ORIGINAL RESEARCH

Life Science Journal of Pakistan
http://www.lifesciencejournal.pk

Computational and Functional Analysis of Hypodermin C Protein against GWFI

Fatima Tahir[1], Almeera Shoukat[1], Abubakar Ahmed[1], Amna Saleem[1], Hinza Asghar[1], Aqeel Ahmad*[1, 2]

[1*]Department of Biotechnology, Faculty of Life Sciences (FLS), University of Central Punjab, 5400 Lahore Pakistan.
[2] Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Pakistan

*CORRESPONDING AUTHOR
Aqeel Ahmad
Address: Department of Bioinformatics and Biotechnology, Government College University Faisalabad, Pakistan
aachaudhary142@gmail.com

ABSTRACT

GWFI (Goat Warble Fly Infestation) is the myiasis caused by the larvae dipterous fly *Przhevalskiana silenus* in goats. This fly colonizes under the skin of an animal and increases its size at each larval stage. The larval gut produces the peptide degrading protein called hypodermin C (HyC). This study suggests that how collagen and HyC interact with each other. FASTA sequences of protein were retrieved from NCBI and the aligned sequences showed the highest similarity with serine proteases. The structure of HyC was modeled using I-TASSER online server followed by its validations using Procheck verification tool and physicochemical properties were confirmed by AA-prop. The protein structure of HyC was docked against Collagen by using their PDB IDs, 1HYL and 1BKV respectively. The structure of HyC showed a more determined hydrophobic nature and both structures were also superimposed to find out similarities and differences between them. Glutamic acid, Aspartic acid, Serine, and Lysine are found as interacting residues that are involved in docking with collagen. Generated structure of HyC that was docked against the collagen protein residues as the HyC produced from the gut of the fly binds with the collagen of the animal body and degrades the collagen, the residues which taken part in the binding process could be blocked which ultimately inhibits the binding of the hypodermin C and collagen thus these residues to control the infection. This present study helps in vaccine development against hypodermis, through inhibition of the binding of the collagen with hypodermin C is the strategy towards vaccine production.

Keywords: Hypodermin C, Myiasis, *Przhevalskiana Silenus*, *Molecular Docking*, Physiochemical

Life Sci J Pak 2021; 3(01):16-26. doi:10
(Received 01 July 2020 – Accepted 23 November 2020 – Published March 31, 2021)

INTRODUCTION

Myiasis originates from the larvae of hypodermatinae subfamily (Diptera, Oestridae) and the disease is called warble fly infestation (WFI). The members of the Diptera family are specific for their host. *Hypoderma lineatum* and *Hypoderma bovis* are host-specific for cattle and buffalo. *Hypoderma diana* causes disease in roe deer. *Hypoderma actaeon* causes disease in red deer. *Hypoderma tarandi* is host-specific for reindeer. *Hypoderma sinense* causes disease in yaks and *Przhevalskiana silenus* is specific for goats in which it causes the disease called GWFI (goat warble fly infestation) (1). *Pzhevalskiana silenus* fly colonize under the skin, legs, and back of the goat. The fly causes massive economic loss by reducing body weight and growth, changes the quality of hides for tanning purposes, lowers milk and meat production, and causes hypersensitivity reactions in goats (2).

At different larval stages, larvae change their size and color. In the first larval stage L1, the size of the larva is 2-7mm, white which inhabits subcutaneous tissues. At the second larval stage L2, the colors remain the

same, but size increases to 8-12mm. At the third larval stage L3, larvae move to the superficial dermis, and color changes from brown to black with an increased size of 13-19mm. The incubation period completes in 3-8 days (3). The parasites produce a large range of serine proteases enzyme that degrades the host body tissue in myiasis (4). HyC is found in the species of hypoderma and many other parasites include Pzhevalskiana silenus. The disease originates by the penetration of larva in the skin of an animal by the activity of collagenase enzymes mainly HyC (5). The hypodermis is caused by the hypodermin C during the incubation period of the larvae. Hypodermin C degrades the subcutaneous tissue of the animal resulting in low performance of the animal by *P.silenus* affects body weight and the immune system of the animal and evokes hypersensitivity reactions in the body (6).

Docking is a computational approach that is used to study the properties and interactions We also studied how Hypodermins bind with collagen which may help devise strategies to develop a vaccine against hypodermosis (7). This study is to model first-time HyC and the generated structure that will be dock against the collagen protein. Molecular docking exhibits the new strategy to make the vaccine for the elimination of disease to overcome the economic loss. As the HyC produces from the gut of the fly binds with the collagen of the animal body and degrades the collagen, the residues which take part in the binding process could be blocked which ultimately inhibits the binding of the HyC and collagen. The inhibition of the binding of the collagen with HyC is the strategy towards the production of the vaccine for the disease.

MATERIALS AND METHODS

The disease-causing protein sequence of the HyC was retrieved from the NCBI (National Centre for Biotechnology Information). The FASTA format of the protein sequence was aligned for the knowledge of sequence similarity with the other proteins using the database BLAST (Basic Local Alignment Search Tool). BLAST showed the highest similarity percentage of HyC with the *Hypoderma lineatum* and serine proteases.

