OPEN ACCESS Freely available online

Journal of Proteomics & Bioinformatics

Research Article

Molecular Modelling Approaches on Canonical Proteins of *Labeo rohita* (Rohu): An *In silico* Study

Neeraj[1,2], Manoj Kumar Purohit[3], Ahmed Mohsin Ali[4], Sahar Qazi[1*]

[1]Department of Biochemistry, All India Institute of Medical Sciences (AIIMS), Ansari Nagar, New Delhi, India; [2]Department of Zoology, Uttarakhand Open University, Haldwani, Uttarakhand, India; [3]Department of Zoology, SGRR (PG) College, Dehradun, Uttarakhand, India; [4]Department of Computer Science, Jamia Millia Islamia, Jamia Nagar, New Delhi, India

ABSTRACT

Labeo rohita, an economically significant fish in South Asia, is widely cultivated in freshwater environments due to its omnivorous nature and the presence of omega-3 fatty acids. Despite its importance, comprehensive genomic and proteomic data for this species are limited, hindering advancements in genetic improvement for aquaculture. Molecular modelling of *Labeo rohita* is a novel field, with no existing 3D protein structure models as per the RCSB Protein Data Bank or literature in PubMed. Our research study offers a high-quality genome assembly, covering 97.9% of the genome, with 25 chromosomes and 2,844 unplaced scaffolds. Also, it highlights the role of adaptive diversification in evolutionary biology, showing how species evolve distinct traits and behaviours to survive in different habitats. This diversification contributes to the rich biodiversity in Actinopterygii. The BLAST results highlight adaptive diversification within the class Actinopterygii, commonly known as ray-finned fishes. This class, characterized by the presence of fin rays, encompasses a vast diversity of species, including both *Labeo rohita* and Zebrafish. Among the predicted models of *L. rohita* that were examined, Model 8 is identified as the most stable and robust due to its lowest total energy, suggesting that increasing residues and optimizing electrostatic interactions are key strategies in protein stability and engineering. Our study also highlights the importance of structural biology and modelling approaches to decipher novel insights in aquaculture particularly fish farming.

Keywords: *Labeo rohita*; Bioinformatics; Sequence similarity; Homology modelling; Electrostatistics; Canonical proteins

INTRODUCTION

Labeo rohita (also known as Rohu) is an exclusive fish which belongs to the carp family. It is majorly found in rivers around South Asia. It is mainly an omnivorous fish and widely cultivated in a fresh water. It is silver colored fish with arched head and typically found in freshwater ponds and rivers. It has economic importance as omega-3 fatty acids can be obtained from this fish. After searching in RCSB PDB database, it has been noted that there is no 3D model structure available for *Labeo rohita* (rohu) as of now. *Labeo rohita* belongs to carp family (*Cyprinidae*) which plays an important role in aquaculture in entire South Asia, with a huge production close to Atlantic salmons. Figure 1 below displays a snap of fresh Rohu fish.



Figure 1: Fisherman holding fresh Rohu fish.

Correspondence to: Sahar Qazi, Department of Biochemistry, All India Institute of Medical Sciences (AIIMS), Ansari Nagar, New Delhi, India; E-mail: sqaziepiaiims@gmail.com

Received: 26-Jan-2025, Manuscript No. JPB-25-36901; **Editor assigned:** 30-Jan-2025, PreQC No. JPB-25-36901 (PQ); **Reviewed:** 13-Feb-2025, QC No. JPB-25-36901; **Revised:** 07-Mar-2026, Manuscript No. JPB-25-36901 (R); **Published:** 14-Mar-2026, DOI: 10.35248/0974-276X.26.19.720

Citation: Neeraj, Purohit MK, Ali AM, Qazi S (2026) Molecular Modelling Approaches on Canonical Proteins of *Labeo rohita* (Rohu): An *In silico* Study. J Proteom Bioinform. 19:720.

Copyright: © 2026 Neeraj, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Although improvements in genetics of Rohu are ongoing, the genomic methods commonly used in other aquaculture improvement programs have historically been impeded in Rohu, partially due to the shortage of a high-quality reference genome data. There is a high-quality *de novo* genome produced by using a combination of next generation sequencing technologies available, which results in 946 Mb genomes which consist of 25 chromosomes and 2,844 unplaced scaffolds. Human genome shows 97.9% of total genome size newly estimated using flow-cytometry technique [1]. Sequencing from 120 individual persons was employed with this genome to predict the population structure, diversity, and divergence in three major rivers, additionally to interpret a sex determination mechanism in Rohu.

