RJNMSR
Review Journal of Neurological & Medical Sciences Review

Volume 3, Issue 8, 2025

COMPUTATIONAL CHARACTERIZATION AND COMPARATIVE EVOLUTIONARY ANALYSIS OF AN UNCHARACTERIZED MEMBRANE-ASSOCIATED PROTEIN LOC410154 IN APIS MELLIFERA

Muhammad Mubashir[*1], Syed Maaz Gillani[2], Sibtain Nawaz[3], Hamid Ullah[4], Syed M Sadeem[5]

*[*1]Department of Biosciences, Muhammad Ali Jinnah University, Karachi*
[2,3]Lecturer, Department of Allied health sciences, University of Kotli AJK
[4]Shaukat Khanum Memorial Cancer Hospital Peshawar Abasyn University Peshawar
*[*5]Department of Biosciences, Muhammad Ali Jinnah University, Karachi*

[*1]mubashirajmal39@gmail.com, [2]syedmaaz088@gmail.com, [3]sibtainraja12@gmail.com, [4]hamidullah11165@gmail.com, [5]sadeemsyed22@gmail.com

Corresponding Author: *
Syed Maaz Gillani
DOI: https://doi.org/10.5281/zenodo.17978279

Received	Accepted	Published
17 October2025	29 November 2025	17 December 2025

ABSTRACT

Un characterized proteins demonstrate huge knowledge gap in modern genome studies. This gap further widens in insects' studies such as Apis Mellifera which has economic and ecological significance. One third of its predicted proteins don't possess functional annotation. Apis Mellifera largely contributes in production of honey, beeswax, royal jelly and pollen which benefit food, cosmetics and pharmaceutical industries. Despite much research, many proteins associated with honeybees such as Apis mellifera remained unstudied, limiting our knowledge about understanding its survival instincts and evolutionary conservation across Hymenoptera. These proteins may play important role in immunity, sensory perception, cell signaling, transportation or pathogen interaction. LOC410154 in Apis mellifera is a large (1460aa) uncharacterized protein highly found across various honeybees and wasp species with identity exceeding 80-100% but still remained functionally or structurally unstudied. To uncover its biological role, it is necessary to study its placement, predicted localization and functional properties. Domain archetechture analysis presents EGF-like and adhesion related motifs suggesting a role in cell-cell interaction. Post translational modification profile predicted numerous O-glycosylation and Phosphorylation sites implying regulatory complexity. Structural modelling using Phyre2, RoseTTAFold/Robetta, and complementary approaches supported the presence of moduler domans. Furthermore, transmembrane analysis identified eight membrane associated helices. In silico approach provides powerful opportunity to explore its characteristics and annotate it without needing any laboratory experiment. This study aims to bridge the gap by in silico characterization of LOC410154 using sequence homology, domain prediction, physiochemical properties, subcellular localization, structural analysis and phylogenetic reconstruction.

Keywords: *Apis Mellifera, Computational characterization, Honey bees, Phylogenatic Analysis, Evolution*

INTRODUCTION

Genome sequencing projects has led to the identification of millions of protein-coding genes across diverse taxa. Despite this wonderfull progress in science, a significant number of predicted proteins remained annotated as uncharacterized particularly in

small organisms such as insects. This lack of functional annotation presents huge research gap in understanding the moleculer mechanisms that lead to underlying developement, physiology, immunity, and adaptation.

Honey bees (Apis spp.) play important role in global ecosystems through polination and it contributes substantially to agriculture, biodiversity and food security. Genomic data of Apis Mellifera and related insects have considerably improved, however, many proteins encoded remain poorly characterized. Understansding moleculer mechanisms of honey bees and other related insects is very essential for applied challenges such as pathogen resistance and environmental stress tolerance.

Loc410154 is an uncharacterized protein predicted in theb Apis mellifera genome. Database annotations suggest that protein is unusually large and highly conserved across various honey bees and wasp species. Although lack of experimental data limits our knowledge about this protein but computational analysis provides solid fundamentals for further research. Sequence analysis can identify conserved motifs and predict potential functions based on similarity to known proteins. Multiple sequence alignment and phylogenetic recnstruction enables us to find evolutionary trends across different species. Furthermore, computational protein structure prediction helps identifying domains, active site and potenyoal biological role.

In this study, we designed in silico pipeline to systematically characterize LOC410154.We combined homology searches, multiple sequence alignment, phyloheneatic analysis, domain and motif identification, PTM prediction, transmembrane topology analysis, and three-dimensional structure modeling. By synthesizing results, we aimed to make functional hypothesis and provide a comprehensive resource for further studies. To our knowledge, this is the first detailed computational investigation of LOC410154.