Protein Modeling and Verification:

The protein structure was modeled by I-Tasser and SWISS modeler to obtain the three-dimensional structure. The best three-dimensional structure was selected on the base of the highest C-Score. Hypodermin C and collagenase structure were verified by using Procheck verification tool. For the verification purpose, PDB ID of the collagenase and hypodermin C was retrieved from the RSCB (Protein Database). The PDB ID used was 1BKV and 1HYL for the collagen and hypodermin collagenase respectively.

Physio-chemical properties

The physicochemical properties of a protein are analyzed using the protein physicochemical properties prediction tool AA-Prop. The FASTA sequence of hypodermin C was retrieved from the NCBI. The amino acid sequence was pasted on the protein sequence bar. The results gave the number of total amino acids, molecular weight, atomic composition, kyte, and Doolittle hydrophobicity plot.

Stereochemical Analysis

Stereochemical analysis was performed by Procheck.

Molecular Docking

The protein structure of hypodermin C was docked against the collagen protein to pursue their binding sites. Protein-Protein docking is also a type of molecular docking. The ligand interaction with protein can be modified by changing its affinity is helpful to design the drugs or new molecular probes. One compound or ligand can be interacting with the protein in many different ways. Protein-protein interactions are necessary to study cellular functions. X-ray crystallography and NMR spectroscopy help us to understand the structure of a single protein and complex protein structure.

The main objective of the docking is to understand which residues of the protein are involved in the interaction of hypodermin C and collagen. For the docking purpose, the Cluspro database was used. For the submission of the job, PDB ID of protein was taken from the RSCB (protein database). The 1HYL PDB ID was used for the hypodermin C and 1BKV PDB ID was used for the collagen. (8).

Analysis of Protein Complex Structures:

The resulting models were analyzed by using the software Dimplot and Ligplot. This software gives information on bind residues between proteins. Different chains represent the protein-protein interaction between the residues.

RESULTS

Secondary Structure Prediction:

PDB structure of salmon trypsin, 1HYL was selected to a model of HyC, with maximum query coverage 95%, residue range covered was from 0-230 and 39.39% identity. The predicted 3D structure of HyC was visualized on Pumol and Jmol. The structures of hypodermin C proteins are predominantly composed of 11% helices, 38% beta-sheet as shown in Fig. 1.

Physio-chemical Analysis:

Various physio-chemical properties of HyC are predicted using AA Prop. It calculates many of the parameters (Table 1). Hypodermin has a sequence of 230 aa. The molecular weight of the protein is 28.57 kilodaltons. It also provides several other

parameters like, Aliphatic index 88.76%, Aromaticity (Y+W+F): 10.3%, Grand average of hydropathicity (GRAVY) 0.049 which indicates its more interaction with water molecules shown in Table I.

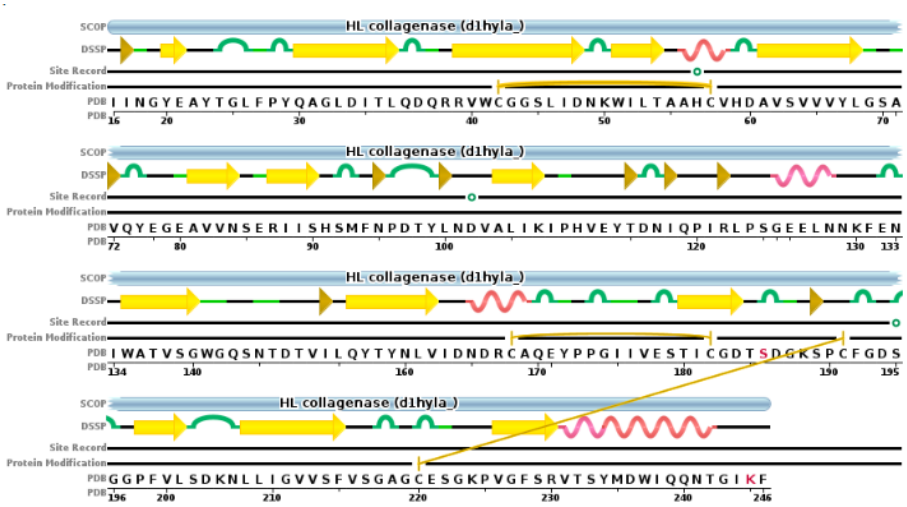

Fig. 1. Secondary structure prediction

Table 1: Composition of residues in Protein

Amino Acid composition	Percentage	Amino Acid composition	Percentage
Ala (A) 19	6.5%	Leu (L) 21	7.2%
Arg (R) 10	3.4%	Lys (K) 7	2.4%
Asn (N) 16	5.5%	Met (M) 6	2.1%
Asp (D) 17	5.8%	Phe (F) 12	4.1%
Cys (C) 7	2.4%	Pro (P) 13	4.5%
Gln (Q) 11	3.8%	Ser (S) 21	7.2%
Glu (E) 16	5.5%	Thr (T) 16	5.5%
Gly (G) 23	7.9%	Trp (W) 5	1.7%
His (H) 7	2.4%	Tyr (Y) 13	4.5%
Ile (I) 24	8.2%	Val (V) 22	7.6%
Leu (L) 21	7.2%	Pyl (O) 2	0.7%

Stereochemical Analysis

The first plot produced by PROCHECK is a Ramachandran plot based on an analysis of 118 structure of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be accepted to have over 90% in the most favored region shown in Fig. 2. The appearance of the plot itself can be modified to some extent by amending the program parameters. Thus, the shading and/or writing of different regions can be switched on or off, then region borders can be drawn in or not drawn in and the individual residues can be labeled as shown in fig.3. Total residues in protein 230 plot statics represented in most favored regions in A, B and L segments 88.5%, additional allowed regions a, b, 1, p 11.5% and residues in generous regions ~a, ~b, ~l, ~p 0.0%.

