ANSF
APPLIED AND NATURAL
SCIENCE FOUNDATION

Research Article

Proteomic insights into the biology of *Clinostomum piscidium* from a fish, *Colisa fasciatus* in India using computational tools

Bhumika Chauhan
Laboratory of Molecular Parasitology, Department of Zoology, Chaudhary Charan Singh University, Meerut-250004 (U.P.), India
Anshu Chaudhary
Molecular Taxonomy Laboratory, Department of Zoology, Chaudhary Charan Singh University, Meerut-250004 (U.P.), India
Anju Tyagi
CSSSPG College Machhra, Meerut-250004 (U.P.), India
Bindu Sharma*
Laboratory of Molecular Parasitology, Department of Zoology, Chaudhary Charan Singh University, Meerut-250004 (U.P.), India
Hridaya S Singh
Molecular Taxonomy Laboratory, Department of Zoology, Chaudhary Charan Singh University, Meerut-250004 (U.P.), India

*Corresponding author. Email: dr.bindusharmazoology@gmail.com

How to Cite

Chauhan, B. *et al.* (2021). Proteomic insights into the biology of *Clinostomum piscidium* from a fish, *Colisa fasciatus* in India using computational tools.. *Journal of Applied and Natural Science*, 13(2), 455 - 462. https://doi.org/10.31018/jans.v13i2.2596

Abstract

Clinostomum is a fish-borne pathogen,digenetic trematode with worldwide distribution. Despite its zoonootic significance, the molecules involved in the host-parasite interaction remains unknown. The present study deals with the proteome profile of the rDNA of *Clinostomum piscidium* using in silico workflow.The physicochemical properties, mass spectrometry ,atomic composition,estimated half-life in different hosts ,Grand average of hydropathy (GRAVY),extinction coefficient and instability index of the rDNA was studied.We identified hypothetical proteins(polar in nature) that showed similarities with proteins of *Trichellina pseudospiralis* and *Melampsora laricipopulina.*The functionality of these proteins revealed their role in intracellular signalling as a substrate for O - linked N - acetylglucosaminetransferases and death transcription factor(DIDO-1). This is the first report of our findings that aims to provide a better understanding of the mechanism by which this digenetic trematode adapts to extreme environments. It is expected that this study will lead to new insights into drug designing strategies and disease control.

Keywords: *Clinostomum piscidium, Colisa fasciatus,* GRAVY, rDNA, In silico Mass spectrometry, Proteomics

INTRODUCTION

Fishes serve as an important component of the ecosystem from the ecological, nutritional, medicinal, and economic perspective. Fishes play an important role in nutrient cycles as they store a large proportion of nutrients in their tissues, transport nutrient and play a significant role in nutrient recycling. Eating uncooked fishes leads to the transmission of parasites to humans or any other host feeding on them and, thus, causes many diseases. Parasitic disease is the single most important factor threatening the fish industry worldwide, particularly in

the tropics (Schmidt and Roberts, 2000). The excysted progenetic larval metacercaria of *Clinostomum piscidium* serves as an infective stage in several genera and species of fish (Southwell and Prashad,1918). In Northern India, the *Colisa fasciata* (Bloch and Schneider, 1801) fish have a very common infection of *C. piscidium* excysted metacercariae. Digenetic trematode of the genus *Clinostomum* is widely distributed in many species of freshwater fishes and cause zoonotic diseases. Eating uncooked fish leads to the entry of *Clinostomum* in humans, an accidental host. At least 19 such cases have been reported from Japan(Hara *et al*., 2014).

Various ailments like pharyngitis and lacramalitis due to *Clinostomum* infection have been reported from Thailand and Korea. In addition to human, freshwater fishes (intermediate host) incur severe damage following heavy infection with metacercariae of *Clinostomum* species(Shareef and Abidi, 2012). The metacercariae have also been reported in stools of human and cause serious digestion problems, respiratory disorders upon entry into the lungs and many other severe consequences (Tiewchaloern *et al.*,1999; Park *et al.*,2009; Zimik*et al.*, 2019). The present study focuses on the physical and chemical properties of the rDNA of *C. piscidium,* which will serve as an important factor for estimating the prevalence range of the parasite, instability index, nature of the proteins and estimated half-life of the parasite in a different environment, 3D structure and functioning of the proteins.These parameters will further help determine the factors that help the parasite sustain and grow within the host, establish host parasite interactions, andserve as an important tool in disease control of *C.piscidium*. Proteomics is crucial for many reasons like early disease diagnosis, prognosis and monitoring the disease development. Molecular mass, theoretical isoelectric point (pI), amino acid composition, atomic composition, extinction coefficient, instability index, aliphatic index and high average hydrotherapy are predicted, which help in determining the various physicochemical parameters of the proteins.

