein to protein interactions. By predicting protein structures, researchers can infer the functional characteristics of proteins and gain insights into their mechanisms of action²⁹

Protein structure prediction allows for comparative genomics, where protein sequences from different organisms are compared to identify conserved structural motifs and functional domains. This analysis can help predict the functions of newly discovered proteins based on their sequence similarity to proteins with known structures. It provides a valuable tool for functional annotation of uncharacterized proteins and helps reveal evolutionary relationships²⁹.

Accurate protein structure prediction helps in elucidating structure-function relationships, enabling researchers to determine how specific amino acid residues and structural elements contribute to protein function²⁹. This knowledge is

necessary for understanding protein mechanisms and designing targeted interventions²⁷.

1.7.2 Challenges in Protein Structure Prediction

Despite significant advancements, accurate protein structure prediction remains challenging. The complexity of protein folding, conformational flexibility, and limitations in computational methods contribute to the difficulty. Protein structure prediction methods often rely on the availability of homologous structures or templates for modeling, limiting their effectiveness for novel proteins or protein families lacking close structural homologs³⁰.

1.7.3 Modeling of Protein Structure Prediction

Homology modeling is a broadly used protein structure prediction method that relies on the assumption that proteins with similar sequences share similar structures. It has been successful in predicting protein structures with high accuracy when suitable templates are available³¹. For example, the SWISS Model and Phyre2 use templates to discover the model of protein, and most of the time they don't provide a complete model of a protein sequence until it matches the template exactly, which is nearly impossible.

Ab initio methods aim to predict protein structures from scratch, without relying on known templates. These methods employ physics-based models,

energy functions, and molecular dynamics simulations to explore the conformational space and identify energetically favorable structures. Recent advances in computational resources and algorithms have improved the accuracy of ab initio methods, making them more useful for novel protein targets.³¹

Hybrid methods combine the strengths of both homology modeling and ab initio methods. They leverage available templates for regions with homology while employing ab initio techniques for regions without close structural homologs. These approaches have shown promising results and are particularly effective for proteins with complex folding patterns³¹.

1.7.4 Improving Accuracy in Protein Structure Prediction

Integration of experimental data, such as cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry with computational predictions can significantly improve accuracy. These experimental techniques provide additional structural information that can be used to validate and refine predicted models.²⁹

Machine learning and deep learning techniques have been increasingly applied to protein structure prediction. These methods leverage large-scale protein structure databases, training models to learn patterns and predict protein structures with improved accuracy. They have demonstrated promising results, particularly in the identification of protein folding patterns and regions of disorder^{27,31}.

1.8 Alphafold, a Modeling Web Server

Understanding the three-dimensional structure of proteins provides valuable insights into their functions and mechanisms. Despite the vast number of protein sequences available in the Universal Protein Resource (UniProt), the Protein Data Bank (PDB) contains a limited number of 3D structures. This scarcity restricts the coverage of the sequence space, posing challenges to global biomolecular research²⁸.

Expanding the coverage of the sequence space through experimentally determined high-resolution structures is a labor-intensive task. It involves significant trial and error to identify suitable protein constructs and crystallization conditions. Despite advancements in electron cryo-microscopy, hybrid methods, and integrative approaches for structure determination, the disparity between known protein sequences and available experimental structures continues to grow²⁸.

To bridge this gap, one approach is to utilize computational methods to predict the structures of numerous proteins. Researchers are increasingly turning to Artificial Intelligence (AI) techniques to computationally determine the structure of a protein solely based on its amino acid sequence. This approach allows for the prediction of protein structures without the need for experimental determination³².

DeepMind's Alphafold is an advanced AI system renowned for its ability to predict protein structures accurately based on their amino acid sequences²⁸. It has

demonstrated exceptional performance and speed, leading to its recognition as a solution to the protein-structure-prediction problem during the CASP14 benchmark in 2020. This recognition has opened new possibilities in protein structure prediction, allowing the creation of a vast database of structure predictions on a large scale. This breakthrough empowers biologists by providing structural models for nearly any protein sequence, revolutionizing their research approach, and expediting their projects. The methodology of AlphaFold, along with valuable insights gained from predicting the complete human proteome, have been recently described²⁸.

The AlphaFold Protein Structure Database (AlphaFold DB, available at <https://alphafold.ebi.ac.uk>) is a collaborative effort between DeepMind and the EMBL-European Bioinformatics Institute (EMBL-EBI). AlphaFold DB aims to provide the scientific community with open access to a vast collection of protein structure predictions.

The predicted structures in AlphaFold DB provide atomic coordinates as well as per-residue confidence estimates, represented by a confidence metric called pLDDT. These confidence scores range from 0 to 100, with higher values indicating higher confidence in the prediction. The pLDDT scores are based on the model's predicted per-residue scores using the IDDT-C α metric, which is a well-established measure in the protein structure prediction field²⁹.

The IDDT metric aims to evaluate the local accuracy of a prediction by assigning high scores to well-predicted regions, even if the entire prediction may not align perfectly with the true structure. This is especially important for assessing multi-domain predictions, where individual domains may be accurate while their relative positions may not align correctly. pLDDT, as a confidence metric derived from IDDT, reflects the local confidence in the predicted structure. It is particularly useful for assessing confidence within a single domain. Similar IDDT-based metrics are employed in other protein structure prediction resources as well^{29,33}.

In AlphaFold DB, the pLDDT values are stored in the B-factor fields of the mmCIF and PDB files available for download. Additionally, the residues of the models in the 3D structure viewer on the structure pages are color-coded based on confidence bands determined by these values. Residues with pLDDT scores of 90 or above indicate very high confidence, while residues with scores between 70 and 90 are classified as confident. Residues with scores between 50 and 70 have low confidence, and those with pLDDT scores below 50 correspond to very low confidence. Recent studies have shown that protein regions with very low confidence pLDDT scores tend to have higher propensities for intrinsic disorder²⁸.

Another output provided by the AlphaFold system is the Predicted Aligned Error (PAE). PAE indicates the expected positional error at residue *x* if the predicted and actual structures are aligned based on residue *y*, using the C α , N, and C atoms. PAE values are measured in Ångströms and are capped at 31.75 Å.

These values can be utilized by scientists to evaluate the confidence in the relative position and orientation of different parts of the model, such as two domains²⁸.

When assessing the PAE values between residues x and y in two distinct domains, lower PAE values suggest that Alphafold predicts well-defined relative positions and orientations for the domains. Conversely, higher PAE values indicate that the relative position and orientation of the two domains are less reliable. In such cases, users should avoid attributing significant biological or structural relevance to these regions. It's important to note that PAE is asymmetric, meaning there can be a difference between the PAE values for (x, y) and (y, x). This discrepancy may arise, for instance, in loop regions where the orientation is highly uncertain²⁸.

1.9 Protein-Protein Docking

The ClusPro server at <https://cluspro.org> is a widely utilized tool for protein-to-protein docking. It offers a user-friendly interface that only requires two files in Protein Data Bank (PDB) format to initiate the docking process. While the basic use is straightforward, ClusPro also provides advanced options for customizing the search³³.

These advanced options include, removal of unstructured protein regions, application of attraction or repulsion forces, accounting for pairwise distance restraints, construction of homo-multimers, consideration of small-angle X-ray scattering (SAXS) data, and location of heparin-binding sites. Depending on the

type of protein, six different energy functions are available for use. Each energy parameter set leads to the generation of ten models, which are defined by the centers of highly populated clusters of low energy docked structures. This ensemble of models provides a range of potential docking solutions³⁴.

The protocol for using ClusPro involves selecting the desired options, constructing auxiliary restraints files if necessary, choosing appropriate energy parameters, and analyzing the resulting models. Despite its widespread usage, the server is known for its efficiency, with most runs typically completed in less than 4 hours³⁵.

To gain mechanistic insights into these interactions, it is often necessary to obtain atom-level details, ideally through X-ray crystallography. Nonetheless, experimental structure determination can be challenging for biologically important interactions that occur in transient complexes, even when the individual protein structures are known. To address this challenge, computational docking methods have been developed. These methods utilize the known structures of the component proteins as a starting point to predict the structure of their complexes. The goal is to achieve a level of accuracy comparable to X-ray crystallography. Docking algorithms generate multiple detailed models that provide the positions of all atoms in the complex. However, current scoring functions used to assess the quality of these models often lack the necessary accuracy for reliable

discrimination. As a result, it is typically not possible to solely rely on computational tools to identify the model that is closest to the native structure³⁴.

To overcome this limitation, additional information can be incorporated into the model selection process. Lower-resolution experimental techniques, such as site-directed mutagenesis or chemical cross-linking, can provide valuable data that helps in the identification of the most accurate model among the generated docking models. By combining computational docking with these lower-resolution methods, it becomes possible to select models that provide atom-level details and are more likely to represent the native structure of the protein complex³⁵.

Direct docking methods aim to find the structure of the target protein complex that corresponds to the minimum Gibbs free energy in the conformational space. These methods require a computationally feasible model for evaluating free energy and an effective minimization algorithm. The success of direct docking depends on the extent of conformational changes that occur upon protein-to-protein association. If the conformational changes are moderate, direct docking methods can yield accurate results³².

Template-based docking, on the other hand, relies on the observation that interacting protein pairs with more than 30% sequence identity often have similar interaction modes. In template-based docking, the structure of the target protein complex can be obtained using homology modeling tools if an appropriate template complex with a known structure is available. While the original application of

template-based docking requires a high degree of sequence identity, it has been expanded to include partial structures representing the interface region as templates. However, the coverage of the template space is still limited, and therefore, direct docking methods are generally more useful in many applications where suitable templates are not available^{30,33,36}.

1.9.1 ClusPro, a Web Server for Protein-Protein Docking

Since its introduction in 2004, ClusPro has undergone significant modifications and expansions. The server performs three computational steps to predict protein to protein complexes. Firstly, it employs rigid body docking by sampling billions of conformations. Secondly, it applies root-mean-square deviation (RMSD) based clustering to identify the largest clusters among the 1,000 lowest-energy structures, which represent the most likely models of the complex. Finally, selected structures undergo refinement through energy minimization³⁴.

The rigid body docking step in ClusPro utilizes PIPER, a docking program based on the Fast Fourier Transform (FFT) correlation approach. The FFT method, pioneered by Katchalski-Katzir et al. in 1992, revolutionized rigid-body protein to protein docking. In this approach, one protein (the receptor) is placed at the origin of the coordinate system on a fixed grid, while the other protein (the ligand) is placed on a movable grid. The interaction energy is expressed as a correlation function or a sum of correlation functions. The computational efficiency of FFT-

based methods lies in the ability to efficiently calculate energy functions using FFTs, enabling exhaustive sampling of billions of conformations and energy evaluations at each grid point. Consequently, the FFT-based algorithm allows docking of proteins without prior knowledge of the complex structure³⁴.

Initially, Katchalski-Katzir et al. used a simple scoring function that considered only shape complementarity. However, subsequent methods based on FFT correlation docking introduced more sophisticated and accurate scoring functions, including terms for electrostatic interactions or both electrostatic and desolvation contributions. One key factor contributing to the success of rigid body methods is the tolerance for some overlaps due to the shape complementarity term. This allows the methods to accommodate moderate differences between the structures of the bound and unbound states^{34,35}.

In the current version of ClusPro, the FFT correlation method implemented through PIPER utilizes a scoring function that incorporates a structure-based pairwise interaction term. This combination of terms in the energy function significantly enhances docking accuracy, resulting in the generation of more near-native structures³⁵.

1.9.2 ZDock, a Protein Docking Program

ZDock is a protein docking program that utilized a fast Fourier transform for efficient 3D searching of spatial degrees of freedom between two molecules. By

incorporating a pairwise statistical potential into the scoring function of ZDock, the recent improvements make ZDock more suitable for computationally intensive tasks such as docking flexible molecules and modeling interactomes. Additionally, the improved efficiency allows ZDock to be more readily utilized by researchers with limited computational resources^{37–39}.

CHAPTER TWO: MATERIALS AND METHODS

Figure 8 on Page 33 displays a flow chart illustrating the experimental procedure. For a comprehensive list of programs and websites utilized, along with concise explanations of their functionalities and availability, refer to the List of Programs and Websites section located in Appendix on Page 78.

2.1.1 Obtaining Primary Sequences

Sequence for mouse EVC (mEVC) protein (Ellis-Van Creveld syndrome protein homolog [Mus musculus]; ACCESSION: NP_067267) and mouse EVC2 (mEVC2) protein (Limbin isoform X1 [Mus musculus]; ACCESSION: XP_036021349) obtained from The National Center for Biotechnology Information (NCBI) advances science and health (<https://www.ncbi.nlm.nih.gov/protein>). Sequence for human EVC (hEVC) protein (EVC complex member EVC isoform 1 [Homo sapiens]; ACCESSION: NP_714928) and human EVC2 (hEVC2) protein (EVC2 [Homo sapiens]; ACCESSION: AAO22066; VERSION: AAO22066.1) obtained from NCBI.

2.2 Alignment

Alignment of mEVC with mEVC2, full sequence mEVC with full Sequence mEVC2 and fragment for EVC with fragment for EVC2, were done in Uniprot

webserver (<https://www.uniprot.org/align>). Then alignment, full sequence mEVC with full sequence hEVC and fragment of mEVC2 with fragment for hEVC2, were done in Uniprot webserver.

2.3 Finding Coiled-Coil Region

The primary sequences of all four proteins, mEVC, mEVC2, hEVC, hEVC2, were uploaded to Prabi, Institute of Biology and Protein Chemistry (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_lupas.html) for coiled-coil prediction.