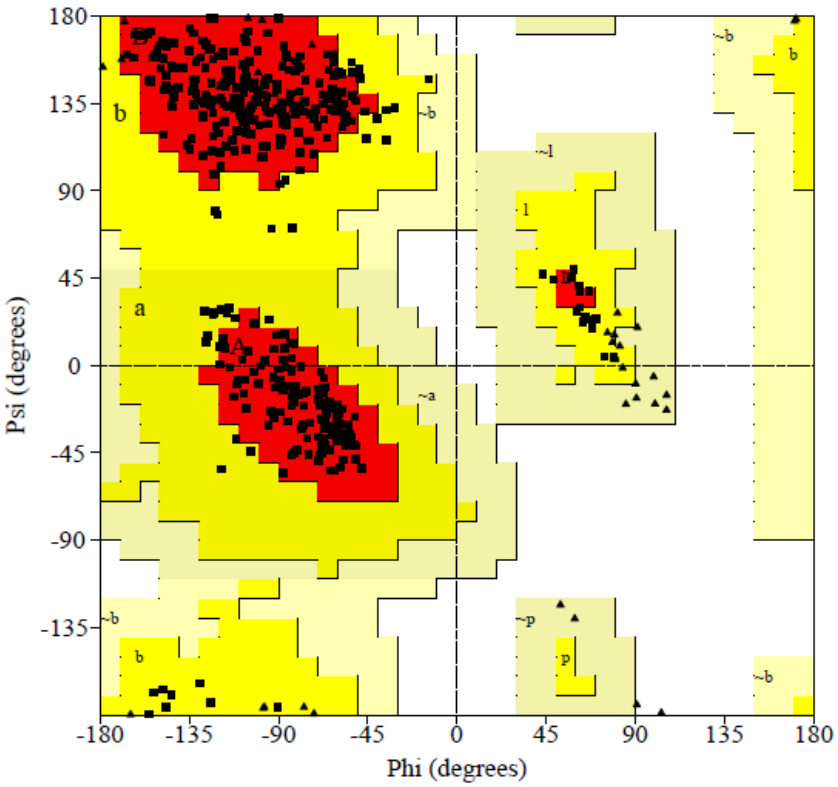

Fig. 2. Ramachandran Plot confirmation of Hypodermin C structure

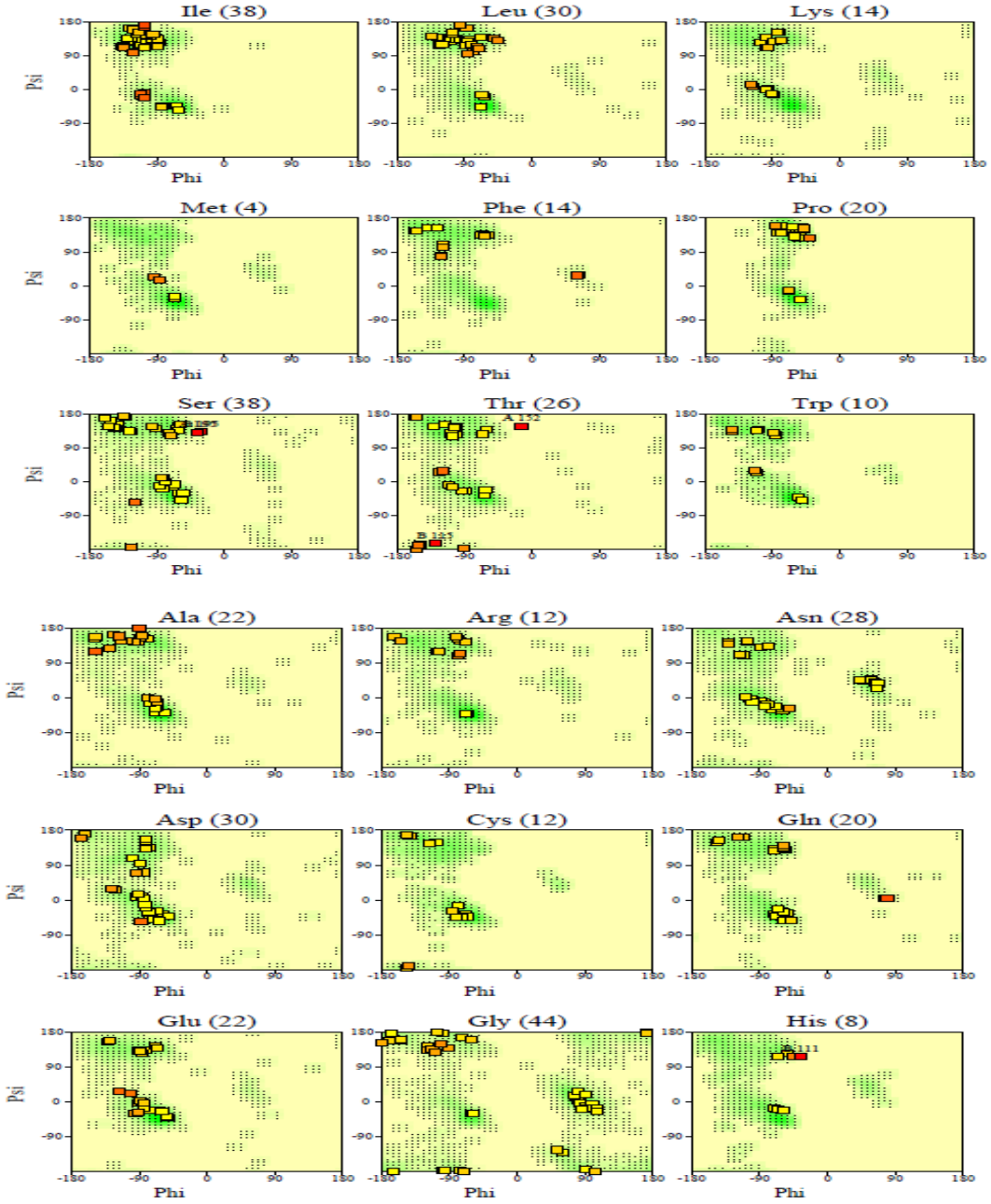

Fig. 3. Ramachandran plot for all residues.

Chi 1 and Chi 2

The numbers of residues are shown in brackets. Those in unfavorable conformations (score < -3.00) are labelled. Shading shows favorable conformations as obtained from an analysis of 163 structures at resolution 2.0A or better shown in Fig. 4. The plots showed a graph of the Chi1 *versus* the Chi2 torsion angles for each residue, where applicable. As each Chi1 and Chi2 can be in one of three preferred configurations *(gauche minus, trans* and *gauche plus)*, there are 3 x 3 =9 combinations for the two angles. The nine 'ideal' positions are marked by crosses on the graph.

The graphs showing the first 111 residues in the bar a, b, and c selected from five possible that contains the three shown with absolute deviation of main chain hydrogen bond energy from the ideal value, the beta value of the lemda atom (O, C or S, whichever is used

in the definition of the X l torsion angle); the average B value of main-chain atoms; and the average B value of side-chain atoms. The bottom part of the diagram illustrates information given in greater detail in the residue-by-residue listing shown in Fig. 5.