The utilization of in silico approaches for the functional prediction of proteins has been successfully used in several helminths like *Ascocotyle longa, Schistosoma mansoni* cercarial secretions, *Schistosome* tegument and its surface membranes, *Fasciola hepatica, S. mansoni* egg and its contents, the miracidium(hatch fluid and secretions) and of parasitic nematodes (Menezes *et al.*, 2109). The purpose of this work was to assign a function to the hypothetical proteins present in the genome of this species for the identification of new parameters that may contribute to the physiochemical properties of *Clinostomum*. This will further improve the understanding of the adaptation of this parasite to the extreme environment.

MATERIALS AND METHODS

Retrieval of the sequence: In this work, we used rDNA*C. piscidium* gene. The sequence was retrieved from NCBI (National Centre for Biotechnology Information) database (http://www.ncbi.nlm.nih.gov)under accession number GQ925911.1.

Protein profiling
The protein profiling was done using ExPASy and six translated peptide sequence of different forward and reverse frames were obtained to unveil the translated protein frames of the rDNA sequence. Further studies were carried out using these peptide sequences to study the physicochemical properties of the zoonotic parasite.

Prediction of physiochemical parameters
Molecular mass, theoretical isoelectric point (pI), massspectrometry, amino acid composition, atomic composition, extinction coefficient, instability index, aliphatic index, estimated half-life and high average hydrotherapy were predicted using the Proteomics analysis tool (Gasteiger *et al.*, 2005) which allows in determining the several physicochemical parameters of the proteins.

Protein statistics and transmembrane orientations
Protein statistics was performed using ExPasY (Gasteiger *et al.*, 2009) and SAPS (Statistical analysis of protein sequence), which provided information about the number of times each amino acid is repeated, compositional analysis, charge distribution analysis and prediction of transmembrane spanning regions.

Phylo-proteomic analysis
For the taxonomy validity of the worm Phylo proteomic tree was constructed using neighbor joining method.

Prediction of secondary structure
The Secondary structure of the peptide sequence was also predicted using hydrophobic periodicity in the sequence and was used to assess the quality of a model built with a tertiary structure prediction method.

Determination of three dimensional structures
Three-dimensional homology modeling of the target proteins was performed by MODELLER. Swiss Model software was used to predict the structures for proteins through Homology Modeling principle. The structure predicted by using the methodology of Homology Modeling provides information insight about the functionality of proteins, dynamics, and interactions with other proteins

RESULTS AND DISCUSSION

Analysis of protein profiling of rDNA of zoonotic trematode
The analysis on the prediction of peptide sequence using the ExPASy platform indicated six forward and reversed frames of the translated peptide sequence. Out of these six frames, 5'3' Frame 1 (Fig. 1) was selected for further observations as it had maximum coding fragment starting with a Start codon and ending with an end codon. The selection of this frame was made on account of the fact that protein synthesis starts from methionine and ends with a stop codon.
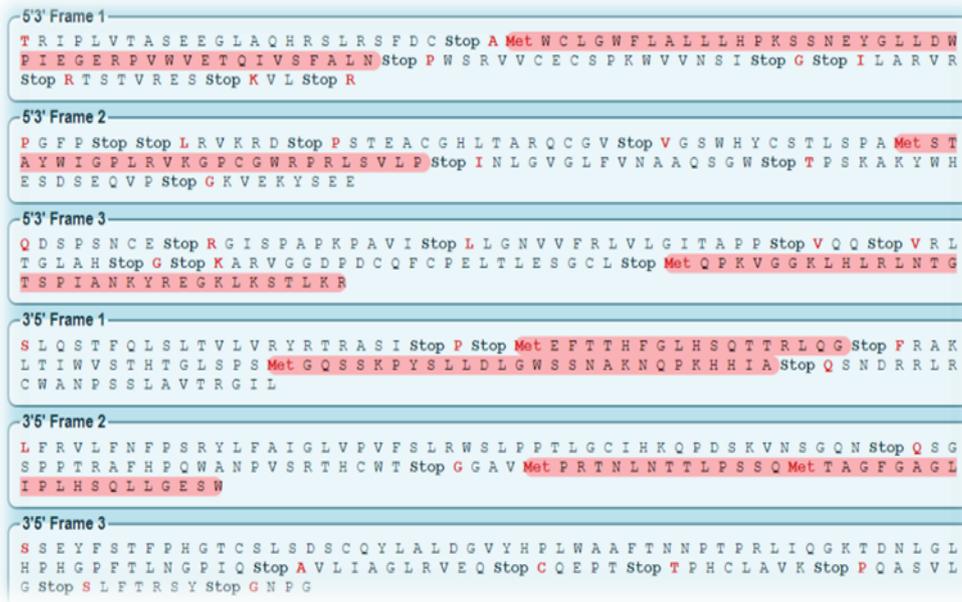
Fig. 1. *Frames of translated protein sequence.*