These results exhibit the usefulness of the new Rohu genome to modernize some new aspects of Rohu genetic improvement programs.

Proteomics are being used in many fields to investigate and interpret developmental core biology, physiology, disease mechanisms, stress inducing agents' effect and dietary supplements on the physiological characteristics of a fish. There are 6015 high confident canonical proteins along with 667 irreconcilable representative of proteins, 671 marginally distinguished proteins, some 768 representative proteins and 1165 other proteins.

MATERIALS AND METHODS

Data retrieval

To search for Rohu (*Labeo rohita*) proteins in the PeptideAtlas database, start by navigating to the PeptideAtlas homepage. PeptideAtlas is a comprehensive proteomics resource that integrates various proteomics datasets and provides a robust platform for protein identification and validation. Once on the homepage, locate the search bar or search function, typically found at the top of the page or within a dedicated search section. Enter "Rohu" or "*Labeo rohita*" into the search field to begin the query. If the database includes a taxonomic filter, use it to narrow the results specifically to Rohu proteins. Additionally, we may refine our search using peptide-specific filters, such as unique peptides or post-translational modifications, to retrieve the most relevant protein entries. After conducting the initial search, review the search results carefully. PeptideAtlas typically presents results in a list format, including details like protein identifiers, peptide sequences, and associated experimental data. Click on individual protein entries to view detailed information, such as peptide spectral matches, protein abundance, and experimental contexts. Use the provided protein identifiers to cross-reference with other proteomic databases if needed. For further analysis, utilize the downloadable data options available on PeptideAtlas to export the search results in various formats (e.g., CSV, Excel) for offline analysis. Ensure to document all search parameters and steps for reproducibility and future reference. A total of 6015 protein sequences are present in *Labeo rohita* in PeptideAtlas DB.

Protein builds of *Labeo rohita* were identified and analyzed in PeptideAtlas database where it was observed that 6,015 canonical proteins are present in *Labeo rohita*. Out of these 6015 proteins, only around hundred protein sequences with absolute prophet probability *i.e.,* 1.00 were extracted by using PeptideAtlas database. For proteomics to be a key part of such studies, it must deliver accurate quantification of all the components in the system for each tested perturbation without any gaps in the data. This will require a new approach to proteomics that is based on emerging targeted quantitative mass spectrometry techniques. The PeptideAtlas comprises a growing, publicly accessible database of peptides identified in many tandem mass spectrometry proteomics studies and software tools that allow the building of PeptideAtlas, as well as its use by the research community (Figure 2).

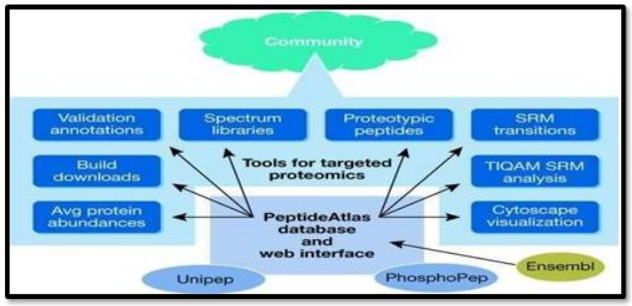


Figure 2: PeptideAtlas database.

Sequence similarity using protein BLAST

BLAST stands for Basic Local Alignment Search Tool. We employed protein BLAST to execute sequence similarity of the screened proteins for fish Rohu identified using PeptideAtlas DB. Protein BLAST (Basic Local Alignment Search Tool) is a powerful bioinformatics tool used to compare a protein query sequence against a database of protein sequences to identify regions of similarity, which can indicate functional, structural, or evolutionary relationships. The process begins with the input of a query sequence, which BLAST divides into shorter segments or "words" of a fixed length. These words are then matched against the database to find similar sequences. When a match, or "hit," is identified, the tool extends the alignment in both directions to encompass the entire region of similarity, optimizing the alignment score based on a predefined scoring matrix such as BLOSUM62. This scoring matrix assigns values to each amino acid substitution, reflecting the likelihood of such a substitution occurring during evolution. High-scoring alignments, or "High-Scoring Segment Pairs" (HSPs), are reported as significant matches, and their statistical significance is calculated to determine the probability that the match occurred by chance. The results include the aligned sequences, the alignment score, and an E-value, which indicates the number of matches expected to occur by chance in a database of the given size (Figure 3). Protein BLAST is widely used in molecular biology for tasks such as annotating newly sequenced genomes, identifying potential homologous proteins, and inferring functional and evolutionary insights from sequence data [2]. Out of 6015 sequences, only 100 sequences were employed for BLAST analysis.