2.4 Model Generation

Then the primary sequences were uploaded to Alphafold colab (<https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>) individually to generate 3D model. Also a fragment of mEVC, amino acid residues number 451 to 750 and mEVC2, amino acid residue 247 to 661, was uploaded to Alphafold colab to generate the model for the fragment.

2.5 Superimposing of Proteins

Superimposition of two proteins, like full sequence mEVC with full sequence mEVC2 and fragment for mEVC with fragment for mEVC2, was done in Pymol. Superimposition of full Sequence mEVC with full sequence hEVC and fragment for mEVC2 with fragment of hEVC2 were done in Pymol.

2.6 Docking of mEVC and mEVC2

Four docking of mEVC and mEVC2, full sequence mEVC, full sequence mEVC2, full sequence mEVC, fragment of EVC2, fragment for EVC, full sequence mEVC2, fragment for EVC, and fragment for EVC2 were done in Cluspro webserver (<https://cluspro.org/home.php>). The protein model generation jobs were performed on a CPU server. As each model consisted of a single chain, there was no need to specify any additional chains for interaction analysis. Furthermore, no advanced options were chosen for the job, indicating that the default settings of the CPU server were used for the modeling process.

2.7 Obtaining K_d Values and Interaction Interfaces

The models generated through ClusPro docking were downloaded and a chain sequence renamed using Pymol. mEVC as chain A and EVC2 as chain C. The molecules were exported from Pymol and saved to the drive. This file was then uploaded to the PRODIGY webserver for the purpose of obtaining K_d values of the models. In the analysis, Interactor 1 was designated as "A" representing mEVC, while Interactor 2 was labeled as "B" representing mEVC2. The temperature was maintained at 25.0°C throughout the process. The job was submitted to the PRODIGY webserver, which subsequently provided a table containing various values such as K_d and ΔG for each model. To facilitate further analysis, the table was copied into an Excel file.

2.8 Modeling of Interaction Interfaces

The modeling of interaction interfaces was conducted using Pymol, a visualization software. Models resulting from the docking of mEVC and mEVC2, which exhibited the desired interactions, were imported into Pymol. The amino acids involved in the interactions were modified to be displayed in a ball-and-stick representation and labeled accordingly. The distance function in Pymol was employed to measure the distances between the atoms of the interacting amino acids, facilitating the identification of the bonding interaction types. The distances were annotated, and a snapshot of the interaction was captured.

2.9 Mutagenesis

The mutagenesis process involved utilizing the Mutagenesis Wizard in PyMol. Only mutations on mEVC, 552-GLU to ALA, were performed, with each model undergoing a single mutation. The mEVC model was opened in PyMol, and the specific residue to be mutated was selected. Using the Mutagenesis Wizard, the desired mutation was chosen and applied to the model. The mutated model was then given a new name and saved. Subsequently, the mutated models were subjected to docking with mEVC2 using ClusPro. The resulting models were further analyzed in PRODIGY to obtain their respective K_d values and interaction interfaces. Interactions involving the mutated residue were documented and compared to the interactions observed in the non-mutated docking models.

2.10 Docking of mEVC and mEVC2 by Zdock

Again, the full sequence mEVC and the full sequence mEVC2, the full sequence mEVC and the fragment of EVC2, the fragment for EVC and the full sequence mEVC2, the fragment for EVC and the fragment for EVC2 were docked by the utilization of a different webserver, Zdock (<https://zdock.umassmed.edu/>). Uploaded the PDB files following the same protocol of ClusPro. The protein model generation jobs were performed on a CPU server. As each model did not consist of a single chain, the chain name for both mEVC and mEVC2 were changed to chain A and chain C respectively. Furthermore, no advanced options were chosen for the job, indicating that the default settings of the CPU server were used for the modeling process.

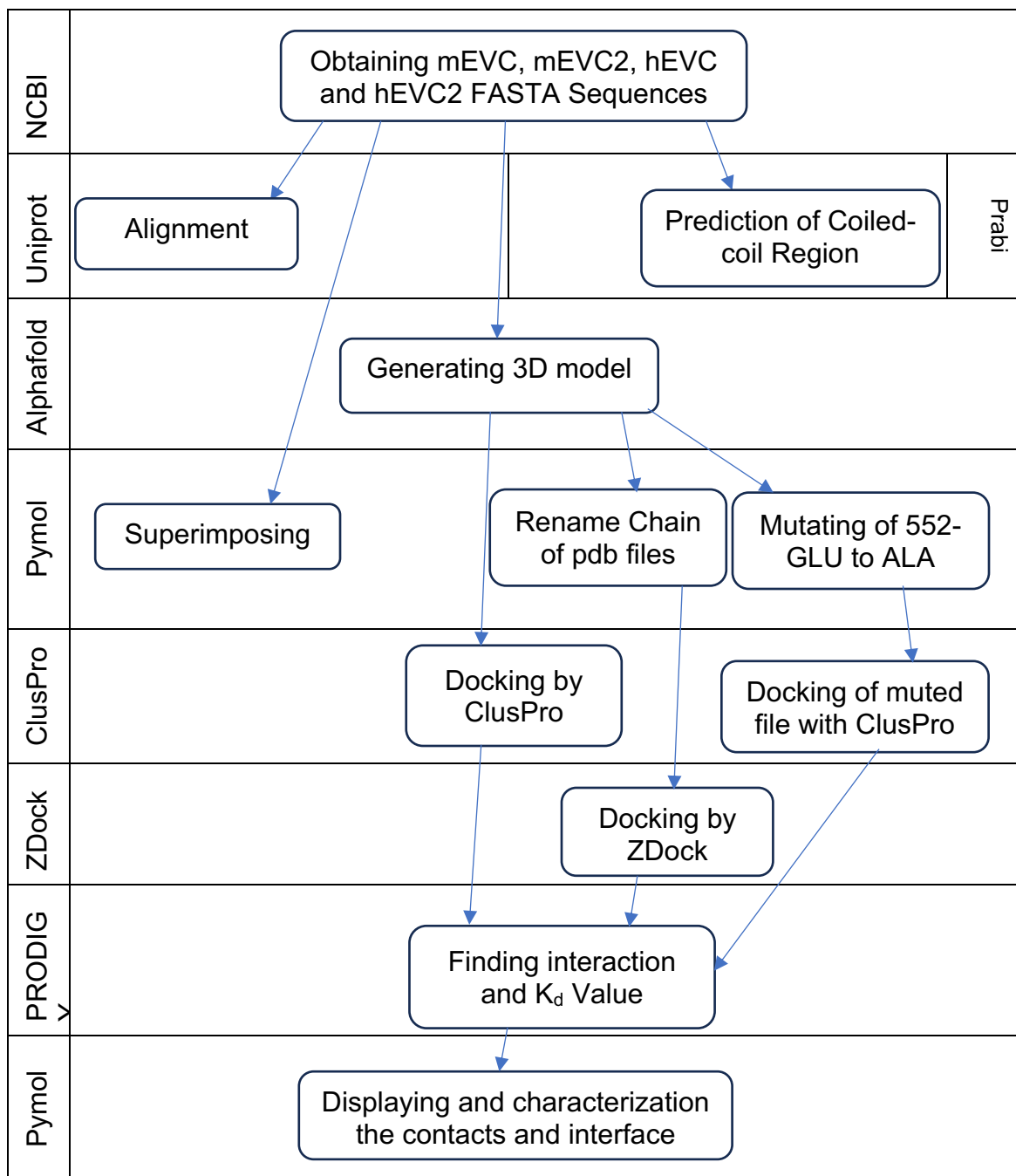


Figure 8 Flowchart of the experimental procedure. Shows the steps taken and what website/program was utilized.

CHAPTER THREE: RESULTS

3.1 Coiled-Coil Region of mEVC and hEVC

Coiled-coils are structures composed of α -helices that are tightly wound around each other, forming super-helical bundles. Typically, a coiled-coil consists of two or three helices arranged in either a parallel or antiparallel orientation.

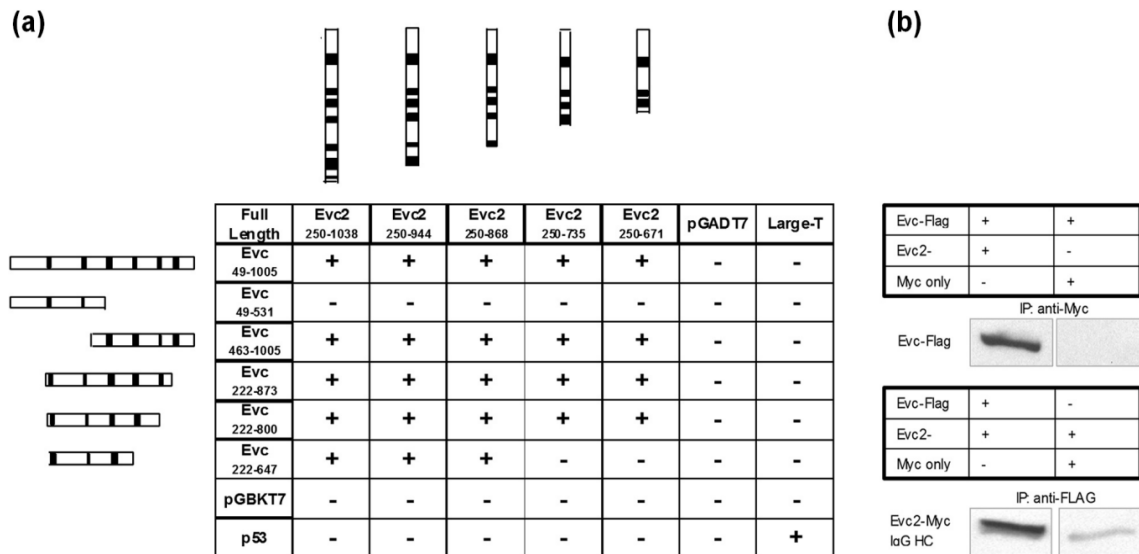


Figure 9 Protein-protein interaction of EVC and EVC2. (a). Directed yeast two-hybrid assay. For each construct, the portion of EVC and EVC2 coding sequence expressed is depicted. Predicted coiled-coil regions are indicated by black boxes. Negative controls (empty vectors pGBKT7 and pGADT7) and positive controls (p53 and Large-T antigen) were included. '+' indicates colony growth/interaction, '-' indicates no colony growth/lack of interaction. (b). FLAG-tagged EVC is immunoprecipitated by Myc-tagged Evc2 but not by the Myc epitope alone. Myc-tagged EVC2 is

immunoprecipitated by FLAG-tagged EVC but not by the Flag epitope alone. A non-specific band corresponding to the IgG heavy chain (HC) is indicated⁴⁰.

To identify proteins interacting with EVC, Blair et al⁴⁰ conducted a yeast two-hybrid assay using a cDNA library derived from E11 mouse embryos. The assay utilized the EVC sequence encoding amino acids 49-1005 as bait⁴⁰. They observed an interaction between EVC and EVC2.

Blair et al. observed significant binding with several EVC constructs, including those expressing amino acids residue number 49-1005, 463-1005, 222-873, and 222-800, when tested against EVC2. However, no contacts were observed with the EVC construct expressing amino acids 49-531, and restricted growth was seen with the construct encoding amino acids 222-647. These results indicate that the interaction primarily occurs in the third and fourth coiled-coil regions of EVC, with limited interaction observed in constructs containing only the first three coiled-coil regions and no interaction detected in constructs containing only the first two coiled-coil regions. Although the fifth and sixth coiled-coil regions were present in constructs showing interaction, they were not individually tested.

Additionally, significant binding was observed between the EVC2 construct expressing amino acids 250-671, which contains the first three predicted coiled-coil regions, and EVC (Figure 9a). This finding suggests that these regions are crucial for the interaction between EVC and EVC2.

Overall, their results shed light on the specific regions involved in the interaction between EVC and EVC2, highlighting the importance of the third and fourth coiled-coil regions in EVC and the first three predicted coiled-coil regions in EVC2⁴⁰.

Based on the experiment of Blair et al, the third and the fourth coiled-coil region of mEVC protein for our research were considered. P-loop and leucine zipper are in this fragment region.

The full length primary sequence of the mouse EVC protein were uploaded to Prabi (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_lupas.html). It predicted six coiled-coil region that situated in the amino sequence number around amino acid sequence 250-280, 450-490, 560-580, 700-750, 790-830, 850-900. So, the fragment from residue number 451 to 750, which include coiled-coil regions second, third and fourth, leucine zipper and p-loop, were considered.

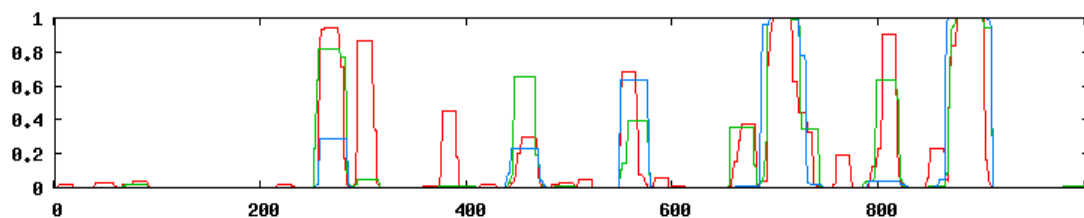


Figure 10 Predicted coiled-coil region of mouse EVC Protein (NP_0672672):

According to the prediction results, it can be observed that mEVC2 protein (as shown in the figure) possesses the first three coiled-coil regions located

between amino acid residue 247-661. These regions were specifically chosen for further experimentation in our study.