Fig. 2. *Mass spectrometry results of peptide sequence.*

Fig. 3. *Amino acid composition of peptide sequence of C. piscidium.*

Prediction of physiochemical parameters

The analysis of physicochemical parameters of the amino acid sequence indicated that the molecular weight was 12013.90 Daltons. Six translated frames of the DNA sequence were obtained. Further Mass spectrometry analysis(Fig.2)predicted the monoisotopic mass and molecular weight to be 12006.19 Da and 12013.90 Da. The isoelectric point was found to be 7.89, which categorizes the proteins to be basic in nature.The theoretical amino acid composition showed that total number of negatively charged residues were 11, positively charged residues were 12. Overall atomic composition(Fig.3) revealed that there are 541 carbon atoms,848 hydrogen atoms,150 nitrogen, 150 oxygen and 05 atoms of sulphur. The chemical formula of the predicted protein is $C_{541}H_{848}N_{150}O_{150}S_5$. The aliphatic was observed to be 99.29. A higher value of aliphatic index depicts that the proteins are stable at higher tem-
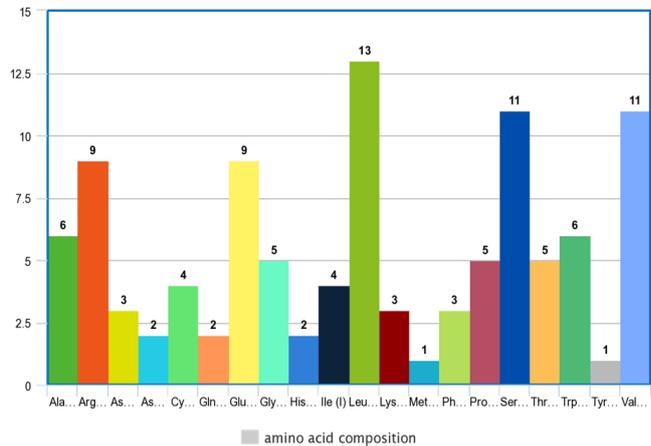
peratures, thereby allowing the parasite to adapt in extreme conditions, making the parasite distribution and diversity wide, thus making it keener to evolve. The grand average of hydropathy (GRAVY) revealed the protein interaction with water, which occurs better with low GRAVY.GRAVY was estimated to be 0.010 and categorized the proteins to be polar in nature. Besides this, the polarity of proteins plays a vital role in the generation and organization of apical surface protrusions (Apodaca,2017), which the parasite might use to establish itself in the host. The Hydropathy index served to be helpful in determining the three-dimensional structures of the proteins and these protein structures, in turn, determine the functionality of the proteins. For determining the instability investigation, the instability index was estimated to be 50.73 and categorized the proteins to be unstable in nature. The extinction coefficient has also been determined using all pairs of Cys

residues that form cystines assuming all Cys residues are reduced, i.e., 34740 and 34490, respectively. The half life of the predicted protein was found to be 5.5 hrs in the mammalian system ,3 min in *Yeast* and 2 min in *Escherichia coli*.

Amino acid composition

The amino acid composition was observed to be as follows:

Total number of negatively charged residues (Asp + Glu): 11

Total number of positively charged residues (Arg + Lys): 12

Atomic composition

Carbon	C	541
Hydrogen	H	848
NitrogenN		150
Oxygen O		150
Sulfur	S	5

Formula:$C_{541}H_{848}N_{150}O_{150}S$

Protein statistics and transmembrane orientations

The predicted protein sequence showed the presence of high scoring hydrophobic segments, which had enabled to predict the transmembrane topology of the proteins, i.e. prediction of what parts of it protrude into the cell and what protrudes out, and how many times the protein chain crossed the membrane that showed the presence of two inside to outside spanning regions and thus proved that the proteins are integral as they were with membrane-spanning regions and thus helpful in cell signalling and exchange gateway between cell to cell.