Main and Side-Chain Confirmation

The plot shows the graphs of six main-chain properties of the structure and how these properties compare with well-refined structures at a similar resolution. The dark band in each graph represents the results from the well-refined structures; the central line is a least-squares fit to the mean trend as a function of resolution, while the width of the band on either side of it corresponds to a variation of one standard deviation about the mean. In some cases, the tendency is dependent on the resolution, and in other cases is independent of it. (a) shows the Ramachandran-plot quality assessment, as measured by the percentage of the protein's residues that are in its most favored, or 'core' regions, (b) shows the peptide bond planarity measured by the standard deviation of the torsion angles; (c) shows the measure of bad non bonded interaction, (d) shows the alpha carbon tetrahedral distortion, measured by the standard deviation of the ~ 'torsion' angle, (e) showed the hydrogen bond energies and (f) showed the overall G-factor shown in Fig. 6.

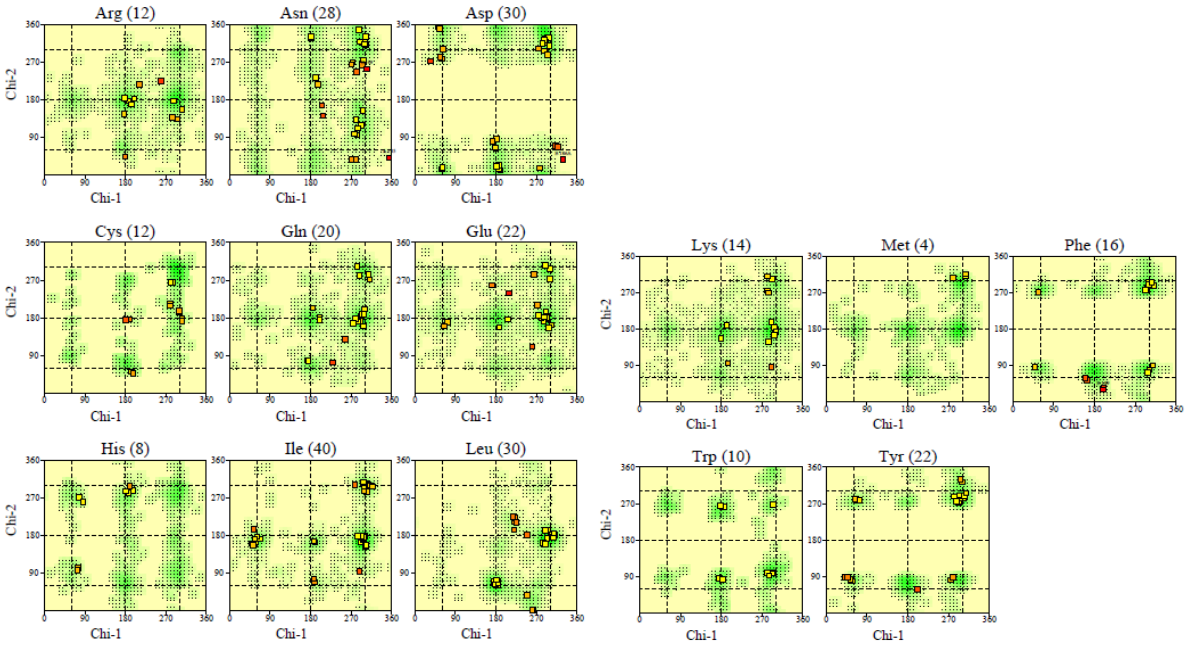

Fig. 4. Ramachandran plot for all residues.