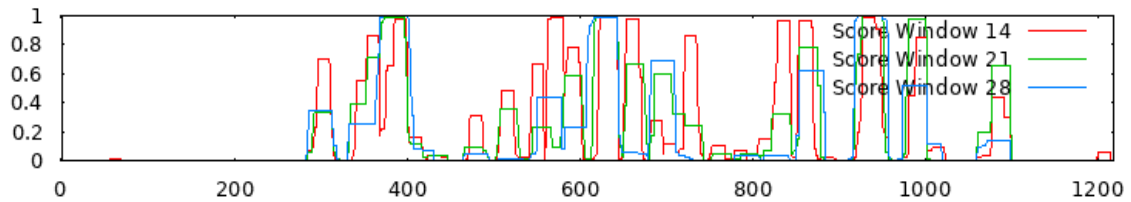


Figure 11 Predicted coiled-coil region of mouse EVC2 Protein (limbin isoform X1 [Mus musculus]
NCBI Reference Sequence: XP_036021349.1)

3.2 Alignment

Table 1 Alignment of mouse EVC, EVC2 and human EVC and EVC2

Protein 1	Identical	Protein 2
Human EVC (NP_714928.1)	68.29%	Mouse EVC (NP_067267.2)
Human EVC2 (AAO22066.1)	71.70%	Mouse EVC2 (NP_666032.1)
Human EVC (NP_714928.1)	24.51%	Human EVC2 (AAO22066.1)
Mouse EVC (NP_067267.2)	21.05%	Mouse EVC2 (NP_666032.1)

Based on the alignment analysis, it appears that the mouse and human EVC and EVC2 proteins exhibit approximately 70% sequence identity. Additionally, the EVC and EVC2 proteins show more than 20% sequence identity in both mouse and human species.

3.3 Modeling of Proteins

The primary sequences of mouse and human EVC and EVC2 proteins were uploaded to AlphaFold, and the platform generated predicted 3D models for analysis.

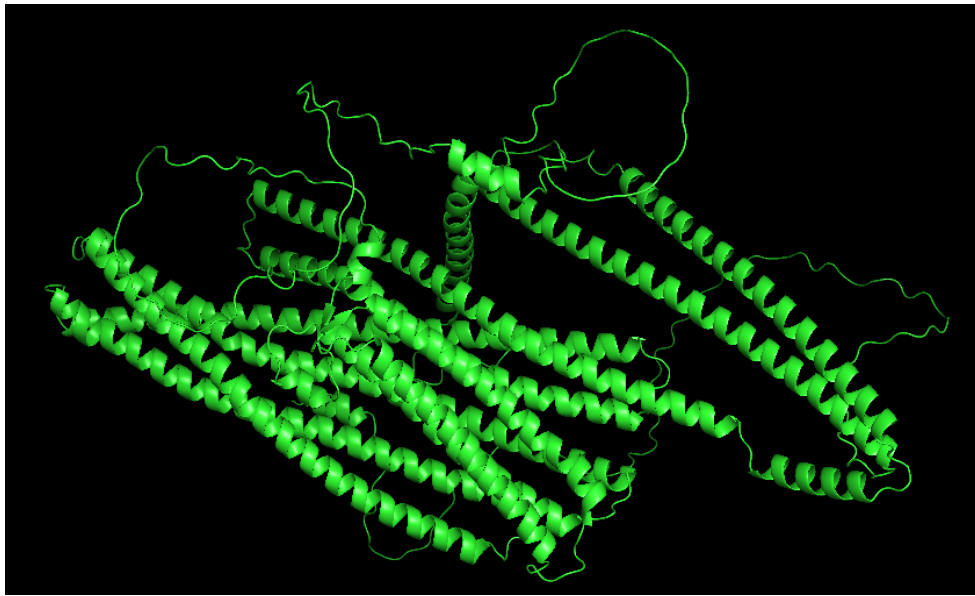


Figure 12 3D model of full-length mouse EVC Protein (NP_0672672_75e86)

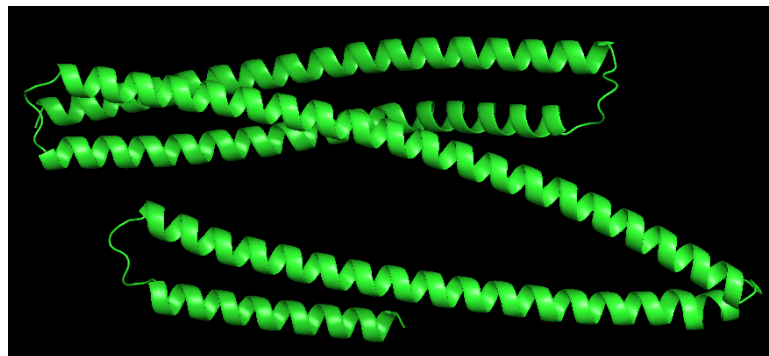


Figure 13 3D model of the fragment of mouse EVC protein (residue 451 to 750) (rank_1_model_5)

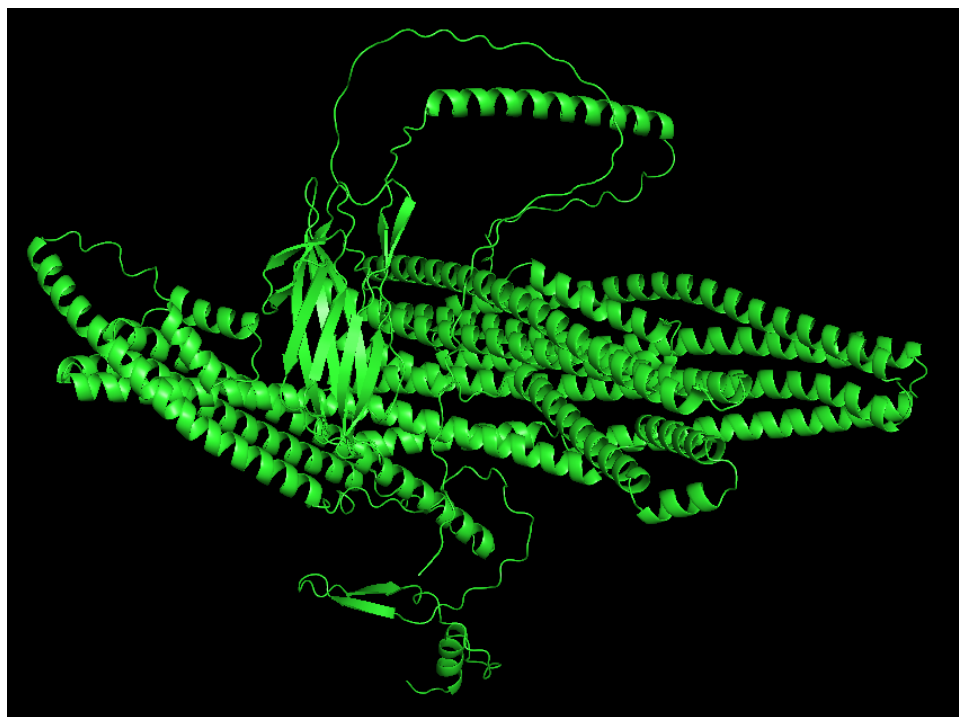


Figure 14: 3D model of mouse EVC2 protein (Source: Uniprot_AF-Q8K1G2-F1-model_v2)

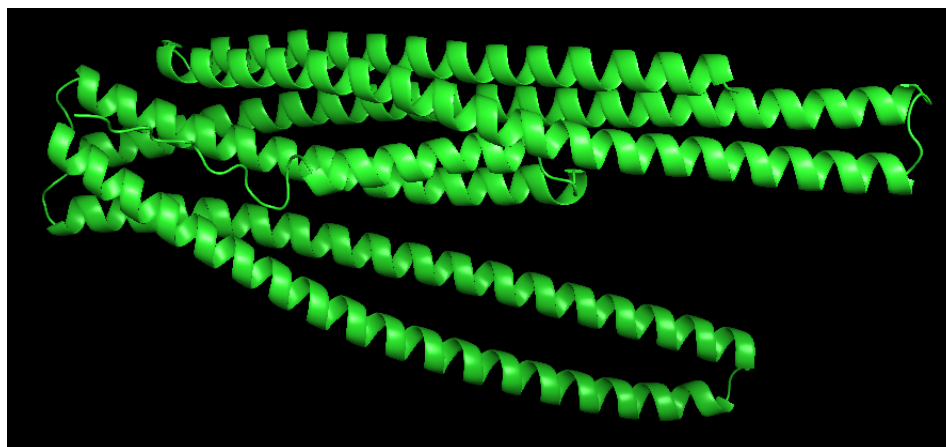


Figure 15 3D model of the fragment of mouse EVC2 Protein (246 to 661) (rank_1_model_3)

3.4 Docking by ClusPro

The generated predicted models of the proteins were downloaded in the PDB file format and subsequently uploaded to ClusPro for further analysis. The following docking configurations were performed: a. Fragment of mouse EVC protein with full mouse EVC2; b. Full mEVC with fragment mEVC2; c. Fragment mEVC with fragment mEVC2; d. Full mEVC with full mEVC2.

The resulting docked configurations were then downloaded in the PDB file format for further analysis.

3.4.1 Interaction between fragment mEVC and full mEVC2

Based on the analysis, it has been determined that there are no contacts observed between the interface of the fragment mEVC (451 to 750) and the fragment mEVC2 (247 to 661).

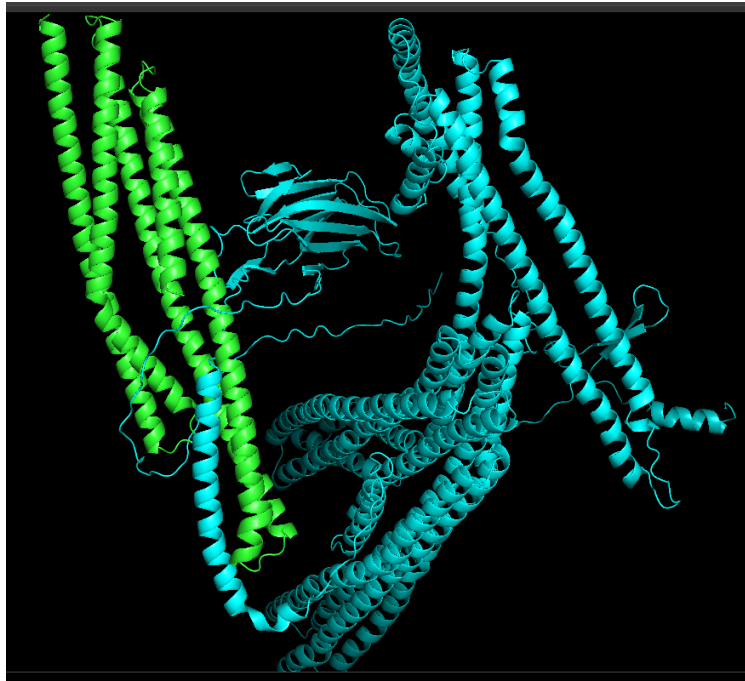


Figure 16 Interaction between fragment mEVC (451 to 750) and full length mEVC2

3.4.2 Ful length mEVC and fragment (247 to 661) mEVC2

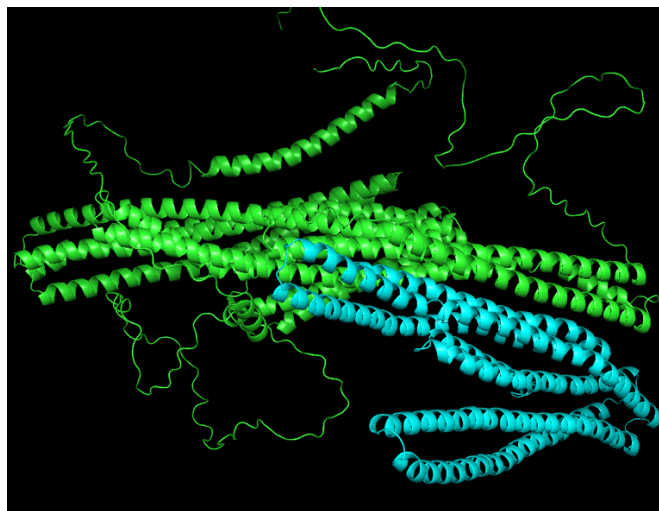


Figure 17 Interaction full length Mouse EVC and fragment (247 to 661) mouse EVC2

Table 2 List of the contacts between full length mEVC and fragment (residue 247 to 661) mEVC2 interface

mEVC	mEVC
HIS-611	GLU-659
THR-615	LYS-656
VAL-619	SER-652, GLU-655
ARG-622	VAL-651
PHE-623	VAL-648
SER-624	ARG-647
GLN-737	SER-652
LEU-738	ARG-580
GLY-745	ARG-584, ASP-645
LEU-748	ARG-584, ASP-645, VAL-648
LEU-749	ARG-587, GLU-588