The peptide cutter software predicted the potential cleavage sites by proteases/chemicals(Fig.3) in the protein sequence. It was found that minimum cleavage was performed by CNBr (01), and maximum cleavage was made by proteinase K (58). The data has also been presented using restriction enzymes in the form of graphical representation (Fig.4).

Charge distributional analysis

The charge distribution analysis of the selected segment further revealed the charge distributional analysis and the following results were obtained -

0+0000000- -00000+00+ 00-0000000 000000000+ 000 -0000-0   00-0-+0000   -000000000   000+000-00 0+00000000 00+0++0000 +-0+00+

+ sign indicates the presence of positively charged amino acid.

- sign indicates the presence of negatively charged amino acid .

0 indicates the presence of uncharged amino acid.

Possible transmembrane orientation of peptide sequence

Two possible transmembrane orientations(Fig.5) were found for the present peptide sequence with positions in their bracket –

Inside to outside helices : 2 found

from	to	score	center
2 (6)	25 (25)	89	15
102 (104)	122 (120)	1486	112

Outside to inside helices : 2 found from   to   score center

Name of enzyme	No. of cleavages	Positions of cleavage sites
Arg-C proteinase	9	13 16 54 72 91 93 94 99 105
Asp-N endopeptidase	2	18 46
Asp-N endopeptidase + N-terminal Glu	11	5 6 18 22 41 46 50 52 58 75 99
BNPS-Skatole	6	25 29 48 57 70 81
CNBr	1	22
Chymotrypsin-high specificity (C-term to [FYW], not before P)	9	18 25 29 30 43 57 65 70 81
Chymotrypsin-low specificity (C-term to [FYWML], not before P)	24	1 9 12 15 18 22 25 27 29 30 31 33 34 35 43 45 46 57 65 67 70 81 89 104
Clostripain	9	13 16 54 72 91 93 94 99 105
Formic acid	2	19 47
Glutamyl endopeptidase	9	6 7 23 42 51 53 59 76 100
Iodosobenzoic acid	6	25 29 48 57 70 81
LysC	3	38 80 102
LysN	3	37 79 101
NTCB (2-nitro-5-thiocyanobenzoic acid)	4	19 25 74 76
Pepsin (pH1.3)	21	1 8 9 17 26 27 29 30 31 32 33 34 44 45 46 64 65 66 88 89 103
Pepsin (pH>2)	30	1 8 9 17 24 25 26 27 28 29 30 31 32 33 34 42 43 44 45 46 48 57 64 65 66 69 81 88 89 103
Proline-endopeptidase [*]	2	37 55
Proteinase K	58	1 2 3 4 6 7 9 10 15 18 21 23 24 25 27 29 30 31 32 33 34 35 42 43 45 46 48 50 51 53 56 57 58 59 60 62 63 65 66 67 70 73 74 76 81 82 83 86 88 89 90 92 95 97 98 100 103 104
Staphylococcal peptidase I	8	6 23 42 51 53 59 76 100
Thermolysin	37	1 3 8 9 14 17 20 21 26 29 30 31 32 33 44 45 49 55 57 61 62 64 65 66 72 73 81 82 85 87 88 89 91 97 102 103
Trypsin	11	13 16 38 72 80 91 93 94 99 102 105

Fig. 4. *Cleavage sites for different proteases and chemicals in protein sequence.*

101 (101) 117 (117)   1013   109
137 (137) 159 (157)   275   147
inside-> outside helices correspond
outside->inside helices.
 Helices shown in brackets are considered insignificant.
 A "+" symbol indicates a preference of this orientation.
 A "++" symbol indicates a strong preference of this orientation.
inside->outside | outside->inside

Phylo-proteomic analysis

Phyloproteomic analysis(Fig.6)revealed that the present protein of interest showed close similarity with *Schistosomamansoni*and*Trichinellapseudospiralis.Clinostomumpiscidium* voucher HS.Meta/2009/09 large subunit ribosomal RNA gene, partial sequence

clustered out at separate node in the form of the separate clade.

