PHYLOGENETIC PREDICTION AND *IN SILICO* COMPARATIVE MODELING OF *IBR DOMAIN PROTEIN* (*OESOPHAGOSTOMUM DENTATUM*) RESPONSIBLE FOR *OESOPHAGOSTOMIASIS*

Santosh Kumar[1]*, Sunena Bamoriya[2], Varsha Chauhan[3], Nem Kumar Jain[4], Motilal[5], Akshay Singh Sengar[6], and O. P. Verma[7]

Abstract:
The roundworms that make up the *Strongylidae* family of parasitic nematodes are members of the genus *Oesophagostomum*. These worms can be found in Indonesia, the Philippines, Brazil, Africa, and China. Human infections caused by *Oesophagostomum* are most common in Ghana and northern Togo. Because the eggs of the species that cause human helminthomas may be indistinguishable from those of hookworms, which are common but rarely cause helminthomas, it is difficult to accurately identify the species that cause human helminthomas. Oesophagostomum, especially *O. bifurcum*, is a common parasite that lives in livestock and non-human primates like humans, pigs, and goats. Despite the fact that the pathways through which humans become infected have not been sufficiently elucidated, it is generally accepted that infected individuals consume soil containing the infectious filariform larvae without realizing it. Oesophagostomum infection is a serious problem for public health that mostly affects northern Togo and Ghana in western Africa. Due to its localization, there has been a dearth of research on effective public health interventions and intervention strategies.
Physiochemical characterization of protein isolated from *Oesophagostomum dentatum* responsible for *Oesophagostomiasis* disease with 204 amino acids was done to interpret properties like pI, EC, AI, GRAVY, and instability index. In silico comparative modeling was performed to generate good-quality models. The assessment of generated three-dimensional structure against structure verification tools PROCHECK showed that model generated by Geno3D was acceptable as compared to SWISS-MODEL. The predicted model can be used in structure-based drug designing and vaccine development.

Keywords: *Oesophagostomiasis, Oesophagostomum dentatum, in silico, PROCHECK, GENO3D*

[1*,3,5]Asst. Professor, Department of Life Sciences, ITM University, Gwalior, M.P, India
[2]Student, Department of Life Sciences, ITM University, Gwalior, M.P, India
[4]Asst. Professor, School of Pharmacy, ITM University, Gwalior, M.P, India
[6]Asst. Professor, Department of Food Technology, ITM University, Gwalior, M.P, India
[7]Assoc. Professor, Jacob School of Biotechnology and Bioengineering, SHUATS, Prayagraj, UP

***Corresponding Author**: Santosh Kumar
*****Asst. Professor, Department of Life Sciences, ITM University, Gwalior, M.P, India
Email: santoshmtech73@gmail.com

DOI: - 10.48047/ecb/2023.12.si10.00505

I. INTRODUCTION

Oesophagostomum is the phylum that includes nematodes. This phylum is composed of five orders: The superfamilies Ancyclostomatoidea, Trichostrongyloidea, and Strongylidea comprise Strongylida. These superfamilies include Ascaridida, Enoplida, Oxyurida, Spirurida, Strongylida, and Rhabditida[1]. The first case of infection with Oesophagostomum spp. was a member of the Strongyloidae group[2]. It was first described by Railliet and Henry in 1905, who described the parasites that were found in a male's tumors of the caecum and colon from the Omo River in southern Ethiopia[3]. The second case of Oesophagostomum stephanostomum was reported as having macroscopical and microscopic pathology by H. Wolferstan Thomas in 1910. His descriptions were founded on what he saw at the postmortem of a Brazilian man who had perished from severe dysentery and was infected[4]. There is no standard clinical picture for the symptoms of oesophageal dysbiosis. However, most patients complain of pain in the lower right quadrant and one or more protruding abdominal masses[5]. Larvae can enter the colon wall in oesophageal stasis, potentially causing two distinct types of nodular pathology. Multinodular disease is characterized by the formation of numerous tiny nodular lesions containing pus and worms along the wall of the colon. About 15% of patients have this kind of oesophagitis[6]. This knob can cause extreme tissue responses, bringing about excruciating, projecting masses. Disease, a ruptured appendix, amebiasis, and tuberculosis are all often misdiagnosed[7]. Most of the time, oesophageal parasitism is thought to be a zoonotic disease, which means that it can spread from one animal to another and from one person to another. Oesophagostomum infects cattle, sheep, goats, wild pigs, and other primates and is mostly carried by non-human animals. Oesophagostomum are primarily carried by non-humans, infecting cattle, sheep, and goats. Humans are generally considered accidental hosts due to their inability to complete the Oesophagostomum development[8]. Oesophagostomum transmission is thought to be oral-fecal for both humans and animals[9]. The IBR domain, which stands for "In Between Ring Fingers," frequently appears to be located in the space between two pairs of ring fingers. On the other hand, the extreme localization of oesophageal parasitism to northern Togo and Ghana in Africa raises the possibility that Oesophagostomum is increasingly showing a preference for human hosts[10]. Additionally, this domain has been referred to as the C6HC domain and the DRIL domain, which stands for "double ring finger linked"[11]. It is accepted that all eukaryotic creatures contain proteins with two ring fingers and an IBR space. The RBR family of proteins is another name for these proteins. Members of the RBR family are involved in protein quality control and can indirectly control transcription.