3.4.3 Interaction between both mEVC and mEVC2 fragment protein

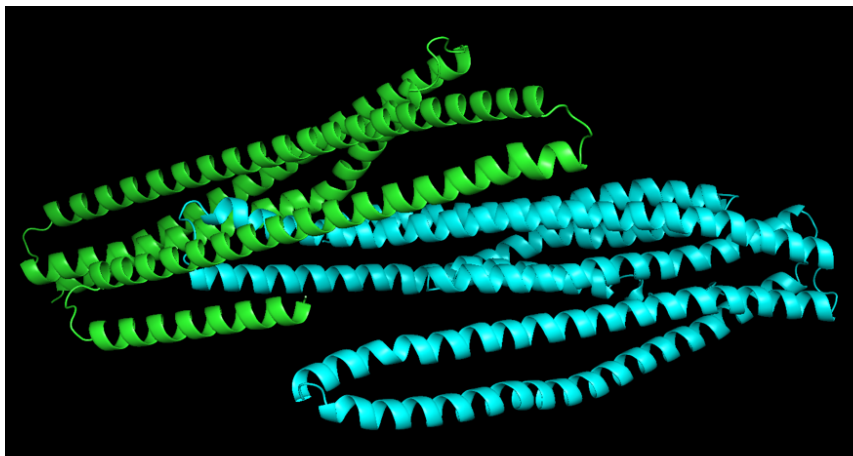


Figure 18 Interaction between the fragment mouse EVC protein and fragment mouse mEVC Protein

Table 3 List of the contacts between the fragment of mEVC (451 to 750) protein and fragment of mEVC2 (247 to 661) protein interface

mEVC	mEVC2
ALA-453	ARG 400, GLU 404
ARG-454	ARG 400, GLU 404
GLU-457	ARG 400, GLU 404
GLN-466	GLU 545
ARG-470	GLU 545, ASP 546
PHE-490	ASP 546
HIS-491	ASP 544, GLU 545, TYR 542, LEU 543, GLY 541
LEU-494	GLU 545, ASP 546, ASP 544
GLU-495	GLU 545, GLY 541, ARG 539, LEU 543
GLN-497	GLU 545
ARG-498	ARG 539, GLU 545, GLY 541, GLU 538, MET 548, ASP 544, TYR 542, LEU 543, ALA 540
ARG-501	GLU 549, GLU 545, LEU 552, MET 548
SER-502	ARG 539, HIS 535
GLU-505	HIS 535, ARG 539
ASP-590	LYS 536, GLN 529, LEU 533
LYS-591	LYS 536
VAL-593	HIS 537
TRP-594	HIS 537, ILE 534, LYS 536, GLN 529, LEU 533
LEU-595	HIS 537
GLU-597	TYR 542, HIS 537
GLY-598	HIS 537, TYR 542
THR-599	HIS 537, TYR 542
SER-601	TYR 542
THR-602	TYR 542
GLN-605	TYR 542
ARG-647	GLY 541, GLN 547, ASP 544, TYR 542, LEU 543
ALA-650	TYR 542, LEU 543
LEU-651	LEU 543, LEU 551, GLN 547
GLY-653	HIS 537
THR-654	GLU 538, HIS 537, TYR 542, LEU 543, LEU 551
THR-655	LEU 551, ARG 554
THR-657	HIS 537, LEU 533
ALA-658	GLN 529, LEU 533, ILE 534, LEU 530

mEVC	mEVC2
LEU-659	GLU 558, LEU 533, LEU 530, LEU 661
GLN-661	LEU 533, LEU 533
MET-662	GLN 529, LEU 533, LEU 530, THR 658, THR 658, THR 559, LEU 661, ALA 527
ARG-663	LEU 661, GLU 659, GLU 660, THR 658
LEU-664	THR 658
SER-665	THR 658, ALA 526, GLN 529, LEU 522
GLY-666	LEU 522, THR 658
LYS-667	THR 658, GLU 659
LYS-668	GLN 529
ARG-669	LEU 522, ARG 518
LEU-670	ARG 518
GLU-673	ARG 518
GLU-713	GLU 659

3.4.4 Interaction between full length mEVC and full length mEVC2 proteins

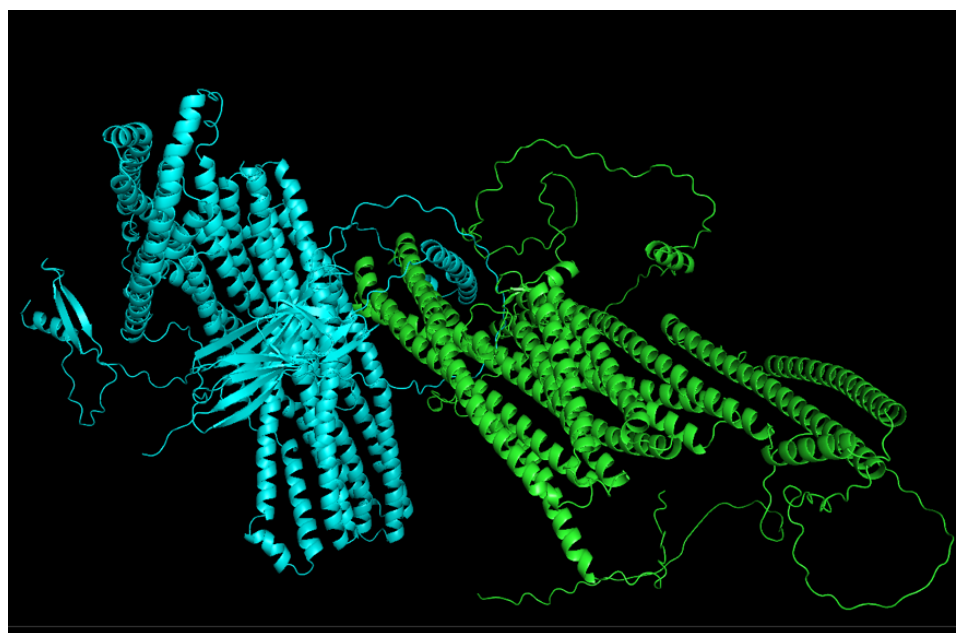


Figure 19 Interaction between full length mEVC and full-length mEVC2 proteins

Table 4 List of the contacts between full length mEVC and mEVC2 interface

mEVC	mEVC2
LEU-540	ASN 262
LYS-541	ASN 262, GLN 354, GLU 358
PRO-544	ALA 261, ASN 262
GLU-545	PRO 255, SER 256, PRO 257, SER 260, ASN 262
SER-548	ALA 261, GLN 597
LEU-549	ALA 261, SER 265, LYS 594
PRO-550	ALA 261, SER 265, LEU 268, LEU 590, LYS 594
VAL-551	ALA 261, ASN 262, SER 265
ALA-552	SER 265, SER 269, ASP 272, ARG 586
GLU-553	ARG 586, LEU 590, LYS 594
THR-556	ARG 586

Upon analysis, it was observed that there is no interaction between the p-loop region (residue 565-572) and the Leucine zipper region (635-656). The K_d and ΔG values of the first four predicted models generated by Alphafold were examined for further evaluation.

Table 5 Compare of K_d value and ΔG for top five docked models

Protein-protein complex	ΔG (kcal mol ⁻¹)	K_d (M) at °C
model.000.00	-21.0	3.7e-16
model.000.01	-20.0	2.2e-15
model.000.02	-20.0	2.3e-15
model.000.03	-21.4	2.1e-16
model.000.04	-21.4	2.1e-16

Table 6 List of the bond length and probable bond type in between the interface of fragment mEVC and mEVC2 proteins when docked both protein's full length.

mEVC	mEVC2	Distance (Å) minimum	Bond/force Type
LEU-540	ASN-262	O-H=2.2, O-H=3.2	Strong bond
LYS-541	ASN-262	O-H=4.2, H-H=2.2	Van der Waals
LYS-541	GLN-354	H-H=3.5	Van der Waals
LYS-541	GLU-358	O-H=1.7	strong
PRO-544	ALA-261	O-H=3.6	weak, electrostatic
GLU-545	PRO-255	O-H=2.9	mostly electrostatic
GLU-545	SER-256	N-O=3.3	Van der Waals
GLU-545	PRO-257	O-H=2.7	mostly electrostatic
GLU-545	SER-260	O-H=4.4, H-H=3.4	Van der Waals
GLU-545	ASN-262	O-H=4.3	Van der Waals
SER-548	ALA-261	O-H=3.3	weak, electrostatic
SER-548	GLN-597	O-H=2.8	strong
LEU-549	ALA-261	O-H=3.2	mostly electrostatic
LEU-549	SER-265	O-O=4.6	Van der Waals
LEU-549	LYS-594	H-H=3.7	Van der Waals
PRO-550	ALA-261	H-H=2.7	Van der Waals
PRO-550	SER-265	O-H=2.4	strong
PRO-550	LEU-268	H-H=3.4	Van der Waals
PRO-550	LEU-590	H-H=3.1	Van der Waals
PRO-550	LYS-594	N-H=2.8, H-H=2.2	electrostatic & Van der Waals
VAL-551	ALA-261	H-H=3.6	Van der Waals
VAL-551	ASN-262	O-H=3.6	weak, electrostatic
VAL-551	SER-265	O-H=2.1	strong
ALA-552	SER-265	O-H=4.1, H-H=2.8	Electrostatic & Van Der Waals
ALA-552	SER-269	H-H=2.7	Van der Waals
ALA-552	ASP-272	O-H=3.6	weak, electrostatic
ALA-552	ARG-586	O-H=3.5	weak, electrostatic
GLU-553	ARG-586	H-H=4.0	Van der Waals
GLU-553	LEU-590	O-H=2.4	strong
GLU-553	LYS-594	O-H=2.1, O-H=1.8	strong
THR-556	ARG-586	O-H=3.8	weak, electrostatic

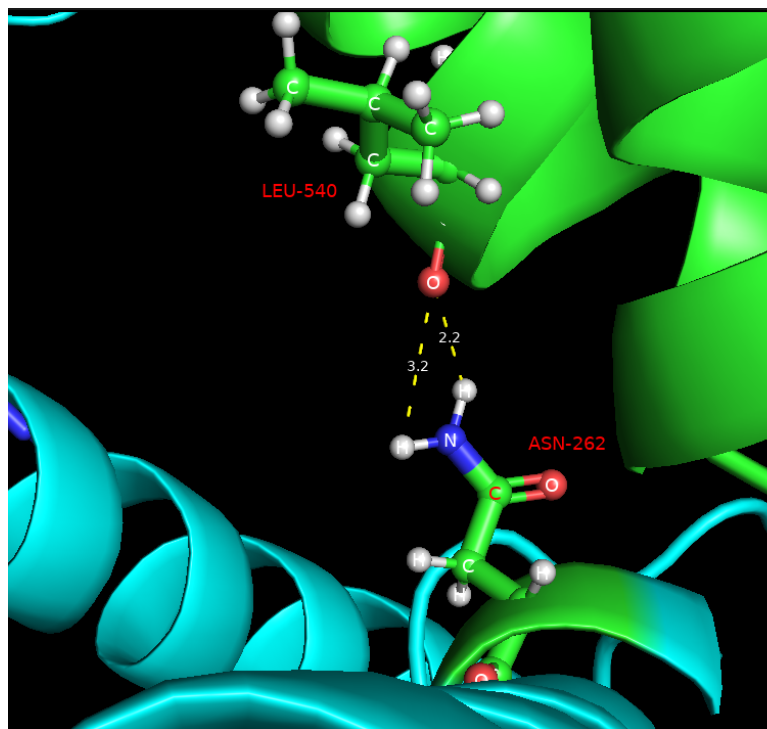


Figure 20 Contacts of LEU-540 of mEVC to ASN-262 of mEVC2 in Model 000

LEU-540 of mEVC shows a potential strong hydrogen bond with the ASN-262 of mEVC2. The minimum distance of Oxygen atom of LEU-540 to the Hydrogen of ASN-262 is only 2.2 Å.

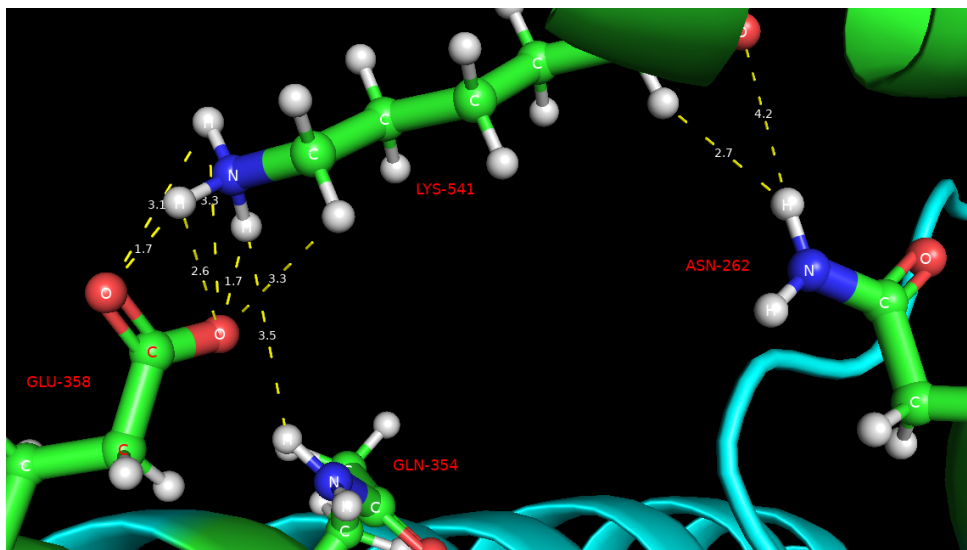


Figure 21 Contacts of LYS-541 of mEVC to ASN 262, GLN 354, GLU 358 of mEVC2 in Model 000.