Three dimensional(3D) structure of proteins

Four identical proteins structure was predicted using Homology modelling method (Fig.7-10). One uncharacterized protein, A0A0V0XCR7_TRIPS A0A0V0XCR7 had 75.4% identity with *Trichinellapseudospiralis*, a viviparous nematode parasite, reported from rodents, bears, hyenas, and humans. 2mbr.1.A was used as a template to construct the 3 D structure of the protein. The sequence identity was 13.33% and the role is associated with UridineDiphospho-N-Aceylenolpyruvyl glucosamine Reductase. The main role of this protein is extensively involved in intracellular signaling as substrate for O - linked N - acetylglucosaminetransferases
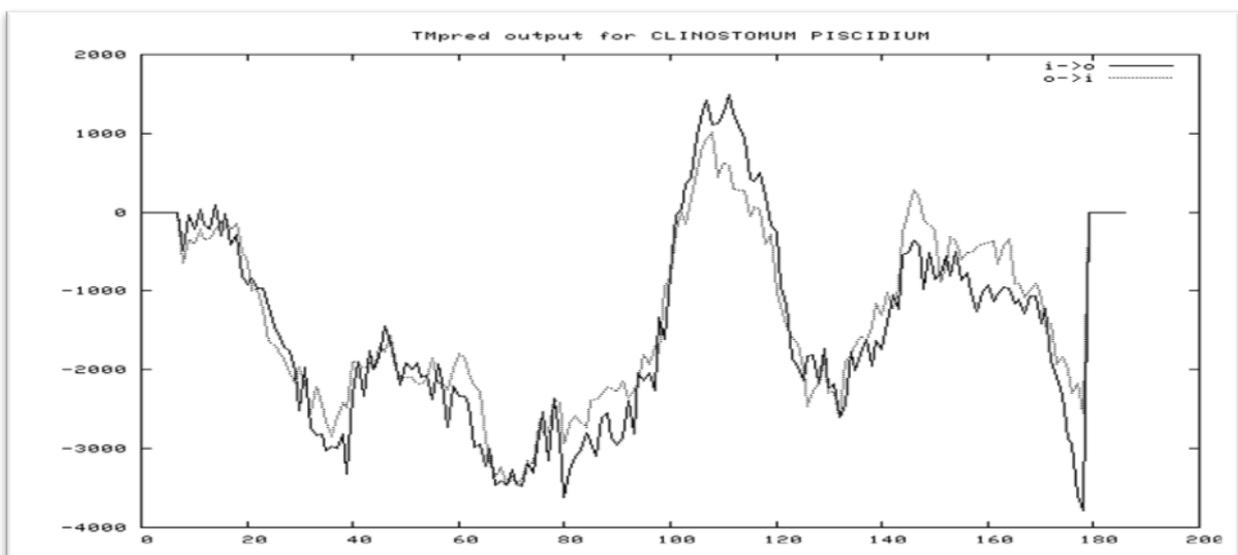

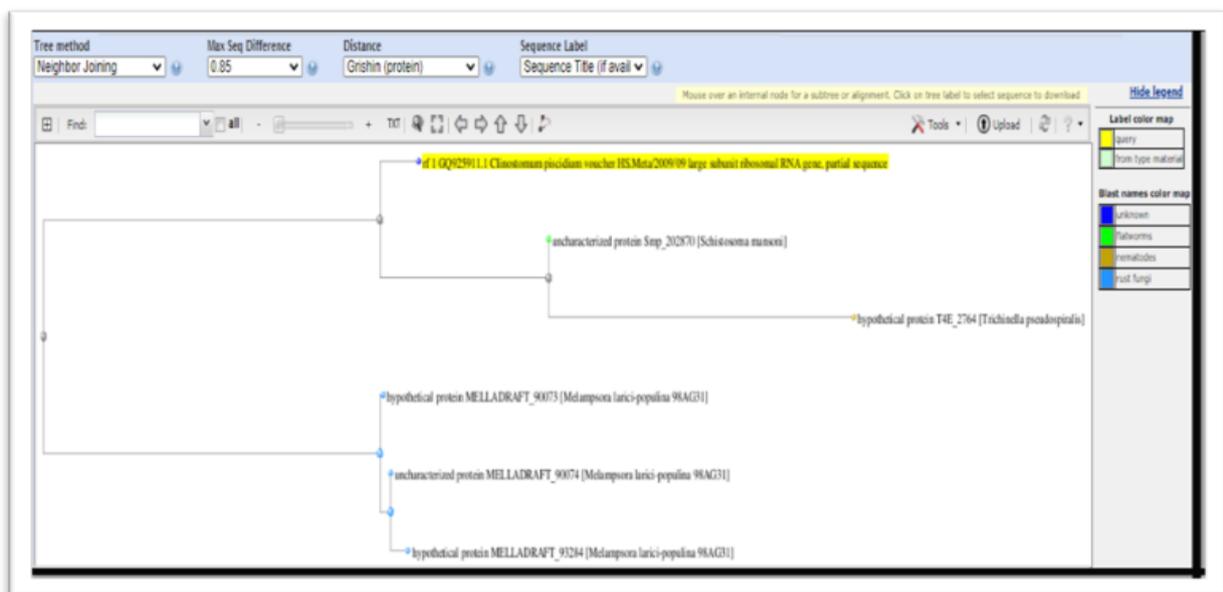**Fig. 5.** *Transmembrane orientation prediction of the peptide sequence.*

