World Journal of Fish and Marine Sciences 5 (4): 426-429, 2013
ISSN 2078-4589

DOI: 10.5829/idosi.wjfms.2013.05.04.7427

Comparative *in Silico* Analysis of Mbl Homologues of Teleosts

[1]Chirag Goel, [1]Ashoktaru Barat, [2]Veena Pande and [1]P.K. Sahoo

[1]Molecular Genetics Laboratory, Directorate of Coldwater
Fisheries Research (ICAR), Bhimtal-263136, Nainital, Uttarakhand, India
[2]Department of Biotechnology, Kumaun University, Nainital, Uttarakhand, India

Abstract: Teleost is a large and extremely diverse group of ray-finned fishes. In this paper, a bioinformatics approach was adopted to explore properties and structure of teleost Mannose Binding Lectin MBL homologue proteins. The teleost species considered under present study were *Carassius auratus, Cyprinus carpio, Danio rerio, Oncorhynchus mykiss, Salmo salar* and *Schizothorax richardsonii*. Physicochemical characterization interprets properties such as theoretical isoelectric point (pI), extinction coefficient (EC), aliphatic index (AL), Grand average hydropathy (GRAVY) and instability index and provides data about these proteins and their properties. Prediction of motifs, patterns, disulfide bridges and secondary structure were performed for functional characterization. Computationally based characterization of the features of the proteins found or predicted in completely sequenced proteomes is an important task in the search for knowledge of protein function.

Key words: Teleost · *Insilico* · MBL Homologue

INTRODUCTION

The fish immune system has always been interesting from an evolutionary biological perspective, since fish evolved at the beginning of the vertebrate radiation. However, during the second half of the last century, the research on the topic has also increased exponentially with the aim of improving fish health care in order to secure aquaculture production in a growing market [1]. Within the fish classes, teleosts have been best studied mainly because of their importance in aquaculture, accumulation of fundamental biological knowledge of immunology, Genetics, development and physiology and also it is the most diversified group of all vertebrates comprising about half of the extant vertebrate species [2]. The innate immune system forms the first line of defense. It detects the entrance of foreign antigens, initiating a series of responses that, not successful at eliminating the infectious agent, will help to determine the type of adaptive immune response to be elicited. The complement is a major humoral system of innate immunity. In mammals, there are three distinct pathways of activating the complement system. The lectin pathway of complement is considered to be the most ancient and Ab-independent activation pathway of the complement system, relying on recognition of pathogen-associated molecular patterns by two distinct groups of lectins; Mannose Binding Lectin (MBL) and ficolins [3].

MBL is a C-type lectin which is composed of an N-terminal region rich in cysteine, a collagenous region characterized by tandem repeats of Gly-X-Y triplet sequences (X and Y stand for any amino acid) and a C-type lectin like carbohydrate recognition domain (CRD) [4]. Through its CRD MBL binds carbohydrates with 3 and 4- hydroxyl groups in the pyranose ring in the presence of Ca^{2+}. During infections, it enhances pathogen clearance, by opsonizing pathogens, activating complement through the lectin pathway and triggering phagocytosis via receptors on the phagocytic cells.

Computational tools provide researchers to understand physicochemical, structural properties of proteins. A large number of computational tools from different sources are available for making predictions regarding the identification and structure prediction of proteins. The major drawback of the experimental methods that have been used to characterize the proteins of various organisms are the time frame involved, high cost and the fact that these methods are not amenable to high

Corresponding Author: Chirag Goel, Molecular Genetics Laboratory, Directorate of Coldwater Fisheries Research (ICAR), Bhimtal-263136, Nainital, Uttarakhand, India. Mob: +91-9927623218.

throughput techniques. *In silico* approaches provide a viable solution to these problems. The amino acid sequence provides the information required to determine, the physicochemical characteristics, structural features and the molecular function of a protein. In the present study, comparative *in silico* analysis of the MBL homologues of teleosts were reported.

MATERIAL AND METHODS

Physicochemical Characterization: MBL homologues of teleosts considered under present study are shown in Table 1. The protein sequences were retrieved from the NCBI in FASTA format and used for further analysis. The physicochemical characteristics, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY) were determined using the Expaxy's ProtParam server. The results were shown in Table 2.