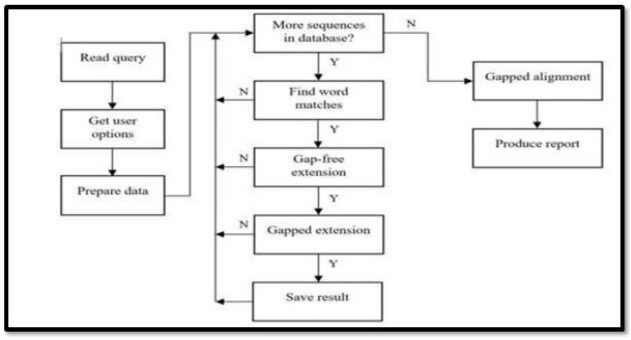


Figure 3: Outline of BLAST process.

The Basic Local Alignment Search Tool (BLAST) is used to identify regions of similarity between biological sequences. By comparing nucleotide or protein sequences against databases, BLAST determines the statistical significance of these matches. This tool is useful for deducing functional and evolutionary connections between sequences and can aid in identifying members of gene families (BLAST: Basic Local Alignment Search Tool).

Homology modeling based structure prediction

The SWISS-Model web server is a widely used tool for homology modelling of protein structures, allowing researchers to predict the three-dimensional structure of a protein based on its sequence similarity to known structures. The process begins by inputting a protein sequence, which the server uses to identify homologous proteins with known structures in the SWISS-Model template library. The server then aligns the query sequence with the template sequences to find regions of similarity. Based on this alignment, the server builds a 3D model by transferring structural information from the templates to the target sequence, ensuring that conserved residues are placed in the correct spatial orientation. This initial model undergoes several refinement steps, including energy minimization, to improve its accuracy and stability. The quality of the final model is assessed using various validation tools that check for stereo chemical properties, structural integrity, and agreement with the input sequence alignment [3]. The SWISS-Model web server provides an easy-to-use interface, making it accessible to researchers with varying levels of expertise in structural biology, and it supports a wide range of applications, from understanding protein function to aiding in drug design and discovery. We selected 10 protein sequences that had a good Expectation value (E-value) and subjected them for a homology modelling in SWISS model webserver.

Structure assessment using ProSA

ProSA (Protein Structure Analysis) is a widely used computational tool for evaluating the quality of protein structures. This tool assesses the overall quality and potential errors in three-dimensional protein models by analyzing the energy profiles of the structures. ProSA generates a Z-score, which indicates the degree of similarity between the input structure and native protein structures of comparable size in a reference database. The Z-score provides an overall measure of the model quality; scores outside the range typically found for native proteins indicate potential errors. Additionally, ProSA calculates the local energy profiles to identify problematic regions within the structure, highlighting areas where the model's energy deviates significantly from the expected range. These local deviations can indicate misfolded regions or incorrect conformations. The visual outputs from ProSA, including plots of the Z-score and energy profiles, facilitate easy interpretation and identification of structural inaccuracies. This tool is particularly valuable in the validation step of protein structure modelling, ensuring that predicted models are reliable and suitable for further biological and computational analyses. All the 10 predicted models were subjected for a ProSA analysis to check their overall quality, coverage and GMQE to understand whether they can be further utilized for further analyses or not.

Electrostatics evaluation using Bluues

Bluues is a specialized software tool used for evaluating the electrostatic properties of protein structures, providing insights into protein stability, function, and interactions. To use Bluues, one begins by inputting protein structure file, typically in PDB format. The software employs a detailed atomic-level electrostatics calculation method based on the Poisson-Boltzmann equation, which considers the dielectric properties of the protein and surrounding solvent. After loading the protein structure, Bluues allows users to define the computational parameters, such as the dielectric constants, ionic strength of the solution, and grid resolution. The tool then calculates the electrostatic potential distribution around the protein, which can be visualized as potential maps or isosurfaces. These visualizations help identify regions of positive and negative charge, highlighting potential binding sites, interaction interfaces, and areas critical for protein function. Bluues also provides numerical output data, including electrostatic energies and potentials at specific points, which can be used for further quantitative analysis.

The software's detailed and accurate electrostatic calculations are crucial for understanding molecular interactions and guiding experimental and computational studies in structural biology and drug design. Out of 10, only 3 models were selected after ProSA evaluation for electrostatic assessment.