LYS-541 of mEVC interacts with ASN 262, GLN 354, GLU 358 of mEVC2. LYS-541 shows a potential strong hydrogen bond with GLU-358 of mEVC2. The minimum distance between the oxygen atom of LYS-541 of mEVC and oxygen of GLU-358 of mEVC2 is 1.7 Å. LYS-541 of mEVC shows a weak Van der Waals force with ASN-262 and GLN-354 of mEVC2.

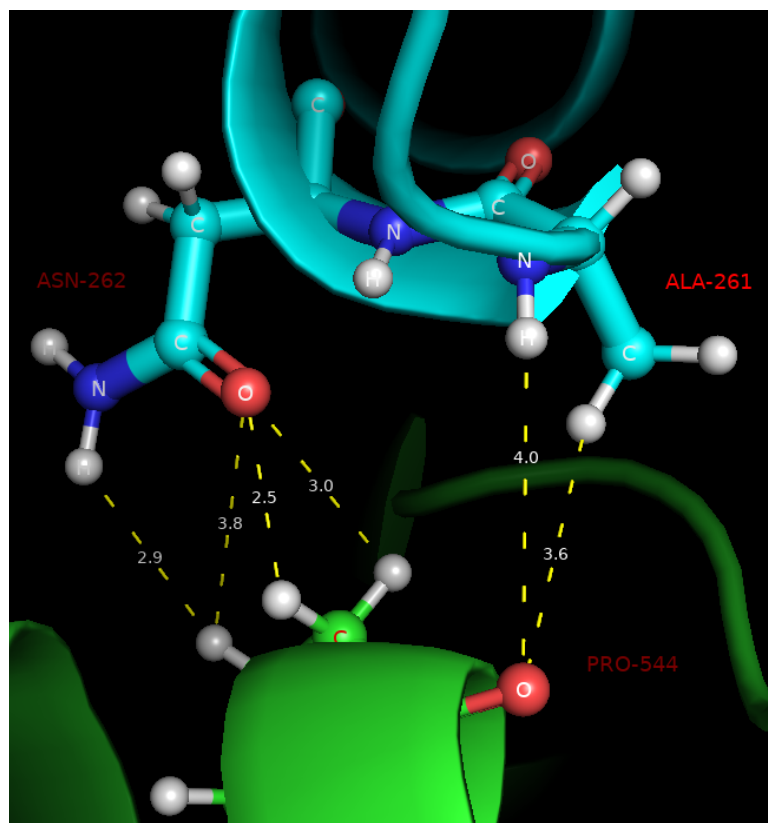


Figure 22 Contacts of PRO-544 of mEVC to ASN 262, ALA-262 of mEVC2 in Model 000

PRO-544 of mEVC shows a potential strong Van der Waals force with ASN-262, and a weak Van der Waals force with ALA-261 of mEVC2. Minimum distance of hydrogen atom of PRO-544 of mEVC to oxygen atom of ASN-262 of mEVC2 is 2.5 Å and to ALA-261 is 3.6 Å.

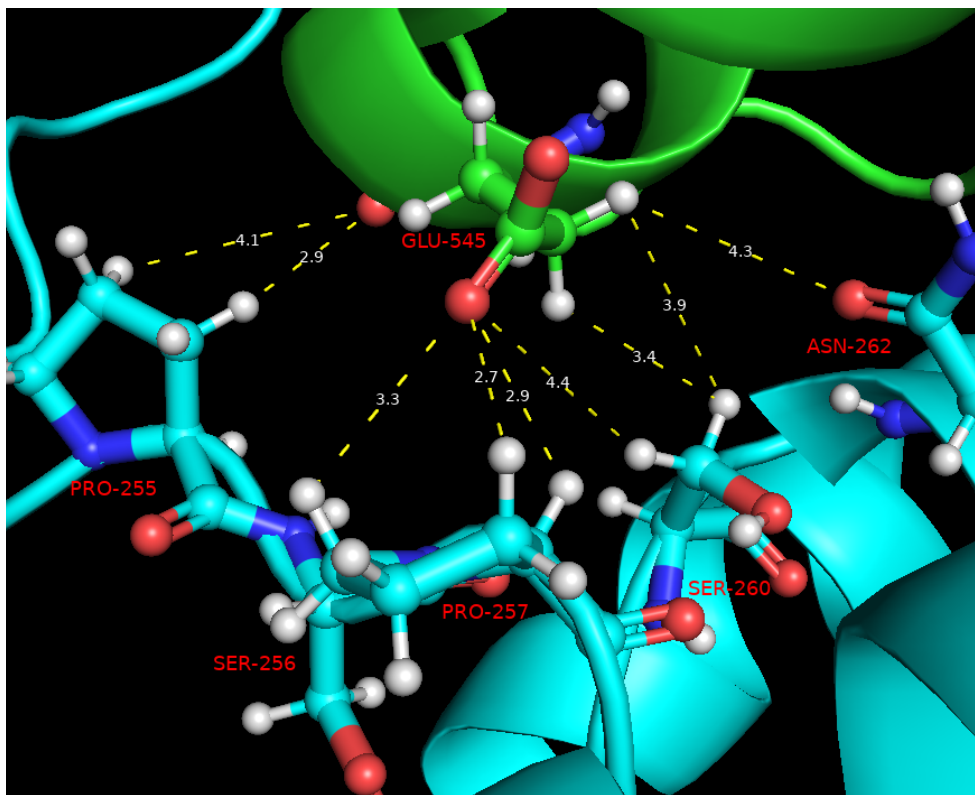


Figure 23 Contacts of GLU-545 of mEVC to PRO 255, SER 256, PRO 257, SER 260, ASN 262 of mEVC2 in Model 000.

GLU-545 of mEVC shows contacts with PRO 255, SER 256, PRO 257, SER 260, ASN 262 of mEVC2. The oxygen atom of GLU-545 has a probable hydrogen bond with the hydrogen atom of PRO-255, SER-256, PRO- 257, and a weak electrostatic force to SER-260. The minimum distance is 2.9 Å, 3.3 Å, 2.7 Å, and 4.4 Å. The hydrogen atom of GLU-545 of mEVC2 has a probable electrostatic force to the oxygen atom of ASN-262 of mEVC2.

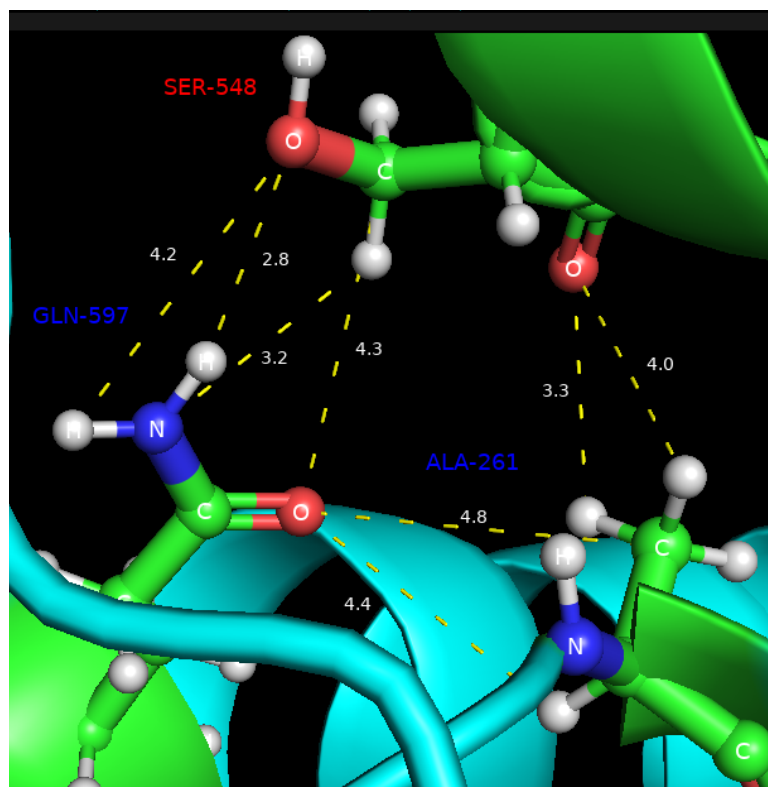


Figure 24 Contacts of SER-548 of mEVC to ALA-261, GLN-597 of mEVC2 in Model 000

SER-548 of mEVC shows contacts with ALA-261, GLN-597 of mEVC2. The oxygen atom of SER-548 of mEVC shows a probable hydrogen bond with the hydrogen atom of GLN-597 of mEVC2. The minimum distance is 2.8 Å. SER-548 also shows a Van der Waals force with ALA-261 of mEVC2. The minimum distance between the oxygen atom of SER-548 of mEVC and hydrogen atom of ALA-261 of mEVC2 is 3.3 Å.

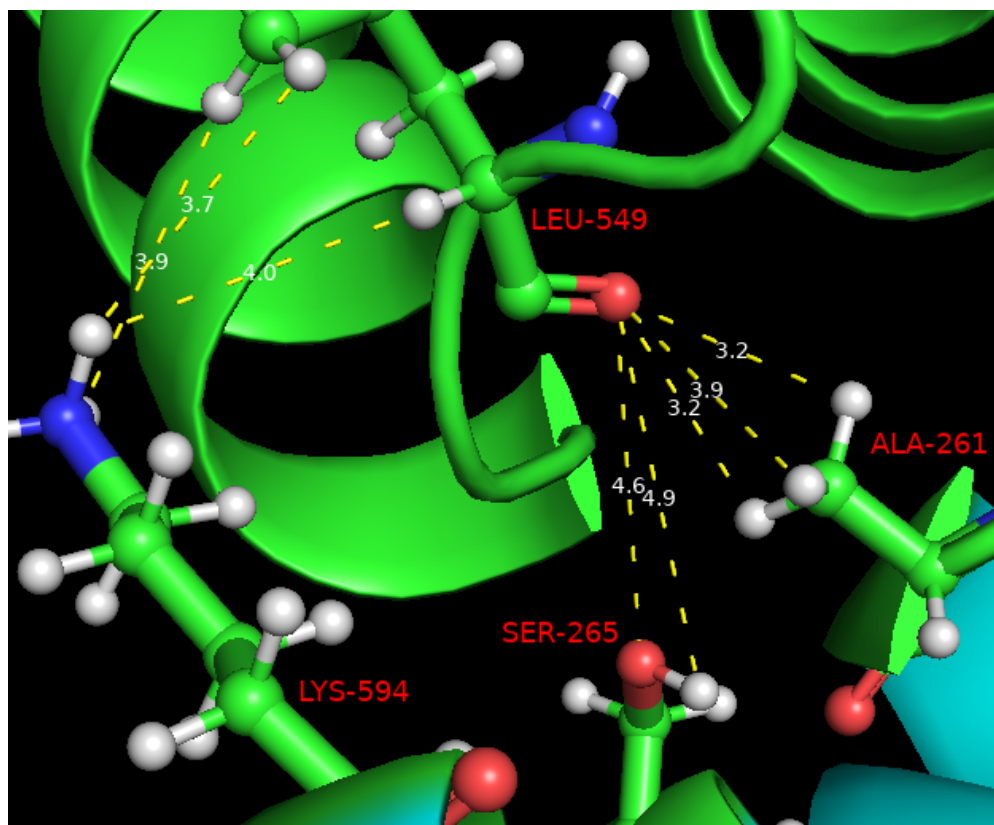


Figure 25 Contacts of LEU-549 of mEVC to ALA 261, SER 265, LYS 594 of mEVC2 in Model 000

LEU-549 of mEVC shows contacts with ALA 261, SER 265, LYS 594 of mEVC2. The oxygen atom of LEU-549 of mEVC has a probable moderate hydrogen bond with the hydrogen atom of ALA-261 of mEVC2. SER 265, and LYS 594 of mEVC2 show a probable Van der Waals force to LEU-549 of mEVC.

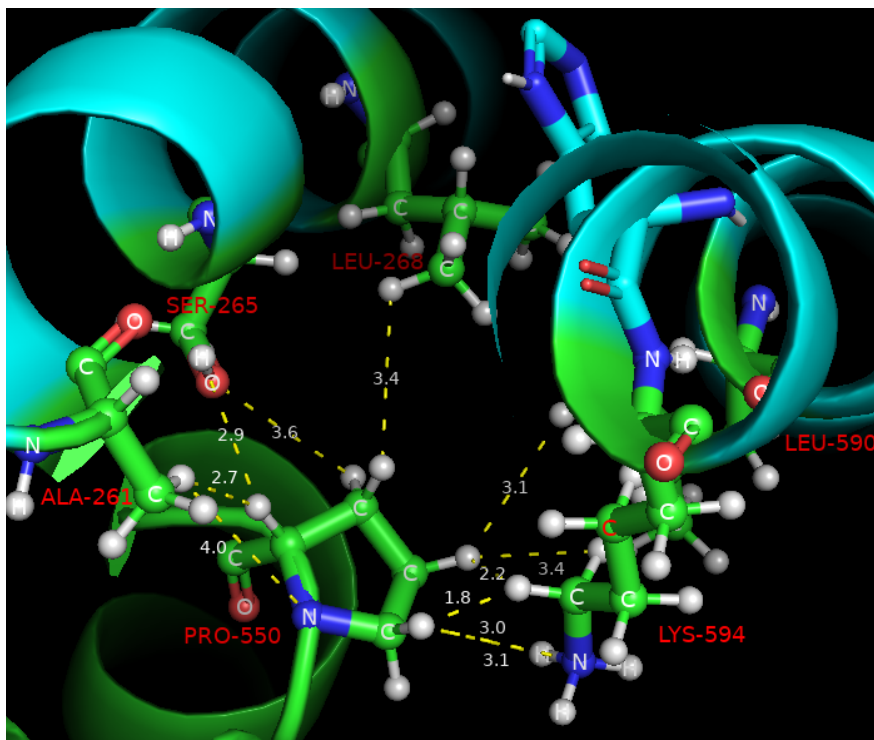


Figure 26 Contacts of PRO-550 of mEVC to ALA 261, SER 265, LEU 268, LEU 590, LYS 594 of mEVC2 in Model 000

PRO-550 of mEVC shows contact with ALA 261, SER 265, LEU 268, LEU 590, LYS 594 of mEVC2. The hydrogen atom of PRO-550 of mEVC shows a probable strong hydrogen bond with the oxygen atom of SER-265 of mEVC2. The distance is 2.9 Å. PRO-550 of mEVC shows a probable Van der Waals force to ALA 261, LEU 268, LEU 590, LYS 594 of mEVC2.