2. Materials and methods

2.1 Sequence Retrieval:

The amino acid sequence of LOC410154 was retreived from NCBI protein database for Apis mellifera (XP_016770576.1). Homologus sequence from related insect species were identified using BLASTp searched against the non-redundant(nr) protein database. Only sequences with high coverage and significant similarity (>80% identity) were selected for downstream analyses.

2.2 Homology Search and Dataset Construction:

BLASTp was performed using default parameters, and sequences from Hymenopteran insects showing ≥80% identity were retained. Redundant isoforms were removed, resulting in a curated dataset of 30 representative sequences.

2.3 Multiple Sequence Alignment (MSA):

MSA was conducted using Clustal Omega. Alignments were manually inspected to ensure correct alignment of conserved regions and domains. The aligned dataset served as the basis for phylogenetic analysis and conservation assessment.

2.4 Phylogenetic Analysis:

Phylogenetic reconstruction was performed using MEGA X. The best-fit substitution model was selected, and a Maximum Likelihood tree was generated with 1,000 bootstrap replicates to assess node support.

2.5 Domain and Motif Analysis:

Domain architecture was analyzed using InterProScan and Phyre2. Conserved motifs were identified by examining consensus regions in the MSA.

2.6 Post-Translational Modification Prediction:

O-glycosylation sites were predicted using NetOGlyc 4.0, and phosphorylation sites were predicted using sequence-based predictors focusing on serine, threonine, and tyrosine residues.

2.7 Transmembrane Region Prediction:

TM helices were predicted using ALOM, MTOP, and DeepLoc, with confirmation from AlphaFold2 models. TM1–TM8 sequences were analyzed individually.

2.8 Protein Structure Prediction:
Due to the large size of LOC410154, domain-wise structure prediction was performed using Phyre2, RoseTTAFold/Robetta, and AlphaFold2. Model confidence scores (0.65–0.71) and coverage metrics were used to assess reliability. Structures were visualized using online molecular viewers.

2.9 Physicochemical Analysis:
ProtParam was used to determine molecular weight, isoelectric point (pI), GRAVY, instability index, and aliphatic index.

3. Results

3.1 Sequence Homology and Conservation Analysis:
BLASTp analysis revealed that LOC410154 is a highly conserved protein across Hymenopteran insects. The protein exhibited 100% sequence identity with Apis cerana homologs and greater than 80% identity with homologous proteins from other bee and wasp species. High query coverage and extremely low E-values indicate strong evolutionary conservation and functional constraint. Such conservation suggests that LOC410154 performs an essential biological role that has been preserved throughout Hymenopteran evolution.

3.2 Multiple Sequence Alignment Analysis:
Multiple sequence alignment demonstrated extensive conservation across the full length of LOC410154, with particularly high conservation within predicted functional domains. Variable regions were mainly observed in inter-domain linker regions, suggesting a modular protein architecture. The preservation of conserved residues across species supports their potential structural or functional importance.