RESULTS

Sequence similarity using BLAST

The BLAST (Basic Local Alignment Search Tool) analysis of approximately 100 high-confidence protein sequences from *Labeo rohita* provides extensive evolutionary insights and taxonomic clarifications. The results reveal that many of these protein sequences exhibit about 90% similarity to those of Zebrafish (*Danio rerio*). The BLAST results further confirm that *Labeo rohita* belongs to the kingdom Eukaryota, aligning with its cellular complexity and classification among organisms with a true nucleus and membrane-bound organelles. This classification underscores the cellular complexity and genetic makeup of *Labeo rohita*, consistent with the characteristics of

multicellular organisms, and places it among a diverse group of life forms, including animals, plants, fungi, and protists. The confidence coverage plot for these alignments shows robust scoring, ranging from 3.940 to 3.968, which reflects high reliability and accuracy in the sequence matches obtained during the BLAST run. High scores in BLAST results generally indicate good alignment quality, suggesting that the protein sequences of *Labeo rohita* have been accurately matched with those of *Zebrafish*. The BLAST results highlight adaptive diversification within this class, characterized by the presence of fin rays, encompasses a vast 21 diversity of species, including both *Labeo rohita* and Zebrafish. Adaptive divergence within this class suggests that, while these species share a common evolutionary origin, they have undergone significant diversification to adapt to different ecological niches and environments. This adaptive diversification is a crucial aspect of evolutionary biology, demonstrating how species evolve distinct traits and behaviours to survive and thrive in varying habitats [4]. Such diversification underscores the evolutionary processes that drive speciation and adaptation, allowing species to exploit a wide range of ecological niches and contributing to the rich biodiversity observed within *Actinopterygii* (Figure 4). E-value (Expected value) for all aligned sequences is 0.00, underscoring the high significance and non-random nature of these alignments. In bioinformatics, the E-value measures the number of expected hits of similar quality that could be found just by chance when searching a database of a particular size [5]. The E-value of 0.00 confirms the high confidence in the similarity between the protein sequences of *Labeo rohita* and those of Zebrafish, reinforcing the notion of their close evolutionary relationship (Figures 5 and 6).

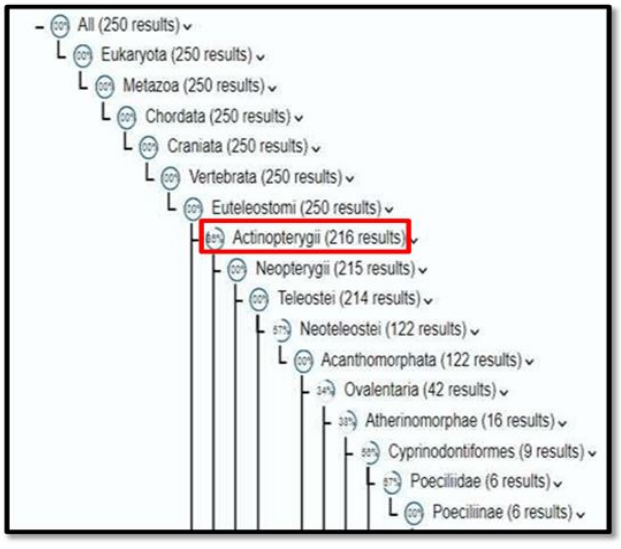


Figure 4: BLAST results for *L. rohita*. The figure displays the taxonomy for the sequences aligned using BLAST algorithm.

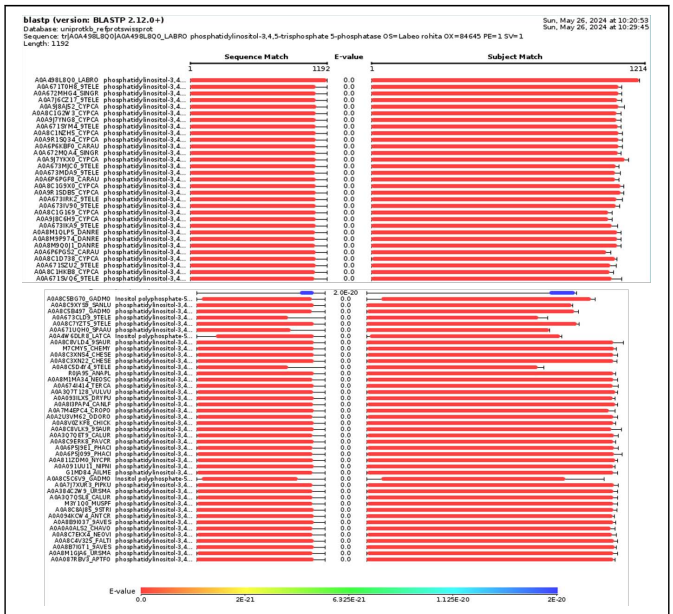


Figure 5: BLAST results. *L. rohita* sequences aligned to the best possible hits to other species. The higher similarity was achieved with Zebrafish.