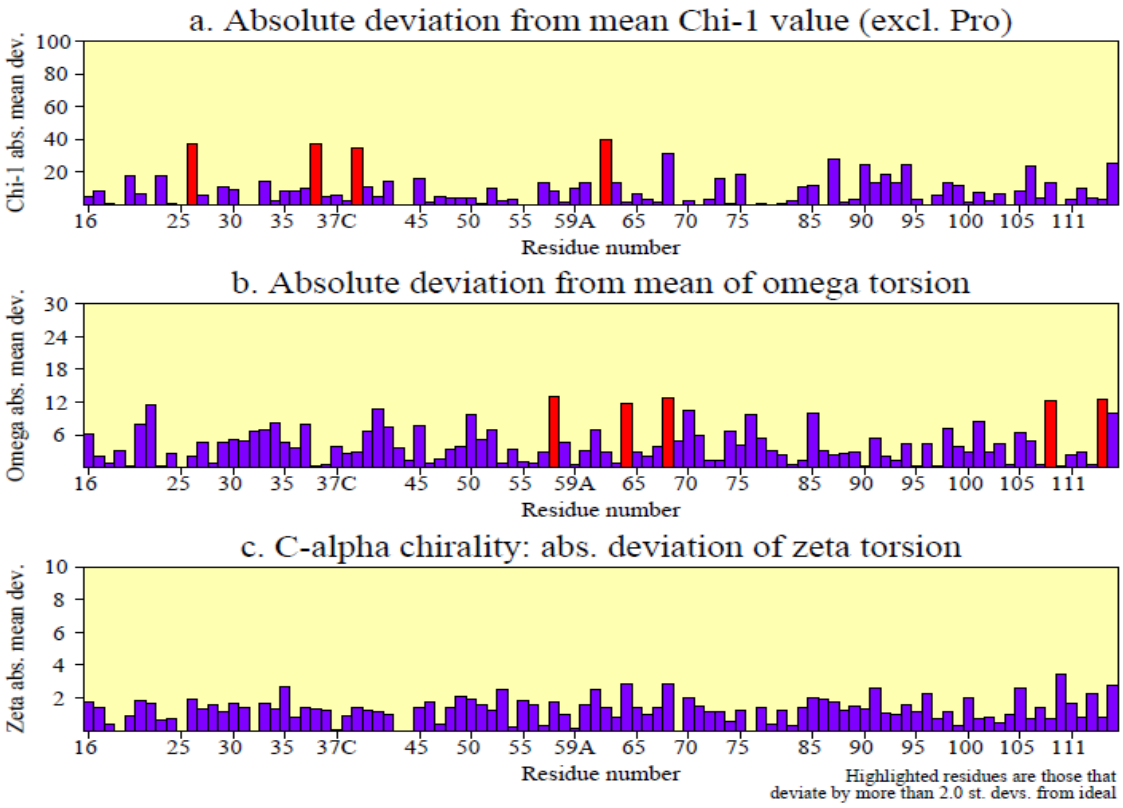

Fig. 5. Residues properties from the Mean

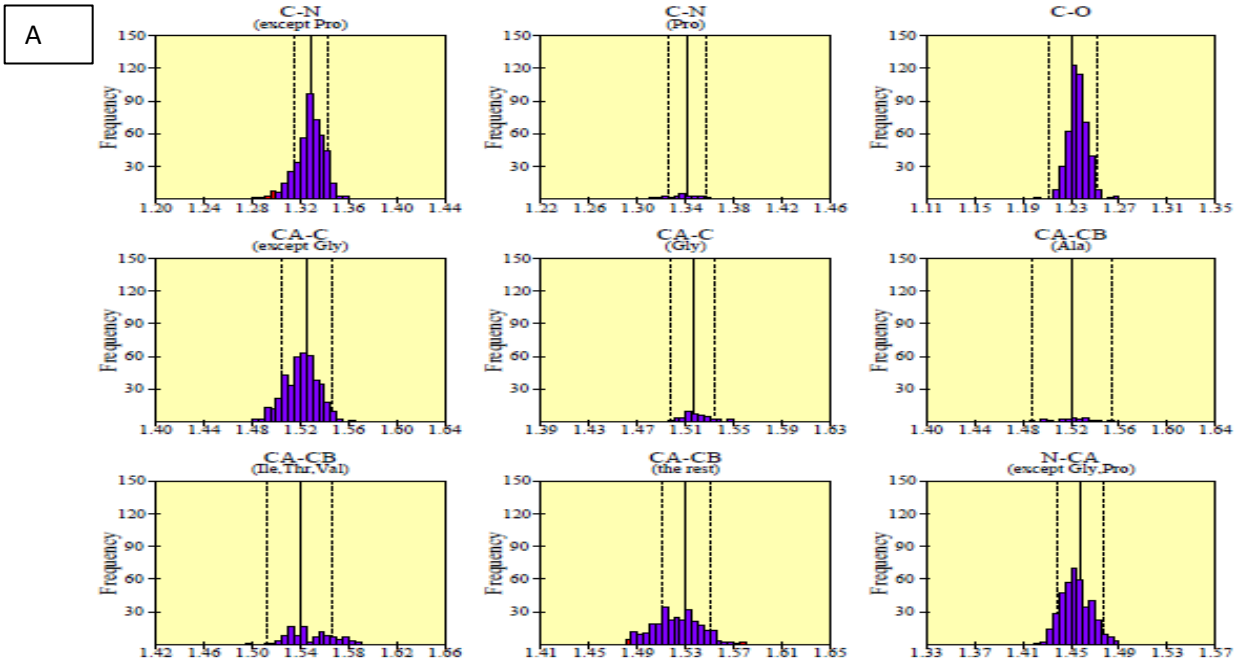

Fig. 6 (a). Bond length side chain different residues (A)

B

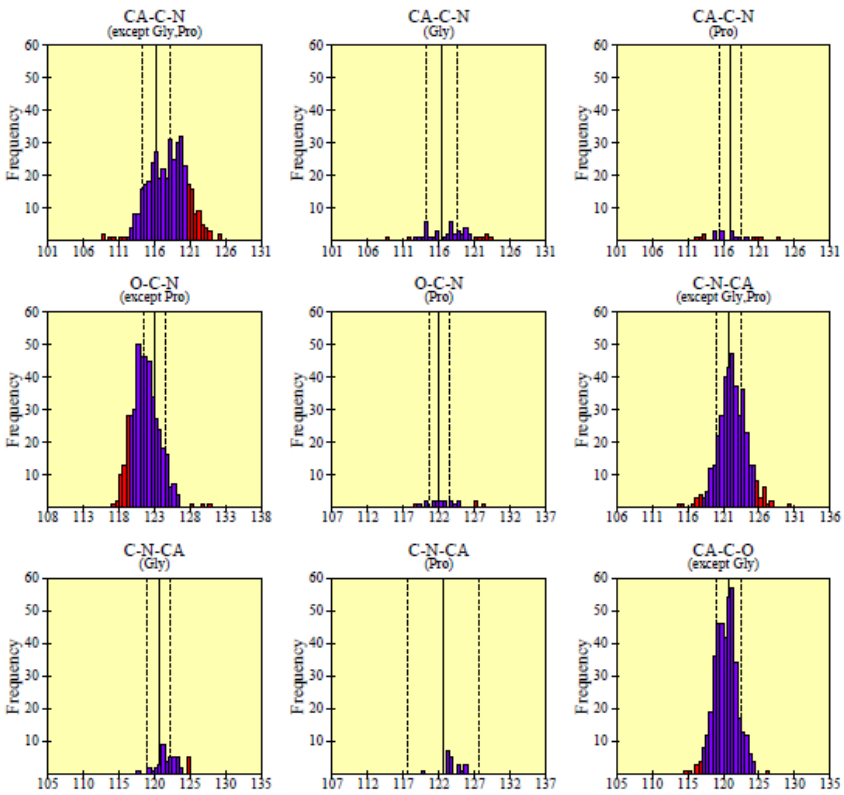

Fig. 6 (b). Bond length Main chain different residues (B)

Sidechain parameters are named chi1, chi2, etc. The chi1 angle is subject to certain restrictions that arise from steric hindrance between the gamma side chain atom(s) and the main chain. The different conformations of the side chain as a function of chi1 are referred to as gauche (+), trans and gauche (-) shown in Fig. 7.

Family and interpretation of domains of Hypodermin C

Interproscan is used for the interpretation of the function of proteins. It classifies the protein into families, subfamilies and gives information about the domains of the protein. Hypodermin C belongs to the family of peptidase S1A, chymotrypsin. The members of this family cause the intracellular digestion of microbes in neutrophils and fibrinolysis. The important catalytic residue is about 230 amino acids. The domains of protein represent the active sites of the protein. In Hypodermin C, this activity is conducted by a charge relay system which includes aspartic acid residues with histidine and histidine is hydrogen-bonded with serine. The biological function of peptidases is proteolysis and the molecular function is the serine-type endopeptidase activity.

Docking Analysis

Interpretation of Protein-protein interaction is the main provocation in the field of proteomics. The docking exhibits the bounded 3D structure of two proteins that interact with each other. The Fourier correlation method covers the translational and rotational spaces between two molecules. The docking results in very few native structures. It incorporates the binding energies of the structures. Other methods are used to purify the interacting regions of the protein (9). The refinement process involves the vanderwaal energy enhances and it improves the surface complementarity between bounded structures.

Cluspro 2.0 selects based on cluster size, given that the lowest energy structures generate the largest clusters. Anyway, the result has been given in the center of the cluster. However, the scores are not a representation of the binding affinity Secondly, even though ClusPro2.0 is CAPRI top-ranked docking software, it does not perform refinement because of the enormous computational power they need for their open-source program (10).