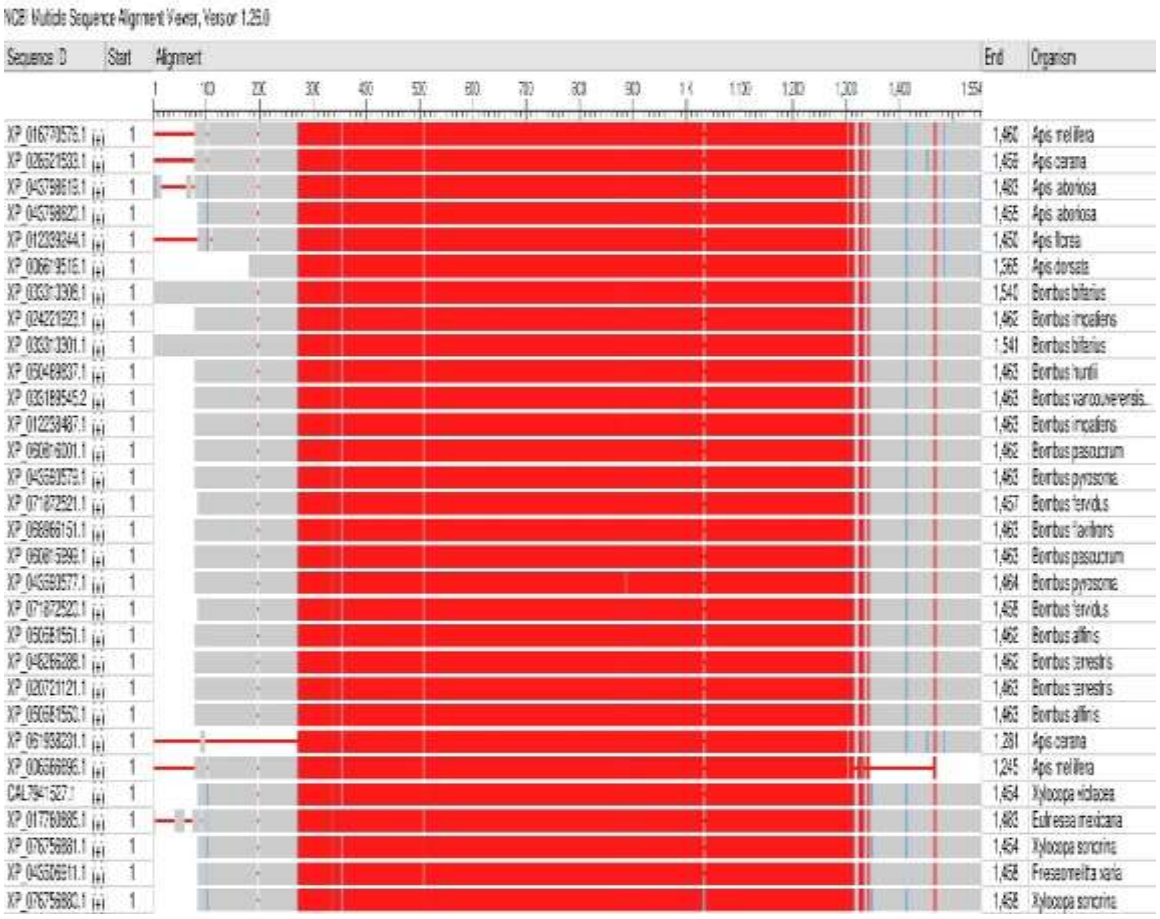


Fig: Multiple sequence alignment results (Clustral Omega)

3.3 Phylogenetic Analysis:

Phylogenetic reconstruction using the Maximum Likelihood method produced a well-resolved tree consistent with known Hymenopteran evolutionary relationships. Sequences from Apis species formed a strongly supported clade, clearly separated from wasp lineages. High bootstrap values (>90%) at major nodes confirm the robustness and reliability of the inferred phylogeny. These results indicate that the evolutionary history of LOC410154 mirrors species divergence patterns.

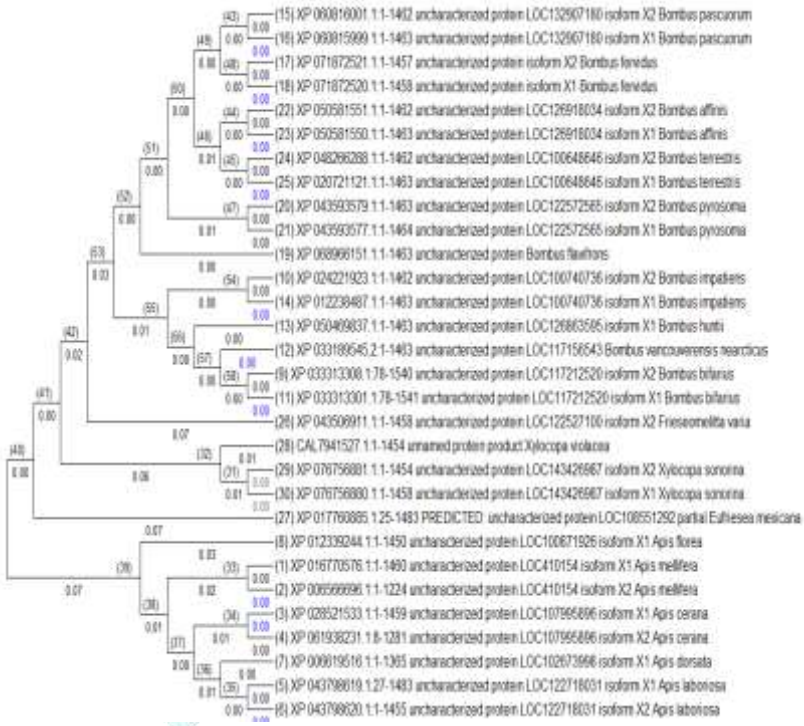


Fig: Maximum Likelihood phylogenetic tree using MEGA 12 software

3.4 Domain Architecture and Motif Identification:

Domain analysis revealed the presence of multiple EGF-like and adhesion-related domains within LOC410154. These domains are commonly associated with extracellular signaling, receptor interactions, and cell–cell adhesion processes. The repeated occurrence of such domains suggests a functional role in mediating molecular interactions at the cell surface or within signaling pathways.