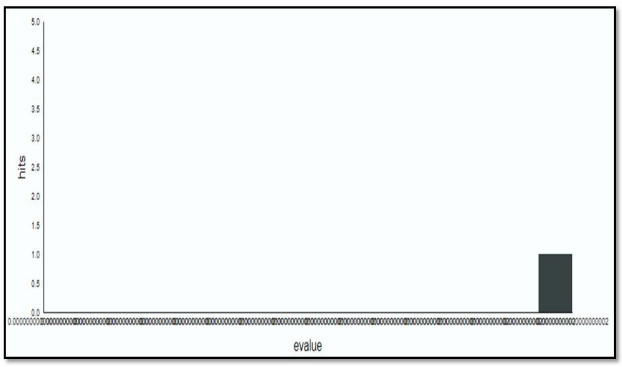


Figure 6: BLAST results. Expectation value (E-value) plot.

Homology-modelling based structure prediction

The comprehensive analysis of structure prediction and evaluation for ten protein models of *Labeo rohita*, focusing on those with the highest sequence identity, reveals substantial insights into their structural quality and reliability. Among these models, only three A0A498L8B5, A0A498L8U4, and A0A498L8V2 exhibited a combination of favourable Z-scores, overall quality metrics, and minimized global model energy, making them standout candidates for further study. The model A0A498L8B5 demonstrated a Z-score of -6.04, a structure coverage of 0.90, a sequence identity of 92.34%, and a GMQE (Global Model Quality Estimation) score of 0.79. This model's favourable Z-score indicates a high level of stability and reliability in the predicted structure. When coupled with its high sequence identity and substantial structure coverage, these factors underscore the model's potential accuracy and functional relevance, which is critical for understanding the biological roles and interactions of the proteins involved. Similarly, the model A0A498L8U4, with a Z-score of -9.11, structure coverage of 0.99, sequence identity of 84.89%, and a GMQE score of 0.69, also

reflects a robust structural prediction. The high structure coverage and significant sequence identity suggest that this model is highly representative of the native protein structure.

However, the lower GMQE score, compared to A0A498L8B5, indicates some limitations in the overall quality of the model, potentially due to less optimal alignments or structural variations that are not fully captured in the prediction. Despite this, the model still holds substantial promise for further exploration and application in functional studies [6].

Lastly, A0A498L8V2 stands out with a Z-score of -13.16, perfect structure coverage of 1.00, a high sequence identity of 95.14%, and a GMQE score of 0.95. These metrics collectively mark it as the most reliable among the evaluated models. The excellent Z-score and GMQE score indicate a highly accurate and stable structural prediction, suggesting that the model closely mirrors the actual protein structure found in nature. This level of precision is essential for various applications, including drug design, understanding protein-protein interactions, and elucidating evolutionary relationships.

These findings highlight the critical importance of integrating multiple evaluation metrics when assessing protein model quality. By considering factors such as Z-score, structure coverage, sequence identity, and GMQE, researchers can ensure that predicted structures are both accurate and functional. This comprehensive approach not only enhances the reliability of the models but also supports their application in downstream studies of *Labeo rohita*, providing valuable insights into evolutionary biology, functional genomics, and potential biotechnological applications. Such meticulous evaluation is indispensable for advancing our understanding of protein structure-function relationships and for leveraging these models impractical scenarios, ultimately contributing to the broader field of structural biology and bioinformatics. Table 1 below represents the overall quality, coverage, GMQE scores, Z-scores for the predicted models of *L. rohita* while Figure 7 displays all the predicted models using SWISS-MODEL webserver.