Fig. 6. *Phyloproteomic Tree analysis of desired protein sequence using neighbour joining method with max seq difference of 0.85.*
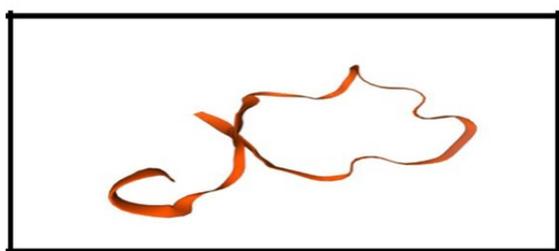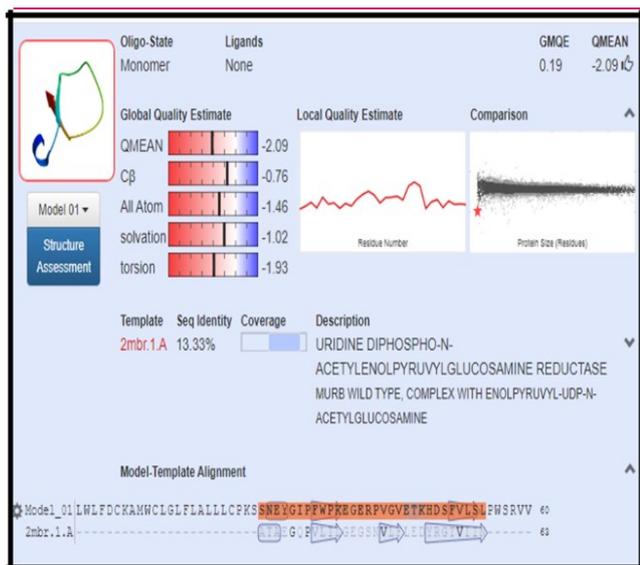
Fig. 7. *Structure prediction of Uncharacterized hypothetical protein A0A0V0XCR7_TRIPSA0A0V0XCR7.*
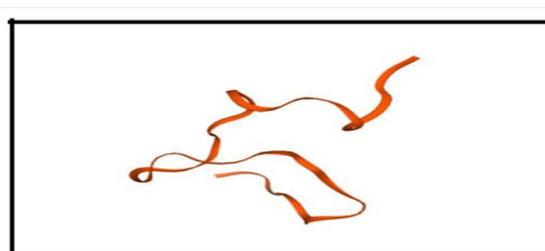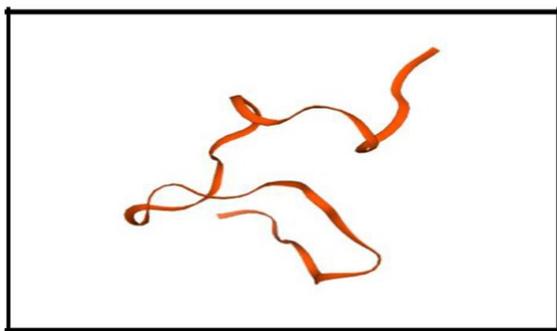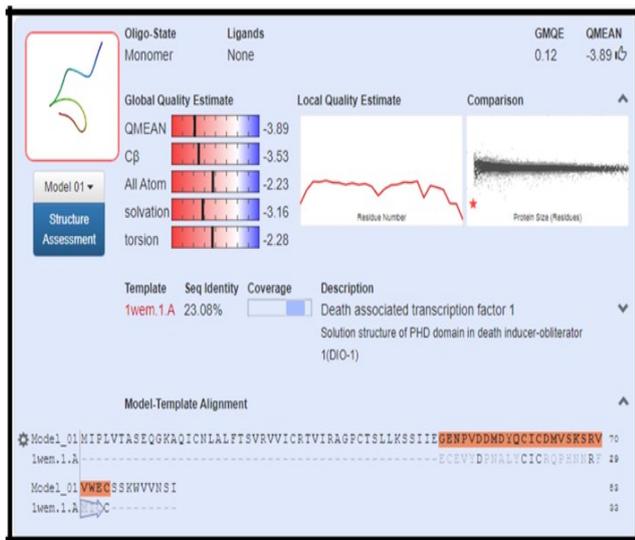