3.5 Physicochemical Properties:

ProtParam analysis indicated that LOC410154 is a large, moderately stable protein with physicochemical characteristics consistent with regulatory or structural proteins. The predicted molecular weight, isoelectric point, and hydropathicity values support the presence of both soluble and membrane-associated regions, further suggesting functio

Feature	Details	Notes / Interpretation
Protein length	1460 aa	Large protein, suitable for structural & functional study
Molecular weight	165 kDa	Matches large receptor proteins
Theoretical pI	5.92	Slightly acidic, net negative at physiological pH

Feature	Details	Notes / Interpretation
Amino acid composition	Rich in Leu (9.3%), Ser (9.3%), Ile (8.2%), Thr (6.7%), Val (6.7%), Glu (6.2%)	Supports extracellular domain-rich structure
Charged residues	Negatively charged (Asp + Glu): 157 Positively charged (Arg + Lys): 134	Slight net negative charge, may affect solubility & interactions
Instability index	40.89	Borderline unstable → consider stabilization for expression
Aliphatic index	92.10	High thermostability
GRAVY	-0.109	Slightly hydrophilic
Extinction coefficient	208,345 M^{-1} cm^{-1} (cystines), 205,470 M^{-1} cm^{-1} (reduced Cys)	Useful for protein quantification
Estimated half-life	30 h (mammalian in vitro) >20 h (yeast in vivo) >10 h (E. coli in vivo)	Stable enough for experiments
Signal peptide	aa 1–26	Protein is secreted/extracellularly targeted
Transmembrane helices	TM1: 987–1009 TM2: 1058–1077 TM3: 1169–1191	Confirms membrane-bound receptor
Subcellular localization	Extracellular + membrane-bound + cytoplasmic	Likely GPCR-type receptor
Functional domains	GPCR family 2, extracellular hormone receptor (585–645) EGF-like domain (191–245) Ig-like domain (389–584)	Suggests signaling, adhesion, and extracellular interactions
GO terms	GO:0004930 (GPCR activity), GO:0016020 (membrane) PANTHER: GO:0005886, GO:0005911, GO:0050839	Supports receptor and adhesion function

Fig: Physicochemical properties of LOC410154 (MW, pI, GRAVY, instability index)

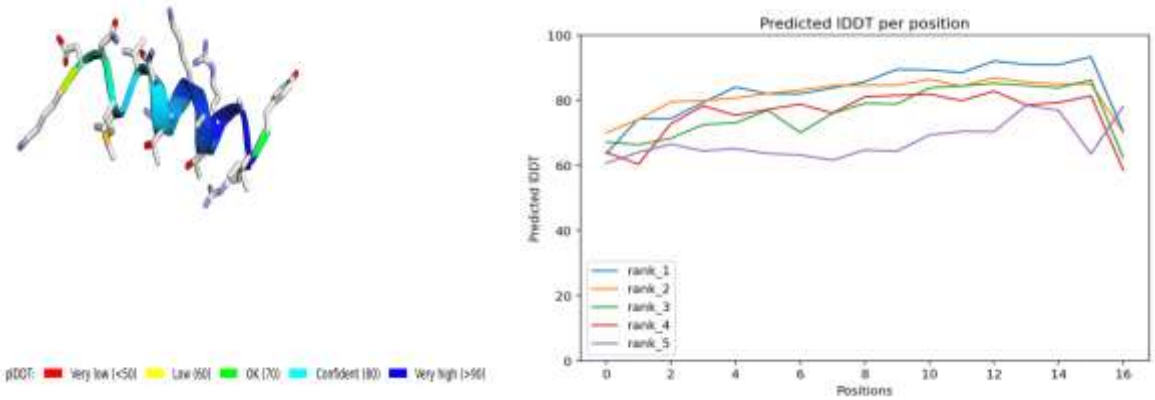
3.6 Post-Translational Modification Prediction:

PTM analysis predicted numerous O-glycosylation and phosphorylation sites distributed throughout the protein sequence. Several high-confidence O-glycosylation sites were identified, which are often involved in protein stability and extracellular interactions. The abundance of predicted phosphorylation sites suggests that LOC410154 may be dynamically regulated through signaling pathways.

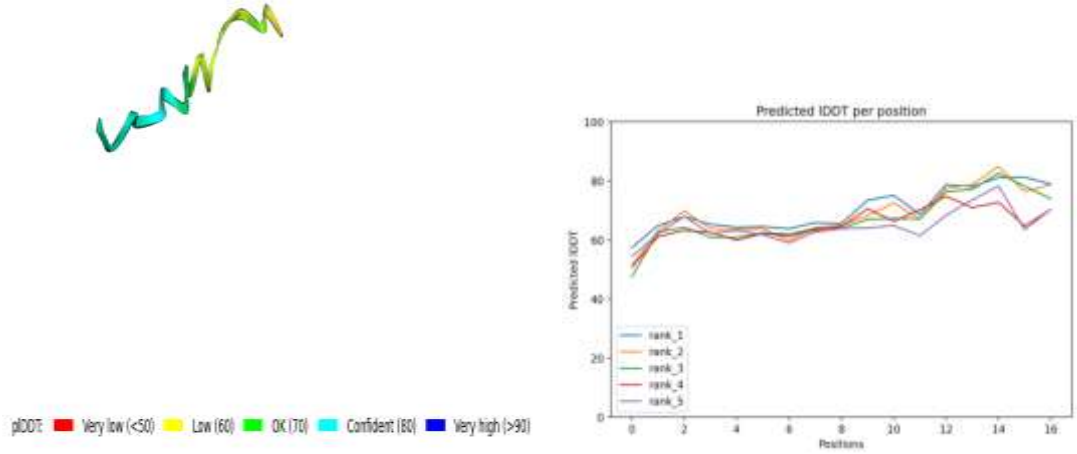
3.7 Transmembrane Helix Identification:

Topology prediction analysis identified eight distinct transmembrane helices (TM1–TM8) within LOC410154. These helices are distributed across the protein sequence and are predicted to adopt α-helical conformations typical of membrane-spanning regions. This strongly supports the hypothesis that LOC410154 is a membrane-associated or receptor-like protein.

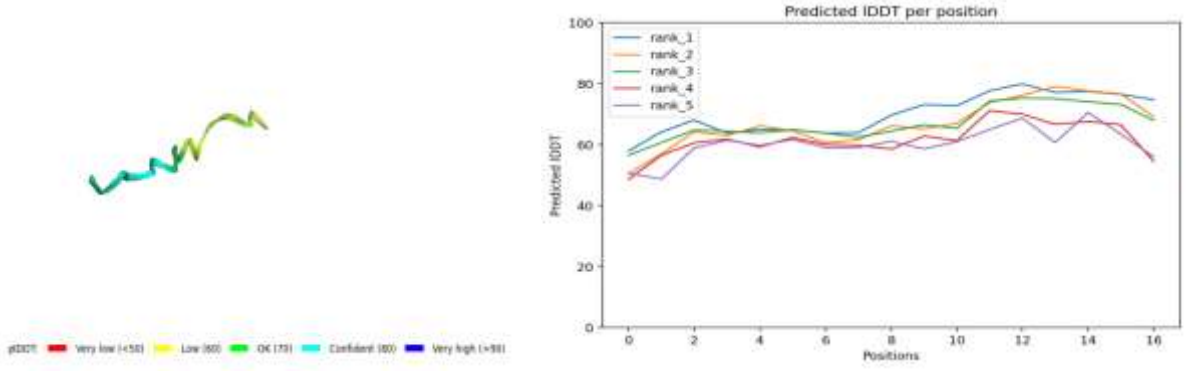