II. MATERIALS AND METHODS

2.1. Retrieval of target sequence: The amino acid sequence of the IBR Domain Protein of Oesophagostomum dentatum was obtained from the sequence database of NCBI (http://www. ncbi.nlm.nih.gov/protein/ KHJ98692.1).It was ascertained that the three-dimensional structure of the protein was not available in Protein Data Bank[12].Hence the present work for developing the 3D model of the IBR Domain Protein of Oesophagostomum dentatum was undertaken. The protein is 204 amino acids in length.

2.2. Physio-chemical characterization: The values of theoretical isoelectric point (pI), molecular weight, and a total number of positive and negative residues, extinction coefficient[13], instability index [14], aliphatic index[15] and grand average hydropathy (GRAVY)[16] were computed. physio-chemical characterization was done using Expasy's ProtParam server[17] (http://us.expasy. org/tools/protparam.html).

2.3. Secondary structure prediction: Secondary structure has been predicted using PHYRE2 (http://www.sbg.bio.ic.ac.uk/phyre2) software where the FASTA format of the sequence was given as input SOPMA[18] was employed for calculating the secondary structural features of the protein sequence considered for this study.

2.4. Model building and quality assessment: The modeling of the three-dimensional structure of the protein was done using two homology modeling programs, Geno3D[19] and the Swiss Model[20]. The overall stereo-chemical property of the protein was assessed by Ramachandran plot analysis [21]. The evaluation of structure models obtained from the two software tools was performed by using PROCHECK[22]. The models were further checked with WHATIF for standard bond length and bond angle determination. The results of PROCHECK were depicted in Table 3.

III. RESULTS AND DISCUSSION

The protein sequence of the matrix protein of the IBR Domain Protein (O*esophagostomum dentatum*) was retrieved in FASTA format from the NCBI Entrez sequence search and used as the query sequence for homology modeling in the current study. Table 1 shows the physiochemical parameters that were calculated using Expasy's

ProtParam tool. A protein with an instability index lower than 40 is predicted to be stable, while a protein with an instability index higher than 40 may be unstable[7]. Protein's 41.76 stability index indicates its stability. Cys, Trp, and Tyr amino acids are present in high concentrations, as evidenced by the high extinction coefficient (25855). The increase in thermal stability is thought to be helped by the aliphatic index. The query protein's high aliphatic index (66.81) suggests that it may be stable over a wide temperature range. The fact that the Grand Average hydropathy (GRAVY) value is low (-0.559) suggests that the interaction with water may be improved.

Table 1: Parameters computed using Expasy's ProtParam tool

S.No.	Property	Value
1.	Number of amino acids	204
2.	Molecular weight	23083.79
3.	Theoretical pl	4.57
4.	Total number of negatively charged residues (Asp+Glu)	38
5.	Total number of positively residues (Arg+Lys)	23
6.	Extinction coefficient	25855
7.	Extinction coefficient*	24980
8.	Instability index	41.76
9.	Aliphatic index	66.81
10.	Grand average of hydropathicity	-0.559

The secondary structure of IBR Domain Protein (*oesophagostomum dentatum*) was predicted by two software's namely SOPMA (Self Optimized Prediction Method with Alignment) and Phyre 2. SOPMA predicts 69.5% of amino acids correctly to describe secondary structure[18] (Table 2).These results show a higher number of random coils in comparison to other secondary structure elements (alpha helix, extended strand, and beta turns). Secondary structure and disorder prediction was made using pyre 2 which is shown in figure 6.