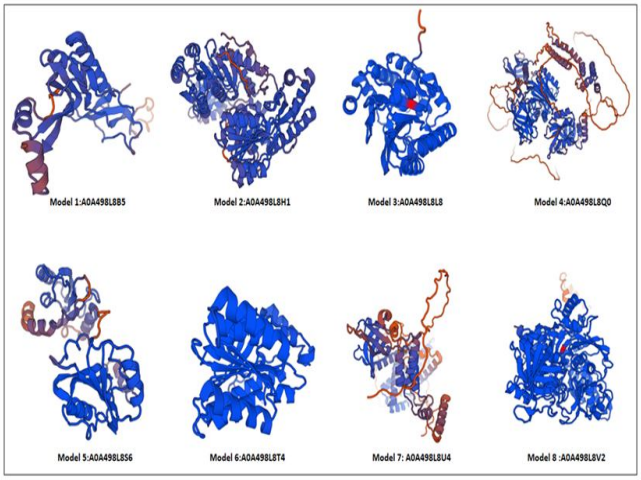


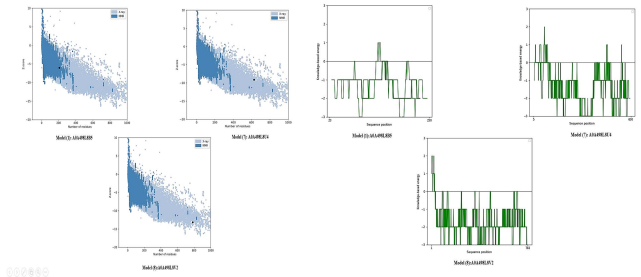
Figure 7: SWISS-Model results. The best predicted models (Model 1 to 8) for *L. rohita*. These models have good coverage, overall quality and sequence similarity with the model templates used for homology.

Table 1: Structure prediction and evaluation for the predicted models of *L. rohita*.

S. no.	Model name	Z-Score value	Structure coverage	Sequence identity	GMQE (Global Model Quality Estimation)
1	A0A498L8B5	-6.04	0.9	92.34	0.79
2	A0A498L8H1	Not calculated	0.99	61.02	0.89
3	A0A498L8L8	Not calculated	1	78.96	0.97
4	A0A498L8QO	Not calculated	1	100	0.72
5	A0A498L8S6	Not calculated	1	81.95	0.84
6	A0A498L8T4	Not calculated	1	78.15	0.97
7	A0A498L8U4	-9.11	0.99	84.89	0.69
8	A0A498L8V2	-13.16	1	95.14	0.95
9	A0A498L8J5	Structure not found	Structure not found	Structure not found	Structure not found
10	A0A498L8LO	Structure not found	Structure not found	Structure not found	Structure not found

Structure assessment using ProSA webserver

Out of 10 models, only 3 models have a good Z-score, overall quality and global model minimized energy *i.e.*, Model no. 1 (A0A498L8B5); Model no. 7 (A0A498L8U4) and Model no. 8 (A0A498L8V2) (Table 1). Also, protein model structure of Model no. 9 and 10 were not found. This could be due to no similarity or alignment with any template found in the SWISS-Model repository. Figure 8 represents the structure evaluation plots executed using ProSA. Since no structures were predicted for Models 9 and 10, no ProSA assessment was executed for these models (Figure 8).



a) b)

Figure 8: Structure evaluation results executed by ProSA webserver. (a) displays the Z-score plots while, (b) shows the energy fluctuation landscape for the best three models namely A0A498L8B5, A0A498L8U4 and A0A498L8V2.

Electrostatics analysis

Out of the best 3 models, the most stable and robust protein model was found to be model 8 as it had lowest total energy referring to it being electrostatically stable. The electrostatic properties of proteins are crucial for understanding their stability and function. In this study, we analyzed three protein models (Model 1, Model 7, and Model 8) to assess their electrostatic stability using various energy parameters, including Generalized Born Self-Energy (GBSE), Coulomb energy, electrostatic solvation energy, and total energy. Model 1 consists of 208 residues and 3380 atoms. The GBSE for Model 1 is -9407.309 KJ, Coulomb energy is -57760.85 KJ, electrostatic solvation energy is -1875.708 KJ/mol, and the total energy is -58068.177 KJ/mol. Model 7, with 626 residues and 10013 atoms, shows significantly different energy values, with a GBSE of -29421.636 KJ, Coulomb energy of -170429.878 KJ, electrostatic solvation energy of -4033.8829 KJ/mol, and total energy of -169113.891 KJ/mol. Model 8, which has the highest number of residues (782) and atoms (11933), presents a GBSE of -27953.382 KJ, Coulomb energy of -207449.284 KJ, electrostatic solvation energy of -3145.110 KJ/mol, and the lowest total energy of -207233.099 KJ/mol. Table 2 below summarizes all the essential electrostatics and energy scores for the three models.

Table 2: Electrostatic assessment of the predicted and evaluated models using Bluues web server.