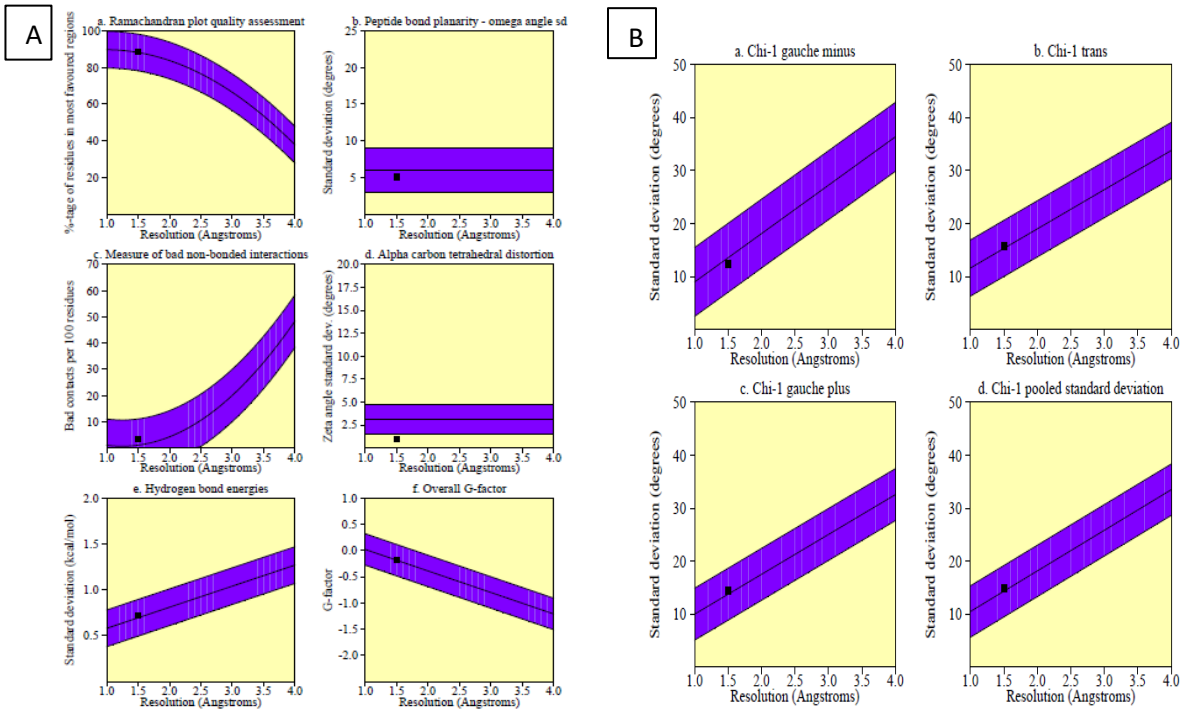

Fig. 7. Main Chain parameters (A)**, Side** Chain parameters (B)

Table 4: Docking Results

Cluster	Members	Representative	Weighted Score
0	141	Center	-858.3
0	141	Lowest Energy	-944.9
1	132	Center	-814
1	132	Lowest Energy	-986.5
2	110	Center	-816.1
2	110	Lowest Energy	-932.4
3	49	Center	-814.3
3	49	Lowest Energy	-877.1
4	46	Center	-863.1

DISCUSSION

Hypodermin C belongs to the family of serine proteases which catalyzes the lysis of protein. This protein comprises 260 amino acids. Hypodermin C causes the lysis of collagen protein. The main goal of this study is to understand the interface interaction between the hypodermin C and collagen. This study explains how the interactive regions of collagen and hypodermin C bind with each other cause the proteolysis in the collagen of bovine animals. The spatial arrangement of aspartate, histidine and serine induce the protein lysis.

Hypodermin C is also called a collagenases enzyme because it belongs to the family of trypsin and chemotrypsin. The members of this family carry out the breakdown of peptide bonds within the amino

acids and also called endopeptidases. These enzymes cause intracellular digestion of microbes in neutrophils. The protein sequence was retrieved from the NCBI and sequence similarity was checked through BLAST which showed the highest query coverage with serine proteases. The main purpose of assessment and analysis was how serine is involved in interaction important to the collagenase action of the enzyme as in these proteases' histidine, aspartate, serine and H-bonded with each other. They split positively charged amino acids like arginine and lysine. The breakdown is always done by the residues that are determined by the histidine, aspartate and serine. Serine plays a major role in breaking bonds of the molecule (11). It attacks Serine is necessary for the direct and indirect breakdown of bonds between the molecules. (12). Their active site contains serine, aspartate and histidine, also known as a catalytic triad, but recent studies have discovered (13) that the glutamate and lysine residues are also essential for the endo-proteolysis. The catalytic triad comprising of Asp-Ser-His is present in the active site of the serine protease. This triad plays an important role in the dissolution of the peptide bond of the substrate. The Substrate binds to the surface of the Hypodermin, serine acts as a nucleophile as it has a hydroxyl group (OH). The hydroxyl group of Serine attacks the carbonyl carbon of the scissile peptide bond of the substrate with the assistance of Histidine. The hydrogen of OH of serine is accepted by the nitrogen of Histidine and the pair of electrons move from double bond of carbonyl oxygen of substrate to the oxygen of Serine forming a tetrahedral intermediate. The peptide bond present in the substrate is now broken. The electrons making this bond now move to attack the hydrogen of Histidine. The connection breaks and the previously moved electrons from carbonyl oxygen move back from negative oxygen to recreate the bond. Water formed during the reaction replaces the N terminus of the cleaved substrate and attacks carbonyl carbon. Again, the electrons move to the oxygen and make it negative oxygen. The bond between the oxygen of water and carbon is formed. This reaction is again assisted by Histidine as it takes a proton from the water forming another tetrahedral intermediate. The electrons that make the bond between the serine and carbonyl carbon move to attack this hydrogen that Histidine just acquired from water. Now carbonyl carbon is electron-deficient (3) and it recreates a double bond with the oxygen and the C terminus of the substrate is also ejected.