Table 2: Calculated secondary structure elements by SOPMA

S.No.	Parameters	Value(%)
1.	Alpha helix	29.41%
2.	310helix	0.00%
3.	Pi helix	0.00%
4.	Beta bridge	0.00%
5.	Extended strand	17.65%
6.	Beta turn	3.92%
7.	Bend region	0.00%
8.	Random coil	49.02%
9.	Ambiguous state	0.00%
10.	Other state	0.00%

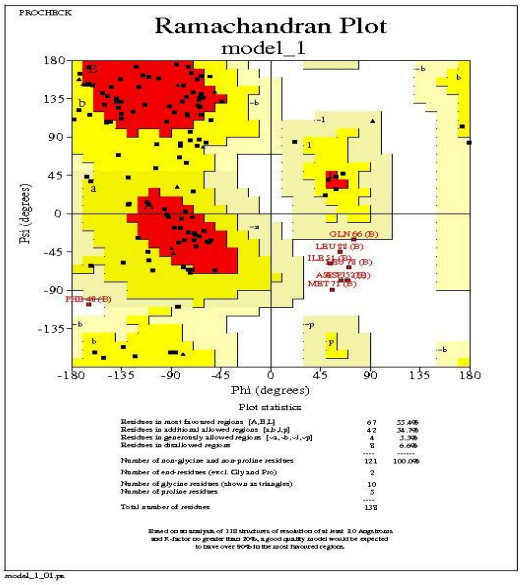
Three-dimensional structures of proteins were predicted due to the unavailability of such data. There is no experimental structure found for the protein considered. The homology modeling of the protein was done using two programs Geno3d and Swiss Model. The results obtained from these programs were compared in Table 3. This comparison suggests that the model generated by the Swiss model is much better than the model generated by Geno3d (Fig. 7). Finally the chosen model was visualized by Rasmol (Figure 3). The evaluation of the predicted structure generated by the Swiss Model for the stereo-chemical quality was done using Ramachandran map calculations validated with the PROCHECK (Figure 1-4). 88.3% of residues were found in the core right-handed alpha helices(A), beta sheets (B), and left-handed alpha helix(L) region.10.7% of residues were found in the allowed right-handed alpha helix(a),beta sheets(b) and left-handed alpha helices regions. 0.7% of the residues were found in the generously allowed alpha helices (~a), beta sheets (~b), left-handed alpha helices (~l), and epsilon (~p) regions. 0.3% of the residues were found to be localized at the disallowed regions. The above results indicate a good quality of the predicted model. The analysis of the modeled structure of the IBR Domain Protein of

Oesophagostomum dentatum by the structure verification server WHATIF[23] revealed RMS Z-score (0.965) which was nearly meeting the required score for a high-quality model.

Table 3: Ramachandran plot calculation and Comparative analysis of the models from Geno3D and Swiss-model computed with the PROCHECK program

Server	Parameters	Value%
Geno 3D	Residue in the most favored region	55.4%
	Residue in additionally allowed region	34.7%
	Residue in generally allowed region	3.3%
	Residue in the disallowed region	6.6%