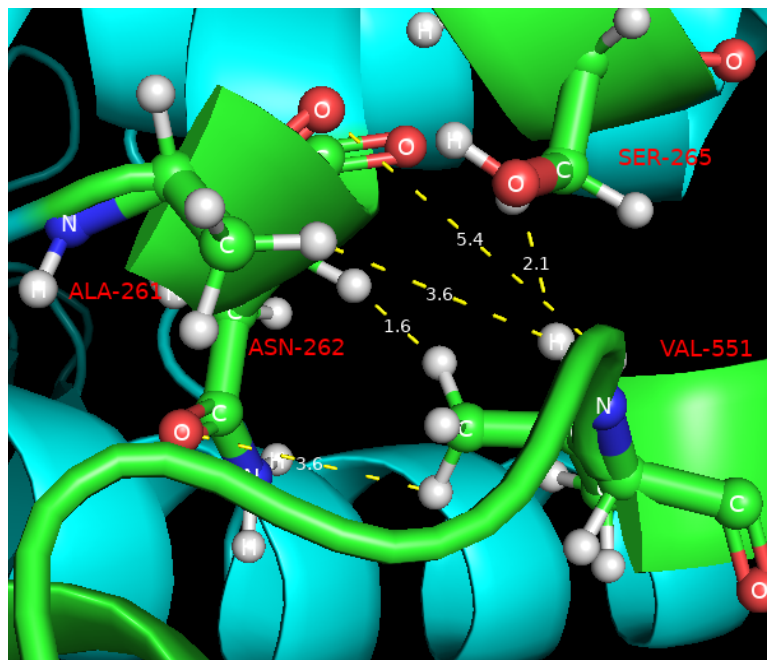


Figure 27 contacts of VAL-551 of mEVC to ALA 261, ASN 262, SER 265 of mEVC2 in Model 000

VAL-551 of mEVC shows a interaction with ALA 261, ASN 262, SER 265 of mEVC2. The hydrogen atom of VAL-551 of mEVC has a probable hydrogen bond with the oxygen atom of SER-265 and a weak electrostatic force with ASN-262 of mEVC2. The minimum distance is 2.1 Å and 3.6 Å respectively. VAL-551 also shows a probable Van der Waals force with ALA 261.

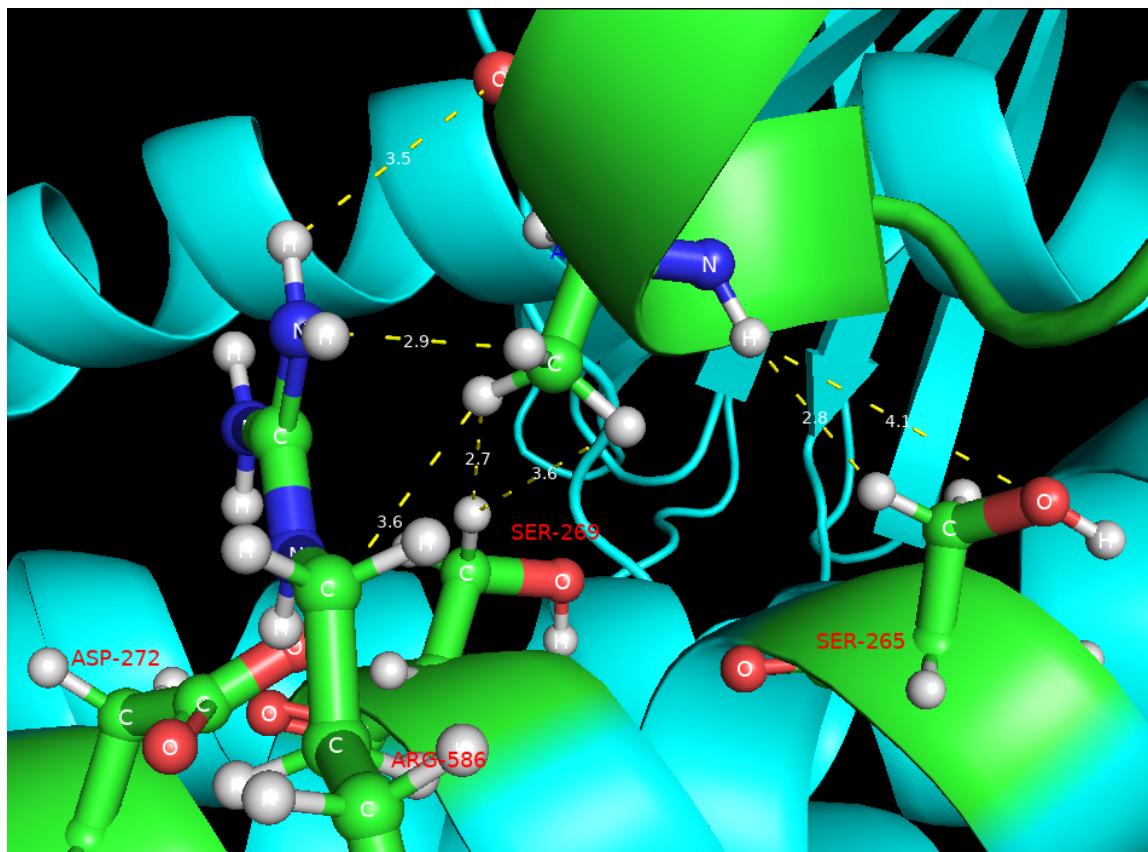


Figure 28 Contacts of ALA-552 of mEVC to SER 265, SER 269, ASP 272, ARG 586 of mEVC2 in Model 000

ALA-552 of mEVC shows contact with SER 265, SER 269, ASP 272, ARG 586 of mEVC2. The oxygen atom of ASP-272 of mEVC2 has a probable electrostatic force with the hydrogen atom of ALA-552 of mEVC. The minimum distance is 3.6 Å. The oxygen atom of ALA-552 of mEVC has a probable electrostatic force with the hydrogen atom of ASP-272 of mEVC2. The distance is

3.5 Å. Also, ALA-552 shows probable Van der Waals forces with SER-269 and SER-265.

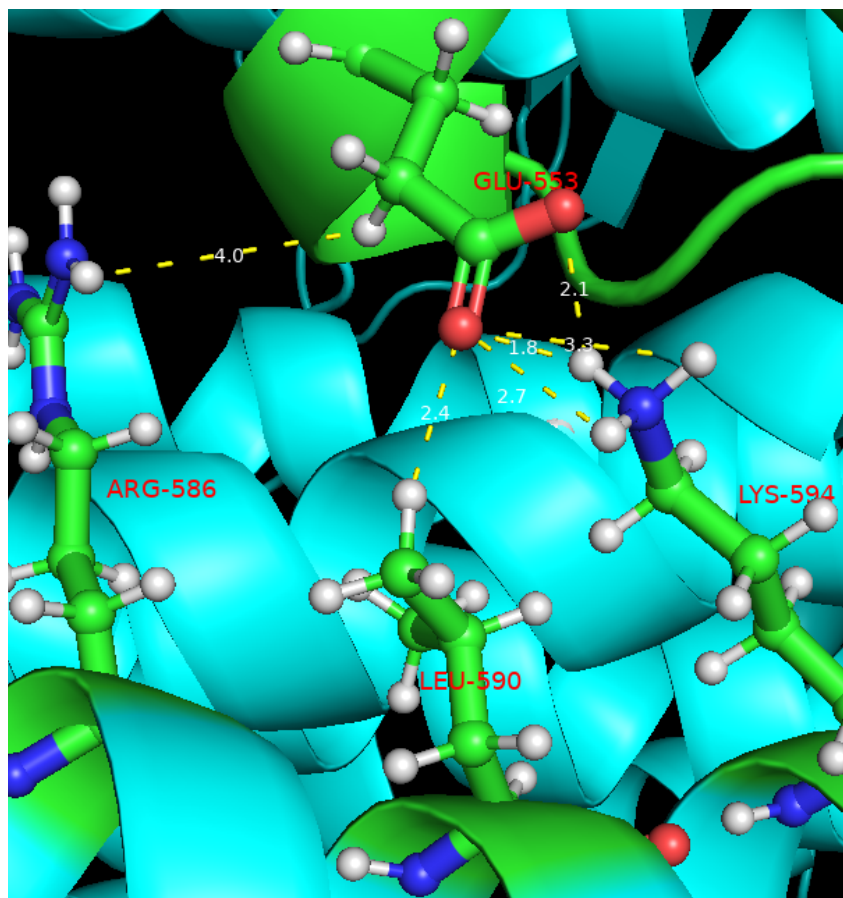


Figure 29 Contacts of GLU-553 of mEVC to ARG 586, LEU 590, LYS 594 of mEVC2 in model 000.

GLU-553 of mEVC shows contact with ARG 586, LEU 590, LYS 594 of mEVC2. The oxygen atom of GLU-553 shows a probable strong hydrogen bond with hydrogen atoms of LYS-594 and LEU-590. The minimum distance is 2.1 Å and 2.4 Å respectively. GLU-553 shows a probable Van der Waals force with ARG-586.

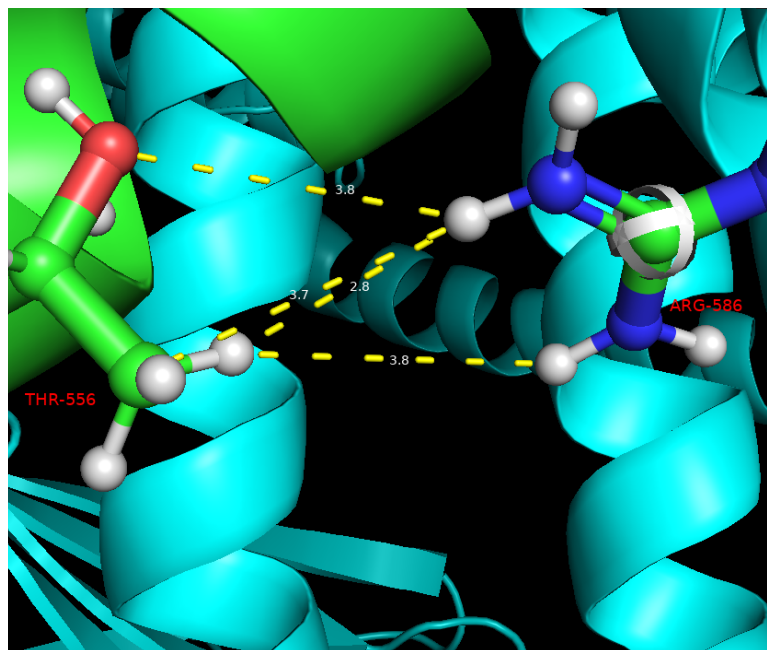


Figure 30 Contacts of THR-556 of mEVC to ARG 586 of mEVC2 in model 000.

THR-556 of mEVC shows probable contact with ARG 586 of mEVC2. The oxygen atom of THR-556 mEVC shows a probable weak hydrogen bond with hydrogen atom of ARG-586 of mEVC2. The minimum distance is 3.8 Å.

3.5 Mutation

To investigate the impact on the interaction between mEVC and mEVC2, a mutation was introduced in mEVC. Specifically, the amino acid GLU-545 was substituted with ALA, as it was observed that GLU-545 exhibited the highest number of interactions with mEVC2. Additionally, ALA was chosen for the mutation due to its relatively smaller volume compared to GLU. This smaller size of ALA may facilitate alterations in the contacts dynamics.

Table 7 Interaction between full length muted (GLU 545 to ALA) mouse EVC and full-length mouse EVC2 docked by ClusPro

mEVC (M)	mEVC2
LEU-540	ASN-262
LYS-541	ASN-262, GLN-354, GLU-358
PRO-544	ALA-261, ASN-262
ALA-545	PRO-255, SER-260, ASN-262
<i>(GLU-545)</i>	<i>PRO-255, SER-256, PRO-257, SER-260, ASN-262)</i>
VAL-546	PRO-255
SER-548	ALA-261, GLN-597
LEU-549	ALA-261, SER-265, LYS-594
PRO-550	ALA-261, SER-265, LEU-268, LEU-590, LYS-594
VAL-551	ALA-261, ASN-262, SER-265, SER-269
ALA-552	SER-265, SER-269, ASP-272, ARG-586
GLU-553	ARG-586, LEU-590, LYS-594
THR-556	ARG-586

According to the results obtained from PRODIGY, the analysis indicates changes in the interaction between two interfaces. Prior to the mutation, the amino acid GLU-545 of mEVC was found to interact with five amino acids of mEVC2:

PRO-255, SER-256, PRO-257, SER-260, and ASN-262. However, after the mutation, when ALA was introduced, it formed interactions with only three amino acids: PRO-255, SER-260, and ASN-262. ALA did not form any bonds with SER-256 and PRO-257.