Fig. 8. *Structure prediction of Uncharacterized hypothetical protein F4R8R4_MELLP F4R8R4.*

(OGTs) to install the O- GlcAc post-translational modification in a wide range of species.

Another three proteins models identified were from *Melampsoralaricipopulina* with 43.2% identity.The proteins have a description of Death associated transcription factor1 (structure of PHD domain in death inducer – obliterator -1, DIO -1)(Rojas *et al*.,2005). *Melampsora* is a genus of *Basidiomycota f*ungi and is the most devastating and widespread pathogen of poplars and has limited the use of poplars for environmental and wood production in many parts of the world. Death inducer obliterator protein 1 [DIDO1; also termed DIO-1 and death-associated transcription factor 1 (DATF-1)] is encoded by a gene; thus, far described only in higher vertebrates. Current gene ontology descriptions for this gene assign its function to an apoptosis-related process. The protein presents distinct splice variants and is distributed ubiquitously. Exhaustive sequence analyses of all DIDO variants identify distant homologues in yeast and other organisms(Rojas *et al*.,2005). These homologues have a role in DNA regulation and chromatin stability and form part of higher complexes linked to active chromatin. Further domain composition analyses performed in the context of related homologues suggest that DIDO induced apoptosis is a secondary effect.

Conclusion

In the present study, in silico approaches predicted the function of hypothetical proteins from the *C. piscidium* genome. The prediction of translated protein sequence from rDNA sequence, physicochemical parameters, mass spectrometry, potential restriction sites within the sequence, compositional analysis and possible transmembrane orientations were useful to reinforce the understanding of the particular characteristics of the proteins annotated. The study identified the presence of proteins that play important roles in the mechanisms of adaptation to adverse environments. The theoretical proteins were unstable, which proved that the thermodynamic equilibrium with the host environment (micro milieu) was still not maintained, so they were unstable. Aliphatic index was computed to be high and a higher value of the aliphatic index depicted that the proteins were stable at higher temperatures. Such stability at higher temperature allowed the parasite to adapt in extreme conditions, making the parasite distribution and diversity wide and keener to evolve. The predicted proteins exhibited a low value of Grand average of hydropathicity, i.e., 0.010, proving that the theoretical proteins were polar in nature and this polarity of proteins plays a vital role in the generation and organization of

Fig. 9. *Structure prediction of uncharacterized hypothetical protein F4RVL4_MELLP F4RVL4.*

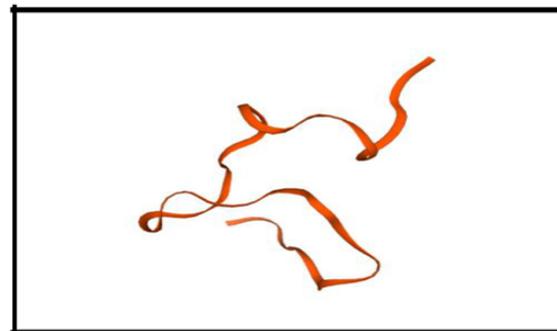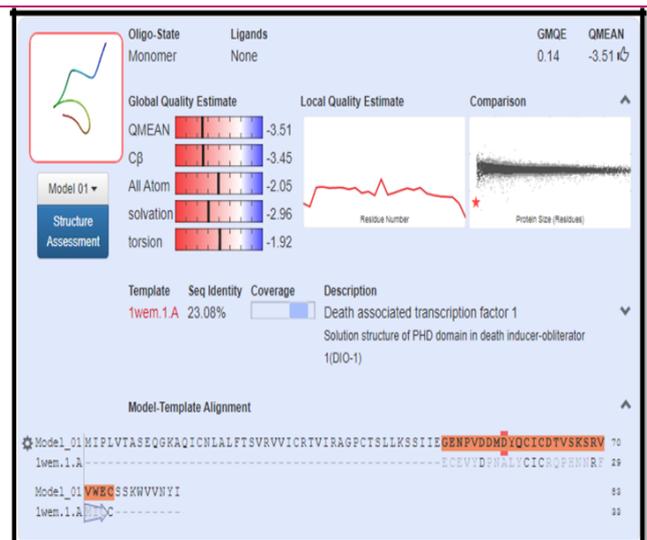**Fig. 10.** *Structure prediction of uncharacterized hypothetical protein F4S4M8_MELLP F4S4M8.*