Ramachandran Plot generated by Geno 3D



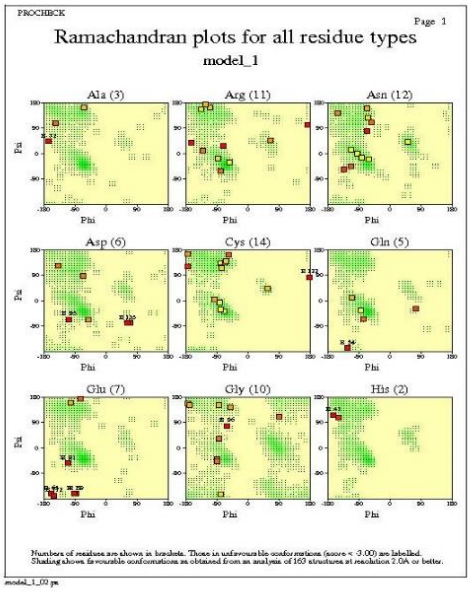


Fig 1: Ramachandran plot for all residues

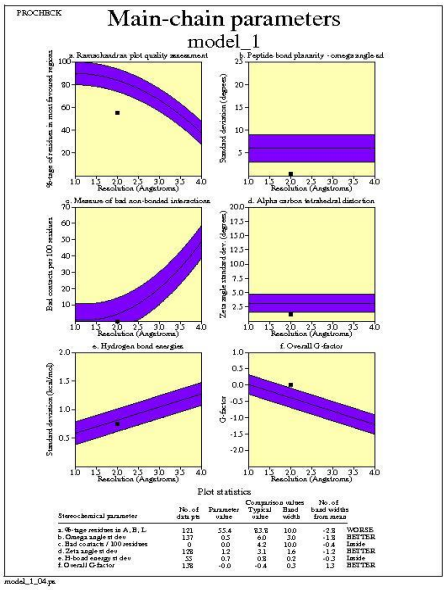


Fig 2: Main chain parameters

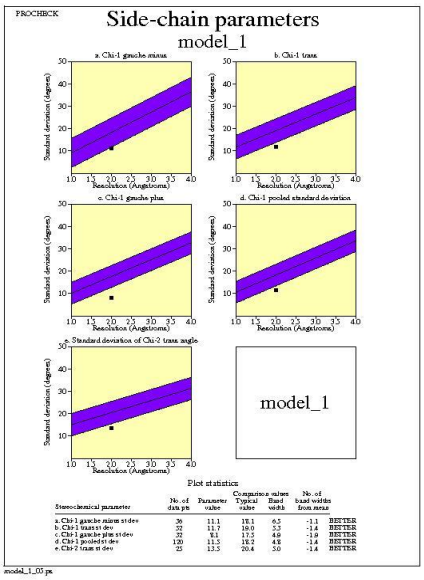


Fig 3: Graph for side chain residues

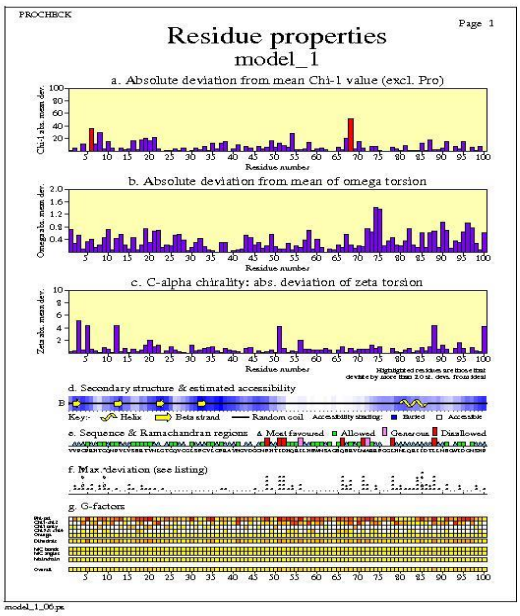


Fig 4: Graph for residual properties

SOPMA RESULT:

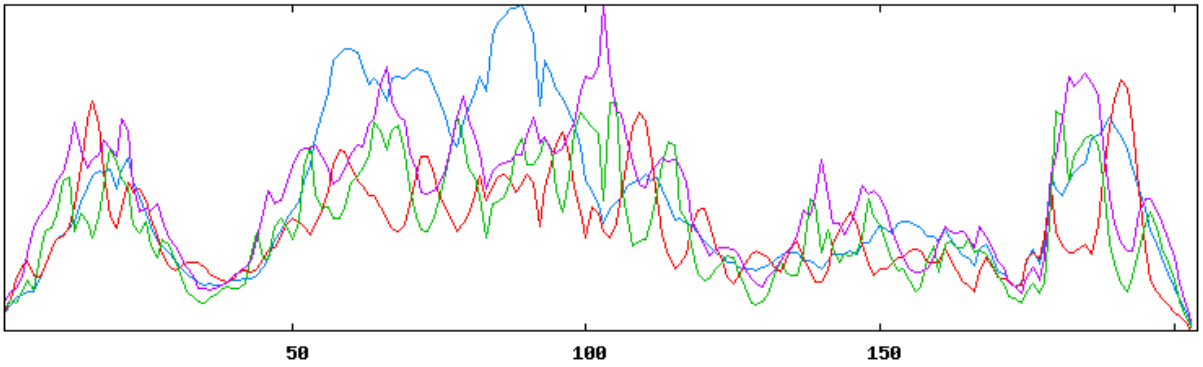


Fig. 5: SOPMA results showing deviations, coils, helix and sheet

PHYRE 2 RESULT: -

Fig. 6: Secondary structure generated by PHYRE 2 software

GENO3D RESULT: -

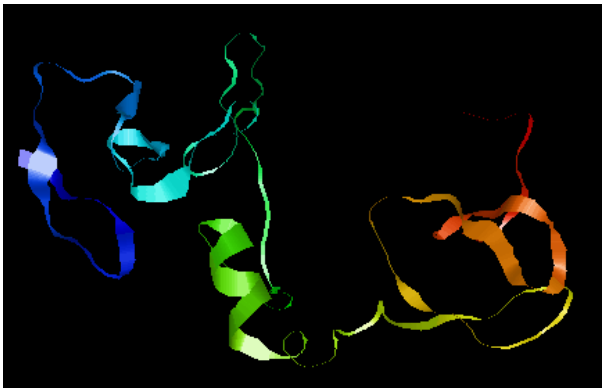


Fig. 7: Tertiary structure generated by Geno3D software.