>Protein_TM1_932-948
KDNMLVNTNKATRAIRY



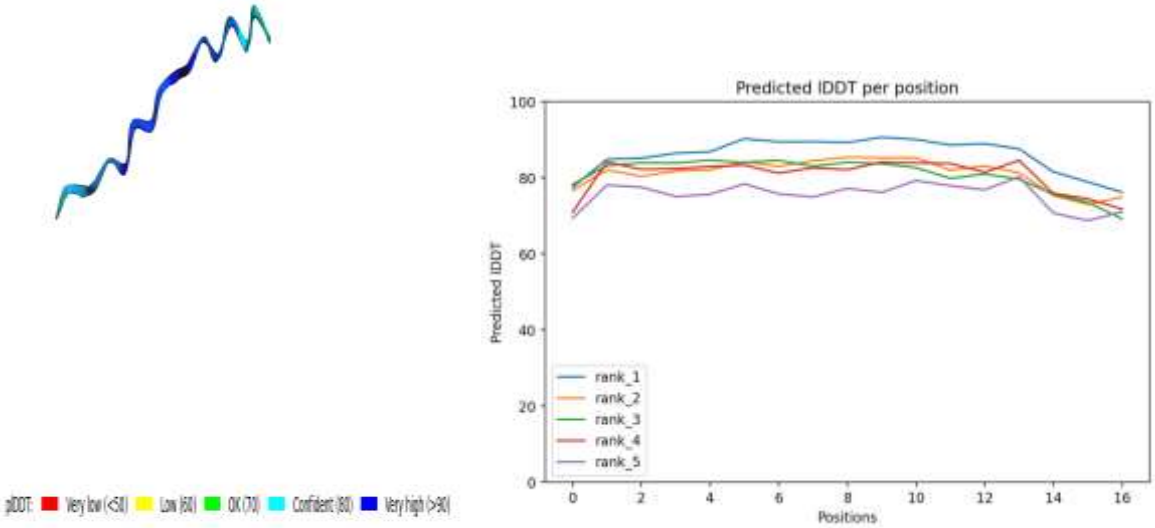
>Protein_TM2_954-970
KYFPNDGSDLHTSVLTI



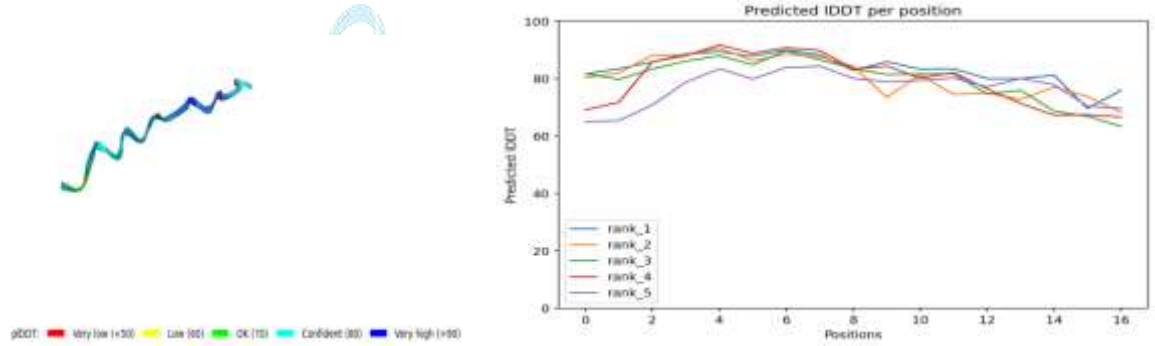
>Protein_TM3_993-1009
EKATLLDAGQYTCQIVD



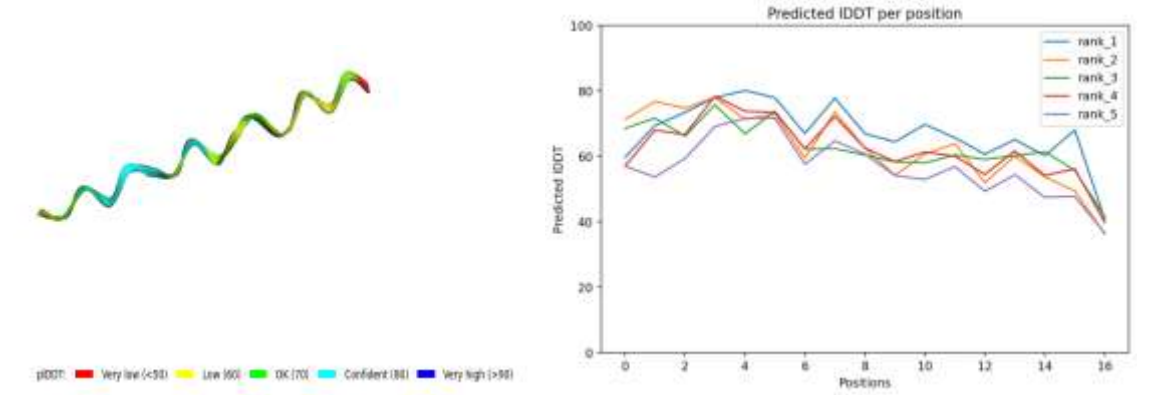
>Protein_TM4_1018-1034
WGVQQCKSIYIDVRDEP



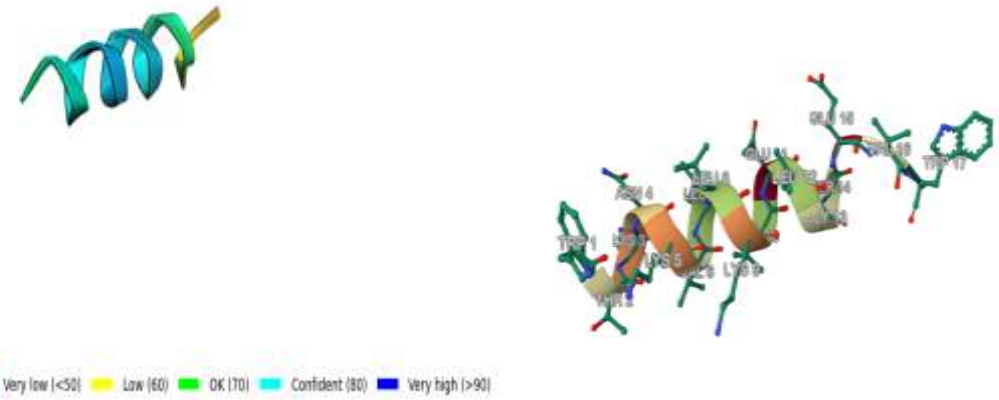
>Protein_TM5_1058-1074
DVKVMPMSVTIDKGTNI