Table 8 K_d value and ΔG when muted mouse EVC docked with full length mEVC2 by ClusPro

Protein-protein complex	ΔG (kcal mol ⁻¹)	K_d (M) at °C
complex.1	-22.0	6.9e-17

3.6 Docking by ZDock

Like ClusPro, the docking program ZDock was employed to perform the docking of the proteins in a similar manner.

3.6.1 Interaction between the Fragment mEVC and the Full Length mEVC2

Table 9 Compare of K_d value and ΔG when fragment mEVC docked with full length mEVC2 by ZDock

Protein-protein complex	ΔG (kcal mol ⁻¹)	K_d (M) at °C
complex.1	-13.6	1e-10

Upon analyzing the interaction between the fragment mEVC (451-750) and fragment mEVC2 (247-661), in which we specifically focused on the coiled-coil

region, it was found that there is no interface interaction between the two fragments.

3.6.2 Interactions between the Full length mEVC and the Fragment mEVC2

Table 10 List of contacts between full length mEVC and fragment mEVC2

mEVC	mEVC2
HIS-611	GLU-659
ALA-727	GLN-529
GLN-730	THR-658, ALA-526GLU-659, LEU-661
GLU-731	GLN-529, LEU-533LEU-530,
ARG-734	ALA-527, ALA-526LEU-533, THR-658, LEU-661, LEU-530, GLU-660, GLN- 529
LEU-735	HIS-537
GLN-737	LEU-661, ARG-554GLU-660, GLU-659
LEU-738	LEU-551, GLU-558ARG-554,
HIS-739	ARG-554
MET-740	ARG-554
GLU-741	ARG-554, GLU-558,
ARG-742	ARG-554

Table 11 Compare of K_d value and ΔG when full length mEVC docked with fragment mEVC2 by ZDock

Protein-protein complex	ΔG (kcal mol ⁻¹)	K_d (M) at °C
complex.1	-20.2	1.6e-15

3.6.3 Interaction between the Full length mEVC and the Full Length mEVC2

Table 12 Compare of K_d value and ΔG of first 10 complex when full length mEVC docked with full length mEVC2 by ZDock

Protein-protein complex	ΔG (kcal mol ⁻¹)	K_d (M) at °C
complex.1	-25.5	2.1e-19
complex.2	-19.6	4.3e-15
complex.3	-14.1	4.8e-11
complex.4	-18.9	1.3e-14
complex.5	-13.8	7.4e-11
complex.6	-22.4	3.7e-17
complex.7	-14.0	5.0e-11
complex.8	-14.0	5.7e-11
complex.9	-16.0	1.8e-12
complex.10	-15.7	2.8e-12

The analysis revealed that the value of ΔG and K_d is random for different models. The analysis also indicated that there were no observed interface contacts between the mEVC, amino acid residue 451 to 750, and the mEVC2, amino acid residue 247 to 661.

3.6.4 Contacts between the Fragment mEVC and the Fragment mEVC2

Table 13 Compare of K_d value and ΔG when fragment mEVC docked with fragment mEVC2 by ZDock

Protein-protein complex	ΔG (kcal mol ⁻¹)	K_d (M) at °C
complex.1	-14.1	4.9e-11

Table 14 List of contacts between fragment of mouse EVC and fragment of mouse EVC2 (ZDocked)

mEVC	mEVC2
GLN-452	THR-365, ASP-369
ARG-454	ARG-373
GLN-455	ALA-366, THR-365, ARG-373, ASP-369
ALA-456	ARG-373
ALA-458	ARG-373
ALA-459	ARG-373
THR-463	GLU-377
GLN-466	GLU-377, TYR-380
GLU-467	TYR-380
ARG-470	GLU-377, MET-384, GLN-381, TYR-380
ARG-471	TYR-380
LEU-474	MET-384, GLU-388
PHE-486	GLU-388
PHE-490	MET-384, GLU-388, GLN-381, VAL-385
LEU-494	GLN-381, VAL-385
GLU-495	GLN-381
GLN-497	GLN-381, GLU-377
ARG-498	GLN-381, GLU-377, ASN-378
ARG-501	GLN-381, ARG-373, ASN-378, THR-376, LYS-374, LEU-370, GLU-377

mEVC	mEVC2
SER-502	LYS-374, LEU-370
GLN-504	LEU-370, ARG-373
GLU-505	MET-372, ARG-373, LYS-374, GLU-371, ASP-369, LEU-370
GLY-506	LEU-370
GLU-508	ARG-373, ALA-366
ASP-509	LEU-370, ALA-366, GLU-367
ILE-512	THR-365, LEU-362, GLU-367, ALA-363, ALA-366
THR-513	ALA-363
ALA-515	LEU-362
MET-516	LYS-360, LEU-364, ALA-363, GLU-356, MET-361, ARG-359, TYR-357, GLU-358, LEU-362
ALA-517	ARG-359
LEU-519	GLU-358, LEU-362, GLU-355
CYS-520	ARG-359, GLU-356, GLU-358, GLU-355
GLN-521	GLU-355
GLU-522	GLU-355
LEU-523	GLU-358, ASN-351, GLU-355, GLN-354
TYR-524	GLU-355
CYS-525	GLU-355
SER-526	GLU-355
THR-527	ASN-351, GLU-355, LEU-348, LEU-347
MET-528	LEU-348, ASN-351
THR-530	ASN-351
PHE-531	LEU-348, GLN-345, LYS-344, PHE-346, LEU-347, ASN-351, LYS-343, LEU-349
GLN-532	LYS-344
PHE-534	LEU-347
VAL-535	LYS-344
LEU-538	ALA-340
PHE-539	LYS-337, ALA-340
LEU-543	GLN-333, LYS-337
THR-547	GLN-333, ARG-336
LEU-549	GLN-333, LYS-337
LEU-557	GLU-463, LYS-337, LYS-461
GLN-560	LYS-461, GLU-463, SER-457
VAL-561	ALA-341, LYS-344, LYS-461, GLN-345

mEVC	mEVC2
GLN-562	LYS-344
GLN-564	ILE-450, THR-453, GLN-345, ALA-458, LYS-461, SER-457, GLN-454, ILE-455
ALA-565	LEU-348, GLN-454, GLN-345
ARG-567	THR-453
GLN-568	SER-449, GLN-345, ILE-450, GLN-454, TYR-452, GLU-446, THR-453
LEU-569	GLU-352
GLN-571	THR-453, GLN-456, SER-449
ALA-572	PHE-442, GLU-446, GLU-352
ASP-573	GLU-352
PHE-575	GLU-446, PHE-442, ASN-445
ARG-576	GLU-352, GLU-356, GLU-446, PHE-442
GLN-579	ASN-445, PHE-442
ARG-647	VAL-385, GLU-389
LEU-651	ARG-382
ARG-669	GLN-438, GLN-434, ARG-437
GLN-672	GLN-438
LEU-680	ASN-445, GLU-446
VAL-684	TYR-452
SER-685	TYR-452, GLN-456
GLU-689	LYS-470

CHAPTER FOUR: DISCUSSIONS

Protein structure prediction plays a vital role in comprehending protein functions and devising potential therapeutic strategies. The advent of accurate protein structure prediction algorithms, such as AlphaFold 2.0, offers a promising alternative to obtain structural information, enabling valuable insights into protein functions and facilitating the discovery of novel therapeutics. The SWISS Model and the Phyre2, were tried but both only provided partial structure. In this study, protein structure was predicted and investigated the interaction between protein EVC and EVC2. As coiled-coil regions 3 and 4 of mEVC were considered for this study, fragment of mEVC, amino acid residue number 451 to 750 was also considered for the analysis. Similarly fragment of mEVC2, residue 247 to 661 was considered for the analysis, because those portions actively contact the coiled-coil region of mEVC2 in the experiment of Blair et al⁴⁰.

In the fragment region, 11 amino acid sequences of mEVC showed probable contacts with mEVC2, when the docking was done with full length both proteins. They are LEU-540, LYS-541, PRO-544, GLU-545, SER-548, LEU-549, PRO-550, VAL-551, ALA-552, GLU-553, and THR-556. They showed a probable Van der Waals force, and electrostatic interaction. Some of them were a hydrogen bond. No interaction was found in the P-loop and Leucine zipper region.

It is inferred that leucine zipper makes hydrophobic interaction by themselves.

The analysis of interfaces revealed that a significant portion of the amino acid interactions identified in PRODIGY were attributed to Van der Waals forces. These contacts were characterized by bond distances that exceeded those typically observed for covalent, ionic, or hydrogen bonds. While some contacts did involve hydrogen bonding, they lacked consistency across different models. It is possible that there are hydrogen bonds and ionic bonds facilitated by water molecules, but they were not observed due to the omission of water in the binding simulations performed by ClusPro and Pymol. Furthermore, many of the contacts appeared to exhibit hydrophobic characteristics.

To investigate the impact of the mEVC mutation on its interaction with mEVC2, a specific mutation was introduced. The amino acid GLU-545 in mEVC, which exhibited the highest number of contacts with mEVC2, was substituted with ALA. ALA was chosen for the mutation due to its relatively smaller volume compared to GLU, which could potentially affect the interaction dynamics. The analysis performed using PRODIGY revealed changes in the contacts between the two interfaces. Prior to the mutation, GLU-545 of mEVC was observed to contact with five amino acids of mEVC2, PRO-255, SER-256, PRO-257, SER-260, and ASN-262. However, after the mutation was introduced, with ALA replacing GLU-545, only three amino acids showed contacts, PRO-255, SER-260, and ASN-262. ALA did not form any bonds with SER-256 and PRO-257.

To compare with the result of ClusPro, ZDock was used to dock mEVC with mEVC2. mEVC sequence was renamed as chain “A” and mEVC2 as chain “C” to dock. The analysis of the interaction between the mEVC fragment (residue 451 to 750) and the mEVC2 fragment (247 to 661), with a particular focus on the coiled-coil region, indicated the absence of a interface interaction between the two fragments. Specifically, the analysis revealed no observable interface interaction between the mEVC sequence spanning from amino acid 451 to amino acid 750 and the mEVC2 sequence spanning from amino acid number 247 to amino acid number 661.

CHAPTER FIVE: CONCLUSIONS

A protein structure prediction is a valuable tool for understanding protein functions and designing potential therapeutics. In this study, protein's structure were predicted and interactions between the coiled-coil regions 3 and 4 of mEVC protein with mEVC2 protein were investigated.

Through docking simulations and interface analysis, we identified 11 amino acids in the mEVC fragment that likely contact mEVC2. These interactions involved Van der Waals forces, electrostatic interactions, and hydrogen bonds. Interestingly, we observed no contacts in the P-loop and Leucine zipper regions, suggesting that the leucine zipper region might engage in hydrophobic interactions internally.

Furthermore, by using PRODIGY, we detected a change of contacts in between the mEVC and mEVC2 interface after introducing a mutation. To validate the result of ClusPro, the ZDock algorithm for docking mEVC with mEVC2 was used. However, no observable interface interaction was found in the coiled-coil region 3 and 4. Overall, our findings provide insights into the specific amino acids involved in the interaction between mEVC and mEVC2, underscoring the significance of certain residues and the impact of mutations on interaction dynamics.

It is worth noting that achieving consistent interaction interfaces across different models proved challenging, potentially due to limitations in generating repeated or sufficient models using the Alphafold, ClusPro and PRODIGY method.

REFERENCES

- (1) Das, D.; Das, G.; Mahapatra, TK. S.; Biswas, J. Ellis van Creveld Syndrome with Unusual Association of Essential Infantile Esotropia. *Oman J Ophthalmol* **2010**, 3 (1), 23. <https://doi.org/10.4103/0974-620x.60017>.
- (2) McKusick, V. A. Ellis-van Creveld Syndrome and the Amish. *Nat Genet* **2000**, 24 (3), 203–204. <https://doi.org/10.1038/73389>.
- (3) Ruiz-Perez, V. L.; Goodship, J. A. Ellis-van Creveld Syndrome and Weyers Acroental Dysostosis Are Caused by Cilia-Mediated Diminished Response to Hedgehog Ligands. *Am J Med Genet C Semin Med Genet* **2009**, 151 (4), 341–351. <https://doi.org/10.1002/ajmg.c.30226>.
- (4) Kamal, R.; Dahiya, P.; Kaur, S.; Bhardwaj, R.; Chaudhary, K. Ellis-van Creveld Syndrome: A Rare Clinical Entity. *Journal of Oral and Maxillofacial Pathology* **2013**, 17 (1), 132–135. <https://doi.org/10.4103/0973-029X.110716>.
- (5) Baujat, G.; Le Merrer, M. Ellis-van Creveld Syndrome. *Orphanet J Rare Dis* **2007**, 2 (1), 1–5. <https://doi.org/10.1186/1750-1172-2-27>.
- (6) Shi, L.; Luo, C.; Ahmed, M. K.; Attaie, A. B.; Ye, X. Novel Mutations in EVC Cause Aberrant Splicing in Ellis-van Creveld Syndrome. *Molecular Genetics and Genomics* **2016**, 291 (2), 863–872. <https://doi.org/10.1007/s00438-015-1151-2>.