These flies are one of the causes of morbidity and mortality in the livestock sector and so far, no control strategies exist. For the last decade, researchers are trying to develop vaccines against myiasis. Although there is a dire need for the rapid development in genomic and proteomic analysis, the alternative control strategies are constantly evolving due to the drug resistance that has a strong impact on animal welfare (13). In this study, the structure of HyC was predicted based on homology modeling and its interaction with the collagen protein was studied via molecular docking technique. We are reporting the 3D structure of HyA and HyB for the first time and have shown how the catalytic triad is spatially arranged in the Hypodermin to bring about the proteolytic activity. The models of HyA and HyB were docked against their substrate collagen which showed for the first time how the nucleophilic triad of the HyA and HyB hydrolyze the peptide bond of the host's collagenase, responsible for puncturing the skin of cattle by breaking down the collagen protein which is an integral part of the epidermis to penetrate the body of cattle (14). Our results have illustrated how the characteristic structure of the HyA and HyB of warble fly larvae has been functionally evolved to proteolysis the skin collagen of its host, essential for parasitic activity. These studies also highlight the possibility of developing vaccines against HyA and HyB antigens as their structure is known.

REFERENCES

1. Yadav, A., Panadero, R., Katoch, R., Godara, R. and Cabanelas, E., 2017. Myiasis of domestic and wild ruminants caused by Hypodermatinae in the Mediterranean and Indian subcontinent. *Veterinary parasitology*, 243: 208-218.

2. Oryan, A., Razavi, S. and Bahrami, S., 2009. Occurrence and biology of goat warble fly infestation by Przhevalskiana Silenus (Diptera, Oestridae) in Iran. *Veterinary parasitology*, 166: 178-181.

3. Khodakaram-Tafti, A., Razavi, S.M. and Gorgipour, S., 2012. Morphopathologic characteristics of myiasis caused by Przhevalskiana silenus larvae in goats. *Veterinarski arhiv*, 82: 25-33.

4. Tabouret, G., Bret-Bennis, L., Dorchies, P. and Jacquiet, P., 2003. Serine protease activity in excretory–secretory products of Oestrus ovis (Diptera: Oestridae) larvae. *Veterinary parasitology*, 114: 305-314.

5. Jan, S., Lateef, M., Abbas, F., Maqbool, A., Jabbar, M.A., Kakar, H., Razzaq, A., Shah, S.H. and Kakar, E., 2014. Sero-

epidemiological studies on goat hypodermosis in northern upland Balochistan, Pakistan. *Pakistan Journal of Zoology*, 46.

6. Åsbakk, K., Oksanen, A., Nieminen, M., Haugerud, R.E. and Nilssen, A.C., 2005. Dynamics of antibodies against hypodermin C in reindeer infested with the reindeer warble fly, Hypoderma tarandi. *Veterinary parasitology*, 129: 323-332.
7. Mihăşan, M., 2012. What in silico molecular docking can do for the 'bench-working biologists'. *Journal of biosciences*, 37: 1089-1095.
8. Tuszynska, I., Magnus, M., Jonak, K., Dawson, W. and Bujnicki, J.M., 2015. NPDock: a web server for protein–nucleic acid docking. *Nucleic acids research*, 43: W425-W430.
9. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J., 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of computational chemistry*, 30: 2785-2791.
10. Comeau, S.R., Gatchell, D.W., Vajda, S. and Camacho, C.J., 2004. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics*, 20: 45-50.
11. Bunker, T., 2009. Time for a new name for frozen shoulder contracture of the shoulder. *Shoulder & Elbow*, 1: 4-9.
12. Polgár, L., 2005. The catalytic triad of serine peptidases. *Cellular and molecular life sciences CMLS*, 62: 2161-2172.
13. Dalton, J.P., Neill, S.O., Stack, C., Collins, P., Walshe, A., Sekiya, M., Doyle, S., Mulcahy, G., Hoyle, D. and Khaznadji, E., 2003. Fasciola hepatica cathepsin L-like proteases: biology, function, and potential in the development of first generation liver fluke vaccines. *International journal for parasitology*, 33: 1173-1181.
14. Rosenberg, G.A., Mun-Bryce, S., Wesley, M. and Kornfeld, M., 1990. Collagenase-induced intracerebral hemorrhage in rats. *Stroke*, 21: 801-807.