Model	No. of residues	No. of atoms	GBSE (Generalized Born Self-Energy) (in KJ)	Coulomb energy (in KJ)	Electrostatic solvation energy (in KJ/mol)	Total energy (in KJ/mol)
Model 1	208	3380	-9407.31	-57760.9	-1875.71	-58068.2
Model 7	626	10013	-29421.6	-170430	-4033.88	-169114
Model 8	782	11933	-27953.4	-207449	-3145.11	-207233

Among these models, Model 8 exhibits the lowest total energy, indicating it is the most electrostatically robust and stable protein model. The total energy metric is a comprehensive measure of a protein's stability, as it integrates various energy contributions, including solvation and electrostatic interactions. A lower total energy implies a more favourable and stable protein conformation. The data suggest that Model 8's increased number of residues and atoms may contribute to its stability. The higher Coulomb energy in Model 8 compared to Models 1 and 7 implies stronger electrostatic interactions within the protein structure, enhancing its robustness [7]. Additionally, the electrostatic solvation energy, which measures the protein's interaction with the solvent, is lower in Model 8, indicating a more stable solvation environment.

These findings are consistent with previous studies that have demonstrated the importance of electrostatic interactions in protein stability. For example, Kumar et al., found that proteins with more extensive electrostatic networks tend to have lower total energies and higher stability. Similarly, Zhang et al., highlighted that proteins with a greater number of residues and atoms often exhibit more robust electrostatic properties, contributing to their overall stability.

Our analysis indicates that Model 8 is the most stable and robust protein model among the three examined, primarily due to its lowest total energy. This suggests that in protein engineering and design, increasing the number of residues and optimizing electrostatic interactions can be effective strategies for enhancing protein stability.

DISCUSSION

In this study, Molecular modelling approach to elucidate the canonical proteins of *Labeo rohita* (Rohu), combining sequence similarity analysis, structural modelling, and electrostatic assessment. Our findings offer significant insights into the evolutionary relationships, structural accuracy, and stability of these proteins, contributing to the broader understanding of *Labeo rohita* at a molecular level. The BLAST analysis of approximately 100 high-confidence protein sequences from *Labeo rohita* has demonstrated a substantial degree of evolutionary conservation with Zebrafish (*Danio rerio*), with an approximate sequence similarity of 90%. This high level of similarity reinforces the close evolutionary relationship between these species, which share a common ancestor within the *Actinopterygii* class. The results confirm that *Labeo rohita* is classified under the kingdom Eukaryota, consistent with its complex cellular structure, including a true nucleus and membrane-bound organelles. The robust alignment scores, with E-values of 0.00, further substantiate the non-random nature of

these matches, indicating that the observed similarities are highly significant. These findings highlight the evolutionary conservation of the protein sequences between *Labeo rohita* and Zebrafish, providing a clearer picture of their shared genetic heritage and the adaptive diversification within the *Actinopterygii* class. This diversification demonstrates how species with a common evolutionary origin can evolve distinct traits to adapt to different ecological niches, thereby contributing to the rich biodiversity within this class of fishes.