- (7) Ruiz-Perez, V. L.; Tompson, S. W. J.; Blair, H. J.; Espinoza-Valdez, C.; Lapunzina, P.; Silva, E. O.; Hamel, B.; Gibbs, J. L.; Young, I. D.; Wright, M. J.; Goodship, J. A. Mutations in Two Nonhomologous Genes in a Head-to-Head Configuration Cause Ellis-van Creveld Syndrome. *Am J Hum Genet* **2003**, 72 (3), 728–732. <https://doi.org/10.1086/368063>.
- (8) Sund, K. L.; Roelker, S.; Ramachandran, V.; Durbin, L.; Benson, D. W. Analysis of Ellis van Creveld Syndrome Gene Products: Implications for Cardiovascular Development and Disease. *Hum Mol Genet* **2009**, 18 (10), 1813–1824. <https://doi.org/10.1093/hmg/ddp098>.
- (9) Blair, H. J.; Tompson, S.; Liu, Y. N.; Campbell, J.; MacArthur, K.; Ponting, C. P.; Ruiz-Perez, V. L.; Goodship, J. A. Evc2 Is a Positive Modulator of Hedgehog Signalling That Interacts with Evc at the Cilia Membrane and Is Also Found in the Nucleus. *BMC Biol* **2011**, 9, 1–13. <https://doi.org/10.1186/1741-7007-9-14>.
- (10) Ruiz-Perez, V. L.; Ide, S. E.; Strom, T. M.; Lorenz, B.; Wilson, D.; Woods, K.; King, L.; Francomano, C.; Freisinger, P.; Spranger, S.; Marino, B.; Dallapiccola, B.; Wright, M.; Meitinger, T.; Polymeropoulos, M. H.; Goodship, J. Mutations in a New Gene in Ellis-van Creveld Syndrome and Weyers Acroental Dysostosis. *Nat Genet* **2000**, 24 (3), 283–286. <https://doi.org/10.1038/73508>.

- (11) García De Viedma, D.; Giraldo, R.; Rivas, G.; Fernández-Tresguerres, E.; Díaz-Orejas, R. A Leucine Zipper Motif Determines Different Functions in a DNA Replication Protein. *EMBO Journal* **1996**, *15* (4), 925–934. <https://doi.org/10.1002/j.1460-2075.1996.tb00427.x>.
- (12) Miller, M.; Shuman, J. D.; Sebastian, T.; Dauter, Z.; Johnson, P. F. Structural Basis for DNA Recognition by the Basic Region Leucine Zipper Transcription Factor CCAAT/Enhancer-Binding Protein α . *Journal of Biological Chemistry* **2003**, *278* (17), 15178–15184. <https://doi.org/10.1074/jbc.M300417200>.
- (13) Abdullah, N.; Balakumari, M.; Sau, A. K. Dimerization and Its Role in GMP Formation by Human Guanylate Binding Proteins. *Biophys J* **2010**, *99* (7), 2235–2244. <https://doi.org/10.1016/j.bpj.2010.07.025>.
- (14) Saraste, M.; Sibbald, P. R.; Wittinghofer, A. The P-Loop - a Common Motif in ATP- and GTP-Binding Proteins. *Trends Biochem Sci* **1990**, *15* (11), 430–434. [https://doi.org/10.1016/0968-0004\(90\)90281-F](https://doi.org/10.1016/0968-0004(90)90281-F).
- (15) Liu, S.; Dontu, G.; Wicha, M. S. Mammary Stem Cells, Self-Renewal Pathways, and Carcinogenesis. *Breast Cancer Research* **2005**, *7* (3), 86–95. <https://doi.org/10.1186/bcr1021>.
- (16) Yang, L.; Xie, G.; Fan, Q.; Xie, J. Activation of the Hedgehog-Signaling Pathway in Human Cancer and the Clinical Implications. *Oncogene* **2010**, *29* (4), 469–481. <https://doi.org/10.1038/onc.2009.392>.

- (17) Ruiz-Perez, V. L.; Blair, H. J.; Rodrigues-Andres, M. E.; Blanco, M. J.; Wilson, A.; Liu, Y. N.; Miles, C.; Peters, H.; Goodship, J. A. Evc Is a Positive Mediator of Ihh-Regulated Bone Growth That Localises at the Base of Chondrocyte Cilia. *Development* **2007**, *134* (16), 2903–2912. <https://doi.org/10.1242/dev.007542>.
- (18) Dorn, K. V.; Hughes, C. E.; Rohatgi, R. A Smoothened-Evc2 Complex Transduces the Hedgehog Signal at Primary Cilia. *Dev Cell* **2012**, *23* (4), 823–835. <https://doi.org/10.1016/j.devcel.2012.07.004>.
- (19) O'Connor, M. J.; Collins, R. T. Ellis-van Creveld Syndrome and Congenital Heart Defects: Presentation of an Additional 32 Cases. *Pediatr Cardiol* **2012**, *33* (4), 491–492. <https://doi.org/10.1007/s00246-012-0155-5>.
- (20) Louie, K. W.; Mishina, Y.; Zhang, H. Molecular and Cellular Pathogenesis of Ellis-van Creveld Syndrome: Lessons from Targeted and Natural Mutations in Animal Models. *Journal of Developmental Biology*. 2020, pp 1–15. <https://doi.org/10.3390/jdb8040025>.
- (21) Caparrós-Martín, J. A.; Valencia, M.; Reytor, E.; Pacheco, M.; Fernandez, M.; Perez-Aytes, A.; Gean, E.; Lapunzina, P.; Peters, H.; Goodship, J. A.; Ruiz-Perez, V. L. The Ciliary EVC/EVC2 Complex Interacts with Smo and Controls Hedgehog Pathway Activity in Chondrocytes by Regulating Sufu/Gli3 Dissociation and Gli3 Trafficking in Primary Cilia. *Hum Mol Genet* **2013**, *22* (1), 124–139. <https://doi.org/10.1093/hmg/dd5409>.

- (22) Takahashi, R.; Yamagishi, M.; Nakano, K.; Yamochi, T.; Yamochi, T.; Fujikawa, D.; Nakashima, M.; Tanaka, Y.; Uchimaru, K.; Utsunomiya, A.; Watanabe, T. Epigenetic Deregulation of Ellis Van Creveld Confers Robust Hedgehog Signaling in Adult T-Cell Leukemia. *Cancer Sci* **2014**, *105* (9), 1160–1169. <https://doi.org/10.1111/cas.12480>.
- (23) Rubin, L. L.; de Sauvage, F. J. Targeting the Hedgehog Pathway in Cancer. *Nat Rev Drug Discov* **2006**, *5* (12), 1026–1033. <https://doi.org/10.1038/nrd2086>.
- (24) Nooren, I. M. A.; Thornton, J. M. Diversity of Protein-Protein Interactions. *EMBO Journal*. July 15, 2003, pp 3486–3492. <https://doi.org/10.1093/emboj/cdg359>.
- (25) Miller, M.; Shuman, J. D.; Sebastian, T.; Dauter, Z.; Johnson, P. F. Structural Basis for DNA Recognition by the Basic Region Leucine Zipper Transcription Factor CCAAT/Enhancer-Binding Protein α . *Journal of Biological Chemistry* **2003**, *278* (17), 15178–15184. <https://doi.org/10.1074/jbc.M300417200>.
- (26) Peng, J.; Xu, J. Raptorx: Exploiting Structure Information for Protein Alignment by Statistical Inference. *Proteins: Structure, Function and Bioinformatics* **2011**, *79* (SUPPL. 10), 161–171. <https://doi.org/10.1002/prot.23175>.
- (27) Bissan Al-Lazikani, J. J. Z. X. and B. H. Protein Structure Prediction.

- (28) Varadi, M.; Anyango, S.; Deshpande, M.; Nair, S.; Natassia, C.; Yordanova, G.; Yuan, D.; Stroe, O.; Wood, G.; Laydon, A.; Židek, A.; Green, T.; Tunyasuvunakool, K.; Petersen, S.; Jumper, J.; Clancy, E.; Green, R.; Vora, A.; Lutfi, M.; Figurnov, M.; Cowie, A.; Hobbs, N.; Kohli, P.; Kleywegt, G.; Birney, E.; Hassabis, D.; Velankar, S. AlphaFold Protein Structure Database: Massively Expanding the Structural Coverage of Protein-Sequence Space with High-Accuracy Models. *Nucleic Acids Res* **2022**, *50* (D1), D439–D444. <https://doi.org/10.1093/nar/gkab1061>.
- (29) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>.
- (30) Peng, J.; Xu, J. A Multiple-Template Approach to Protein Threading. *Proteins: Structure, Function and Bioinformatics* **2011**, *79* (6), 1930–1939. <https://doi.org/10.1002/prot.23016>.

- (31) Ginalski, K. Comparative Modeling for Protein Structure Prediction. *Current Opinion in Structural Biology*. April 2006, pp 172–177. <https://doi.org/10.1016/j.sbi.2006.02.003>.
- (32) Kozakov, D.; Beglov, D.; Bohnuud, T.; Mottarella, S. E.; Xia, B.; Hall, D. R.; Vajda, S. How Good Is Automated Protein Docking? *Proteins: Structure, Function and Bioinformatics* **2013**, 81 (12), 2159–2166. <https://doi.org/10.1002/prot.24403>.
- (33) Ghani, U.; Desta, I.; Jindal, A.; Khan, O.; Jones, G.; Hashemi, N.; Kotelnikov, S.; Padhorny, D.; Vajda, S.; Kozakov, D. Improved Docking of Protein Models by a Combination of AlphaFold2 and ClusPro. <https://doi.org/10.1101/2021.09.07.459290>.
- (34) Kozakov, D.; Hall, D. R.; Xia, B.; Porter, K. A.; Padhorny, D.; Yueh, C.; Beglov, D.; Vajda, S. The ClusPro Web Server for Protein-Protein Docking. *Nat Protoc* **2017**, 12 (2), 255–278. <https://doi.org/10.1038/nprot.2016.169>.
- (35) Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. ClusPro: A Fully Automated Algorithm for Protein-Protein Docking. *Nucleic Acids Res* **2004**, 32 (WEB SERVER ISS.). <https://doi.org/10.1093/nar/gkh354>.
- (36) Källberg, M.; Wang, H.; Wang, S.; Peng, J.; Wang, Z.; Lu, H.; Xu, J. Template-Based Protein Structure Modeling Using the RaptorX Web Server. *Nat Protoc* **2012**, 7 (8), 1511–1522. <https://doi.org/10.1038/nprot.2012.085>.

- (37) Wiehe, K.; Peterson, M. W.; Pierce, B.; Mintseris, J.; Weng, Z. *Protein-Protein Docking: Overview and Performance Analysis*.
- (38) Chen, R.; Li, L.; Weng, Z. *ZDOCK: An Initial-Stage Protein-Docking Algorithm*; 2003. <http://zlab.bu.edu/>.
- (39) Pierce, B. G.; Hourai, Y.; Weng, Z. Accelerating Protein Docking in ZDOCK Using an Advanced 3D Convolution Library. *PLoS One* **2011**, 6 (9). <https://doi.org/10.1371/journal.pone.0024657>.
- (40) Blair, H. J.; Tompson, S.; Liu, Y. N.; Campbell, J.; MacArthur, K.; Ponting, C. P.; Ruiz-Perez, V. L.; Goodship, J. A. Evc2 Is a Positive Modulator of Hedgehog Signalling That Interacts with Evc at the Cilia Membrane and Is Also Found in the Nucleus. *BMC Biol* **2011**, 9. <https://doi.org/10.1186/1741-7007-9-14>.

APPENDIX

List of Programs and Websites

AlphaFold	AlphaFold, developed by DeepMind, is an artificial intelligence system that utilizes amino acid sequences to predict the 3D structure of proteins. Its accuracy is often comparable to experimental results. The AlphaFold webserver, available at https://alphafold.ebi.ac.uk/ , allows users to access this system and generate protein structure predictions based on their input sequences.
ClusPro	ClusPro is a webserver for protein docking, developed and maintained by the Vajda Lab and ABC Group at Boston University and Stony Brook University. It allows users to dock protein structures or amino acid sequences and generate a series of models based on various conditions. You can access it at https://cluspro.bu.edu .
PRODIGY	PRODIGY is a webserver for protein binding energy prediction, created by the Computational Structural Biology group/NMR Research Group at Utrecht University. It analyzes

protein models and provides values such as K_d (dissociation constant) and ΔG (Gibbs free energy change), as well as information about interacting amino acids between protein chains. The PRODIGY webserver can be found at <https://bianca.science.uu.nl/prodigy>.

Pymol

Pymol is an open-source molecular visualization software that is maintained and distributed by Schrödinger. It allows users to open and edit PDB files, enabling tasks like renaming or mutating models. To access Pymol, visit <https://pymol.org>.

VITA

Md Atikul Islam Mamun was raised in Bagerhat, Bangladesh. He completed his high school education at Govt. Science College, Dhaka, in 2007. Subsequently, in the academic year of 2007-2008, he enrolled at the University of Dhaka, Bangladesh. In 2013, he proudly earned a Bachelor of Science in Applied Chemistry and Chemical Engineering. In the Fall of 2021, he embarked on a new journey at Stephen F. Austin State University, where he pursued his graduate studies under the expert guidance of Professor Dr. Odutayo Odunuga. Throughout this period, he also served as a Graduate Teaching Assistant. This endeavor culminated in his amazing achievement of obtaining a Master of Science in Natural & Applied Sciences (Concentration: Chemistry) in August 2023.

Permanent Address: West Dhuligati Primary School Road, Bagerhat-9322, Bangladesh.

Style manual designation: The ACS Chemical Biology Style Guide

This thesis was typed by Md Atikul Islam Mamun in 2023