IV. CONCLUSION

The predicted three-dimensional structure of the IBR Domain protein of *Oesophagostomum dentatum* is stable based on the evaluation of various structural and physiochemical parameters. The data from the Ramachandran plot indicated that the core region contains nearly 55.4% of the residues in each protein. The best stereo chemical quality of the protein structure is represented by the maximum percentage of residues in the core region, which is further supported by nearly 55.4% of residues in the most preferred and permitted regions. Only 6.6% of residues are found in the disallowed regions, which typically involve steric inhibition between the main chain atoms and the side chain C-beta methylene group. The generated structures exhibit the highest stereo chemical quality and stability due to the significantly lower proportion of residues found in the disallowed region than in the core and allowed regions. The best optimized and superimposed model was thought to be the three-dimensional structures created using the SWISS Model software and validated by the PROCHECK software, which displayed the highest identity, lowest energy in Kcal/mol, and lowest root mean square deviation (RMSD). Due to the fact that the IBR Domain Protein of *Oesophagostomum dentatum* does not have a vaccine or effective treatment. This model's structural information can be effectively utilized and incorporated into subsequent drug design.

V. ACKNOWLEDGEMENT

We would like to acknowledge ITM University, Gwalior for its timely support directly or indirectly in making this research a success. We would also like to thank the software without whose support this work would have not been completed.

VI. REFERENCES

1. John, David T., Krotoski, Wojciech A. and Markell, Edward K. Markell and Voge's Medical Parasitology. St. Louis: Saunders Elsevier, 2006. 9th Edition.
2. Wooten, Darcy. *Oesophagostomum.* 5 Feb 2009.
3. Polderman, A. M.; Gigase, P.; Krepel, H. P.; Blotkamp, J.; Baeta, S. (1991). "Oesophagostomiasis, a common infection of man in northern Togo and Ghana". *American Journal of Tropical Medicine and Hygiene*. 44 (3):336–344. doi:10.4269/ ajtmh. 1991.44.336. PMID 2035755.
4. Thomas, H.W. (1910). The pathological report of a case of oesophagostomiasis in man. Annals of Tropical Medicine and Parasitology. 4: 57-88.
5. Ash, Lawrence R. and Orihel, Thomas C. Parasites in Human Tissues. Chicago: ASCP Press, 1995.
6. *GIDEON Infectious Diseases – Diseases (oesophagostomiasis).* GIDEON Infectious Disease Database. 5 Feb 2009.
7. Sun, Tsieh. Parasitic Disorders: Pathology, Diagnosis, and Management. Baltimore: William and Wilkins, 1999. 2nd ed.
8. Ziem, J.B. *Controlling human oesophago-stomiasis in northern Ghana.* (Doctoral thesis) Leiden University. 2006.
9. Gasser, R. B.; de Gruijter, J. M.; Polderman, A. M. (April 2006). "Insights into the epidemiology and genetic make-up of *Oesophagostomum bifurcum* from human and non-human primates using molecular tools".
10. TRIADs: a new class of proteins with a novel cysteine-rich signature. van der Reijden BA, Erpelinck-Verschueren CA, Lowenberg B, Jansen JH. (1999), *Protein Science* 8, 1557-61.
11. View article PMID: 10422847. Parkin and relatives: the RBR family of ubiquitin ligases. Marin I, Lucas JI, Gradilla AC, Ferrus

A. *Physiol.Genomics* 17, 25363, (2004). Vie w article PMID: 15152079.
12. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE (2000), The Protein Data Bank. Nucleic Acids Res, 28 (1): 235-42.
13. Gill SC, Von Hippel PH (1989) Extinction coefficient. Anal Biochem 182: 319-328.
14. Guruprasad K, Reddy BVP, Pandit MW (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. Prot Eng 4: 155-164.
15. Ikai AJ (1980) Thermo stability and aliphatic index of globular proteins. J Biochem 88: 1895- 1898.
16. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology 157: 105- 132.
17. Gasteiger E (2005) Protein Identification and Analysis Tools on the ExPASy Server. In: John M.Walker, The Proteomics Protocols Handbook, Humana Press 571-607.
18. Geourjon C, Deléage G (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Computational and Applied Bioscience 11: 681-684.
19. Combet C, Jambon M, Deleage G, Geourjon C (2002) Geno3D: Automatic comparative molecular modelling of protein, Bioinformatics 18: 213-214.
20. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22: 195-201.
21. Ramachandran GN, Ramakrishnan C, Sasisekhran V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7: 95-99.
22. Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. JBiomol NMR 8: 477-486.
23. Vriend G (1990) WHAT IF: A molecular modeling and drug design program. J Mol Graph 8: 52-56.