apical surface protrusions, which the parasite might use to establish itself in the host. The half-life of the predicted protein was found to be 5.5 hrs in the mammalian system,3 min in the yeast and 2 min in *E. coli*. By determining the possible transmembrane orientations, it was found that the proteins were integral as they were with membrane spanning regions. Proteins were found to be integral proteins that are helpful in cell signalling and exchange gateway between cell to cell. Since the present proteins showed similarities with *T. pseudospiralis* and *M. aricipopulina* drug designing for the present helminth would also play a significant role in treating infections from both *T. pseudospiralis* and *M. laricipopulina*. Our findings open possibilities for better investigation of this zoonotic parasite for disease diagnostic in the field of parasitology.

Conflict of interest

The authors declare that they have no conflict of interest.

REFERENCES

1. Apodaca,Gerard.(2017). Role of Polarity Proteins in the Generation and Organization of Apical Surface Protrusions. *Journal of Cold Spring Harbor Perspective in Biology*,13(3),46-57.

2. Bloch M.E.,& Schneider J.G. (1801). Systema Ichthyologiae iconibus CX illustratum. Post obitum auctoris opus inchoatum absolvit, correxit, interpolavit Jo. Gottlob Schneider, Saxo. Berolini. Sumtibus Austoris Impressum et Bibliopolio Sanderiano Commissum. *Systema Ichthyol.,* 10(1),12-20

3. Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D & Bairoch A. *(2009).ExPASy. the proteomics server for in-depth protein knowledge and analysis Journal of Nucleic Acids Research*, 31,3784-3788.

4. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A & John M. Walker *(2005.). Protein Identification and Analysis Tools on the ExPASy Server.In: Walker, J.M.(Ed.). The Proteomics Protocols Handbook*, Totowa Humana Press , 571-607.http://archiveouverte.unige.ch/unige:37793

5. Hara,H., Miyauchi,Y; Tahara,S & Yamashita,H. (2014).Human laryngitis caused by Clinostomum complanatum Nagoya. *Journal.Medical Science*,76,181-185

6. Menezes,Fraga Alana; Junior,Cavalcanti; Martis,Semen& Filho,Moraes.(2019). In Silico Characterization of Meloidogyne Genus Nematode Cellulose Binding Proteins.Journal of *Brazilian Archives of Biology and Technology,* 62,67-75

7. Park, C.W; Kim, J.S; Joo, H.S & Kim,J. (2009). A human case of *Clinostomum complanatum* infection in Korea. *Korean J. Parasitol.*, 47, 401–404

8. Rojas,M.Ana., Pulido,L.Luis., Futterer, Agnes., Weley,van

Karel., Martinez, Carlos & Valencia Alfonso (2005).Death inducer obliterator protein 1 in the context of DNA regulation Sequence analyses of distant homologues point to a novel functional role. *Federation of European Biochemical Societies*,27,3505-3511

9. Schmidt,G. D. & Roberts,L. S.(2000).Foundations of Parasitology .Boston: McGraw-Hill .

10. Shareef,Ahammed P.A.& Abidi,S. M. A.(2012). Incidence and histopathology of encysted progenetic metacercaria of *Clinostomum complanatum* (Digenea: Clinostomidae) in *Channa punctatus* and its development in experimental host. *Tropical Biomedicine*,2(6),421-426

11. Southwell,T & Prashad,B.(1918). Notes from the Bengal fisheries laboratory 5. Parasites of Indian fishes, with a note on carcinoma in the climbing perch. *Rec. Ind. Mus.*,15, 341-355

12. Tiewchaloern, S, Udomkijdecha, S,Suvouttho, S, Chunchamsri, K & Waikagul, J. (1999). Clinostomum trematode from human eye. *Southeast Asian J. Trop. Med. Public Health*, 30, 382–384.

13. Zimik, Philayung., Sharma, Sunil & Roy, Bishnupada (2019). Characterization of *Clinostomum* metacercariae using microscopic and molecular approaches. *Annals of Parasitology,* 65, 87-97.