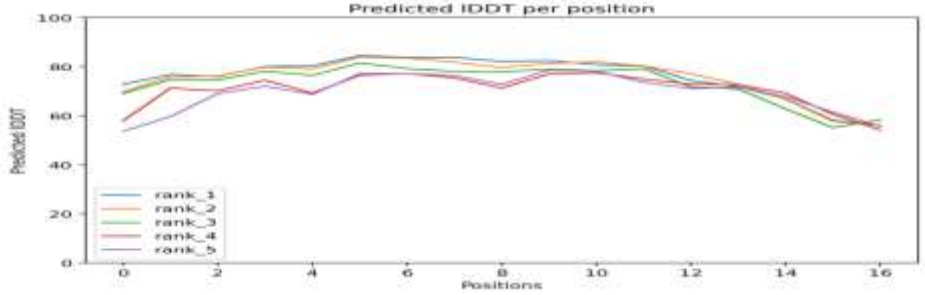


>Protein_TM6_1103-1119
QLTCMTPNVPNIGIGFG

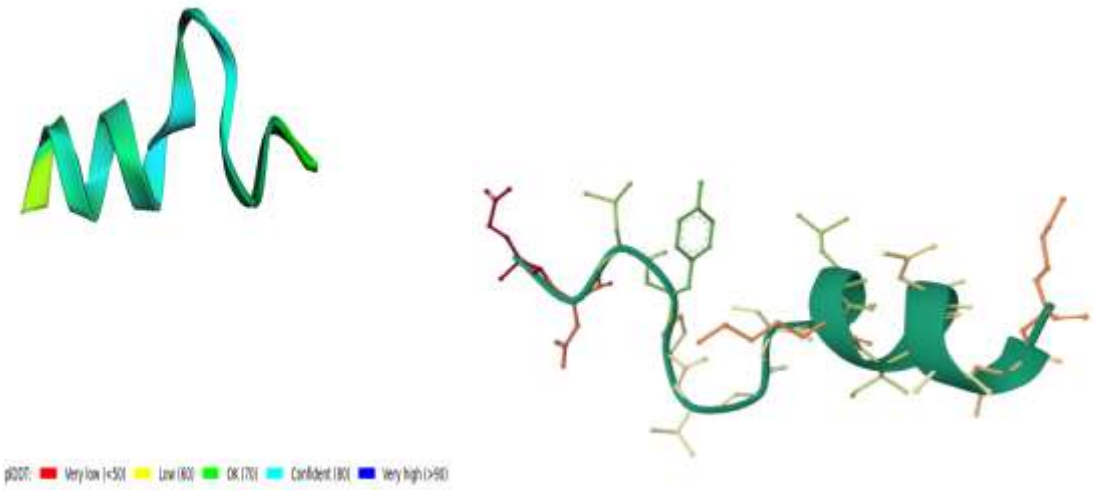


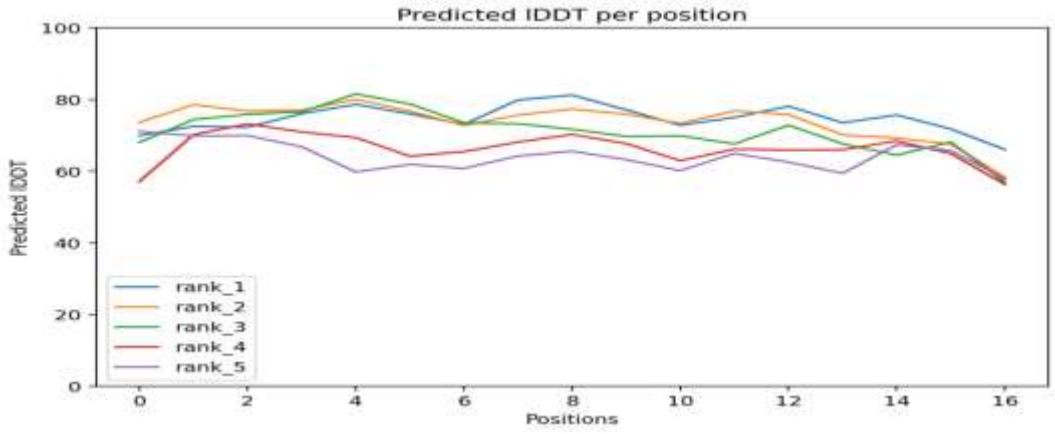
>Protein_TM7_1146-1162
WTKNKVLLKLELGSEVW





>Protein_TM8_1176-1192
EDLYPTGSILKITNAQK





3.8 Structural Modeling of LOC410154 and TM Segments:

Due to the large size of LOC410154, domain-wise structure prediction was performed. Structural models generated using AlphaFold and RoseTTAFold/Robetta showed moderate to high confidence scores (0.65–0.71) for individual domains. Predicted structures were predominantly α-helical, particularly within transmembrane regions, while extracellular domains exhibited folded architectures consistent with adhesion and signaling proteins.