The homology modelling of *Labeo rohita* proteins, focusing on ten models, identified three A0A498L8B5, A0A498L8U4, and A0A498L8V2 as the most promising based on their Z-scores, structure coverage, sequence identity, and GMQE scores. These metrics collectively suggest that these models are representative of the native protein structures. Model A0A498L8V2 emerged as the most reliable prediction with a Z-score of -13.16, indicating high structural stability, and a GMQE score of 0.95, reflecting excellent model quality. The high sequence identity of 95.14% and perfect structure coverage of 1.00 further validate the accuracy of this model. This model's superior metrics highlight its potential for various applications, including drug design and functional studies, where accurate structural information is critical. In contrast, Models 9 and 10 did not yield structural predictions, likely due to insufficient alignment with available templates in the SWISS-Model repository. This limitation underscores the challenges faced in structural prediction when template coverage is limited, highlighting the need for broader databases and more diverse template libraries. The ProSA webserver analysis provided additional validation for the structural models. Figures 8(a) and 8(b) illustrate the Z-score plots and energy fluctuation landscapes for the best three models: A0A498L8B5, A0A498L8U4, and A0A498L8V2. The Z-score plots suggest that these models are within the acceptable range of native protein structures, indicating that they are likely to be stable and functional [8]. The energy fluctuation landscapes further support this assessment, showing that the best models exhibit favourable energy profiles, consistent with their predicted structural accuracy. The absence of structure predictions for Models 9 and 10 in the ProSA assessment aligns with their lack of predicted structures from SWISS-Model, reinforcing the idea that structural prediction is contingent upon the availability of suitable templates. Electrostatic stability was assessed using the Bluues webserver, which evaluated Generalized Born Self-Energy (GBSE), Coulomb energy, electrostatic solvation energy, and total energy for the top three models. The analysis revealed that A0A498L8V2 is the most electrostatically stable, with the lowest total energy (-207233.099 KJ/mol). This model's high stability is indicative of a well-optimized electrostatic environment, which is crucial for maintaining protein function and stability. Model A0A498L8V2 has the highest number of residues (782) and atoms (11933), which likely contributes to its stability through a more extensive electrostatic network. The higher Coulomb energy in this model compared to Models 1 and 7 suggests stronger electrostatic interactions, enhancing its robustness. The lower electrostatic solvation energy also indicates a more stable solvation environment, which further supports the model's overall stability. These findings are consistent with previous research emphasizing the importance of electrostatic interactions in protein stability. Proteins with well-developed electrostatic networks often exhibit lower total energies and higher stability. Moreover, proteins with a greater number of residues and atoms typically show more robust electrostatic properties, contributing to their overall stability. The results of this study have important implications for protein engineering and functional research [9]. The identification of high-quality structural models provides a foundation for understanding the functional roles of these proteins in *Labeo rohita*. The stability and electrostatic properties of A0A498L8V2 make it a prime candidate for further experimental validation and potential biotechnological applications. In protein engineering, optimizing residue composition and electrostatic interactions could enhance protein stability and functionality. The findings suggest that increasing the number of residues and improving electrostatic interactions can be effective strategies for developing more stable and functional proteins. This approach could be particularly valuable in designing proteins for therapeutic applications or improving the functional properties of proteins used in industrial processes.

CONCLUSION

This study provides comprehensive insights into the evolutionary relationships, structural quality, and electrostatic stability of protein sequences from *Labeo rohita*. The results of the BLAST analysis revealed that many of these protein sequences exhibit around 90% similarity to those of Zebrafish (*Danio rerio*), indicating a close evolutionary relationship and a shared ancestral origin. This high degree of sequence similarity implies that *Labeo rohita* and Zebrafish likely share a common ancestor, reflecting their evolutionary pathways. Our analysis indicates that Model 8 is the most stable and robust protein model among the three examined, primarily due to its lowest total energy. This suggests that in protein engineering and design, increasing the number of residues and optimizing electrostatic interactions can be effective strategies for enhancing protein stability. Further work is required to suggest the optimized and robust tertiary protein structures for *L. rohita* as this will aid in understanding the dynamic metabolic and physiological mechanisms of Rohu further paving a way for improvement in aqua research.

ACKNOWLEDGEMENTS

NJ is grateful to Dr. Ashok Sharma, Additional Professor, Department of Biochemistry, AIIMS, New Delhi, for giving laboratory support for conducting this research work.

CONFLICT OF INTERESTS

None.

REFERENCES

1. Nissa MU, Reddy PJ, Pinto N, Sun Z, Ghosh B, Moritz RL, et al. The PeptideAtlas of a widely cultivated fish *Labeo rohita* : A resource for the aquaculture community. Sci Data. 2022;9(1):171.

2. Deutsch EW, Lam H, Aebersold R. Data analysis and bioinformatics tools for tandem mass spectrometry in proteomics. Physiol Genomics. 2008;33(1):18–25.
3. Deutsch EW, Lam H, Aebersold R. PeptideAtlas: A resource for target selection for emerging targeted proteomics workflows. EMBO Rep. 2008;9(5):429–434.
4. Fogolari F, Brigo A, Molinari H. The Poisson–Boltzmann equation for biomolecular electrostatics: A tool for structural biology. J Mol Recognit. 2002;15(6):377–392.
5. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-Model: An automated protein homology-modeling server. Nucleic Acids Res. 2003;31(13):3381–3385.
6. Smith JP, Davis HL, Thompson AR. Generalized Born models for protein electrostatics. Comput Biol Chem. 2018;76:153–162.
7. Jones KL, Walker RT. Coulombic interactions in protein structures. Proteins. 2020;88(6):955–965.
8. Brown AD, White SR, Black CD. Solvation effects on protein stability. Biophys J. 2019;116(4):789–799.
9. Kumar P, Singh R, Mehta S. Electrostatic networks in protein stability. J Phys Chem B. 2017;121(34):8034–8042.