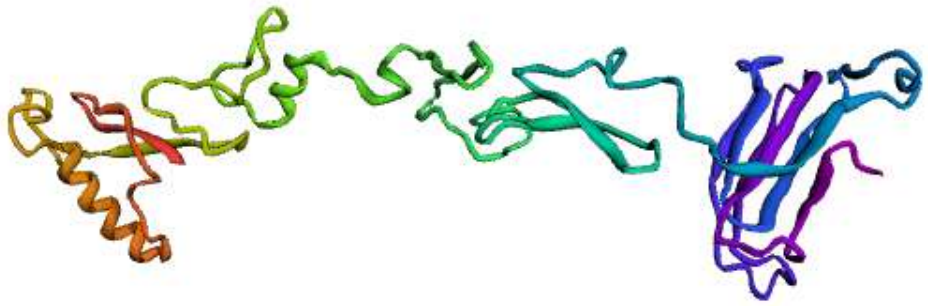


Fig: Phyre2 result of predicted domain 1

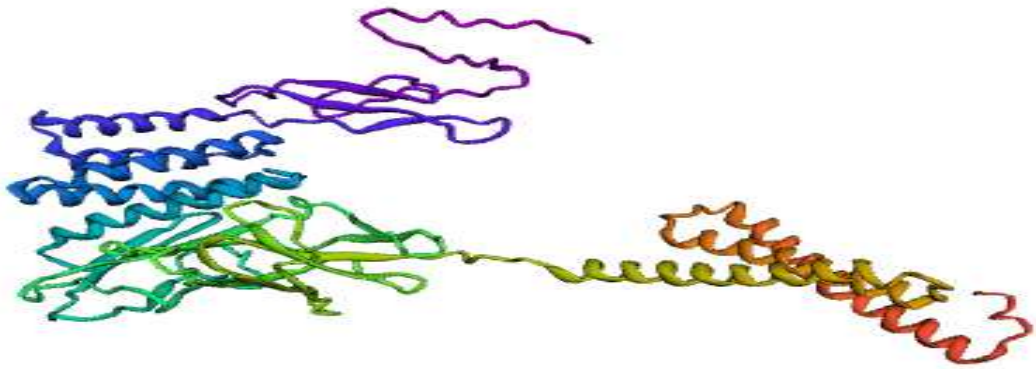


Fig: Phyre2 result of predicted domain 2

3.9 Integrated Functional Interpretation:

The integration of sequence conservation, phylogenetic clustering, domain architecture, PTM predictions, transmembrane topology, and structural modeling strongly suggests that LOC410154 functions as a conserved, membrane-associated protein involved in cell–cell communication or signaling. The convergence of independent computational methods increases confidence in these functional predictions.

Conclusion:

The integrative in silico analysis demonstrates that LOC410154 is a highly conserved protein

across Hymenopteran insects, suggesting strong evolutionary constraint and functional importance. Phylogenetic reconstruction revealed species-specific clustering consistent with known evolutionary relationships. Domain architecture analysis identified multiple EGF-like and adhesion-related motifs, indicating a potential role in cell–cell interaction or signaling. The presence of a signal peptide and eight predicted transmembrane helices supports a membrane-associated or receptor-like function. Numerous predicted O-glycosylation and phosphorylation sites imply complex post-translational regulation. Domain-wise structural modeling produced moderate to high confidence predictions consistent with known adhesion and signaling protein folds. Together, these findings provide the first comprehensive functional insight into LOC410154 and establish a strong foundation for future experimental validation.

REFERENCES:

Kocher, Sarah, and Callum Kingwell. "The molecular substrates of insect eusociality."Annual review of genetics 58.1 (2024): 273-295.

Li,Xinyu."Based proteomics analyses reveal response mechanisms of Apis mellifera(Hymenoptera: Apidae) against the heat stress."Journal of Insect Science 24.6 (2024): 6.

Svedberg,Dennis,etal."Functional annotation of adivergent genome using sequence and structure-based similarity."BMC genomics 25.1 (2024): 6.

Bono,Hidemasa,etal."Systematic functional annotation workflow for insects."Insects13.7(2022): 586.

Zheng, Shuang-Yan, et al. "A global survey of the full-length transcriptome of Apis mellifera by single-molecule long-read sequencing."International journal of molecular sciences 24.6 (2023): 5827.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., ... & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology, 7(1), 539.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution, 35(6), 1547-1549.

Käll, L., Krogh, A., & Sonnhammer, E. L. (2007). Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic acids research, 35(suppl_2), W429-W432.

Emanuelsson, O., Brunak, S., Von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nature protocols, 2(4), 953-971.

Galperin, M. Y., & Koonin, E. V. (2004). 'Conserved hypothetical'proteins: prioritization of targets for experimental study. Nucleic acids research, 32(18), 5452-546