Functional Characterization: The SOSUI server performed the identification of transmembrane regions. Table 3 represents the transmembrane region identified for these MBL homologues proteins. Disulphide bonds are important in determining the functional linkages. Table 4 shows prediction of "SS" bonds using the primary structure (protein sequence data) by the tool CYS_REC (http://sunI. softberry.com/ berry.phtml?topic). CYS_REC identifies the position of cysteins, total number of cysteins present and pattern, if present, of pairs in the protein sequence. Secondary structure prediction SOPMA [5] was employed for calculating the secondary structural features of the MBL homologues protein sequences considered for this study. The results were presented in Table 5.

RESULTS AND DISSCUSION

MBL homologues of teleosts considered under present study are shown in Table 1. The protein sequences were retrieved from the NCBI in FASTA format and used for further analysis. The physicochemical parameters were computed using Expasy's Protparam tool (Table 2). The computed pI value of MBL homologues ranges from 4.97 to 6.22 (pI<7) revels that these proteins were basic in nature. The Isoelectric point (pI) is the pH at which the charge of protein is zero. The theoretical pI will be useful for developing buffer system for purification by isoelectric focusing method. As at pI, solubility is least and mobility in an electrofocusing system is zero. The extinction coefficient of MBL homologues at 280nm ranges from 18700 to 27305 M-1cm-1 with respect to the concentration of Cys, Trp and Tyr. These coefficients helps in the Quantitative study of Protein-Protein and protein-ligand interactions in solution. The instability index provides an estimate of the stability of protein in a test tube. There are certain dipeptides, the occurrence of which is significantly different in the unstable protein compared with those in the stable ones. This method assigns a weight value of instability. Using these weight values it is possible to compute instability index (II).

Table 1: Protein Sequences considered for the study.

Protein Name	Organism Name	Accession No.
Goldfish GBL	*Carassius auratus*	AF227739
Common carp MBL	*Cyprinus carpio*	BAD02477
Zebra fish GBL	*Danio rerio*	AF227738
Oncorhynchus mykiss H2	*Oncorhynchus mykiss*	NM_001160480
Oncorhynchus mykiss H3	*Oncorhynchus mykiss*	EU118768
Salmo salar	*Salmo salar*	ACI67500
Sch-GBL	*Schizothorax richardsonii*	KC620459
Common carp GBL	*Cyprinus carpio*	AF227737

Table 2: Parameters computed using Expasy's ProtParam tool.

Organism Name	Sequence length	Molecular weight	pI	-R	+R	EC	II	AI	GRAVY
Goldfish GBL	246	25708.8	5.21	31	26	18825	31.94	69.31	-0.503
Common carp MBL	245	26220.8	6.22	29	27	27305	24.58	77.55	-0.321
Zebra fish GBL	251	26829.6	5.63	32	30	20315	31.63	80.44	-0.355
Oncorhynchus mykiss H2	242	25783.5	5.22	29	25	22960	41.37	75.70	-0.334
Oncorhynchus mykiss H3	242	25686.3	4.97	29	23	22960	43.97	70.45	-0.342
Salmo salar	240	25468.1	5.97	26	24	22835	41.08	73.50	-0.342
Sch-GBL	254	26928.7	6.18	30	29	21680	31.77	70.98	-0.522
Common carp GBL	256	26934.3	5.36	31	27	18700	31.62	70.43	-0.442

Table 3: Transmembrane regions identified by SOSUI server

Organism Name	Type of Protein	Length	Transmembrane region
Goldfish GBL	Soluble	-	-
Common carp MBL	Transmembrane	22	LLMHRLCAALLLTEFVLLLAAA
Zebra fish GBL	Transmembrane	23	ALLKLFLGALLLLQLVLQLMAGA
Oncorhynchus mykiss H2	Transmembrane	21	MALSRPLTLGLCVLSLLVLPG
Oncorhynchus mykiss H3	Transmembrane	22	MALSRPLTLGLCVFSLLVLPGQ
Salmo salar	Transmembrane	22	MALSRLLTLGLCVLSLLVLPGQ
Sch-GBL	Transmembrane	23	ALFKLFLGTLLLLQFALQLLDGA
Common carp GBL	Transmembrane	23	ALFKLFLGTLLLLQFALQLLDGA

Table 4: Disulphide (SS) bond pattern of pairs predicted by CYS_REC

Organism Name	CYS_REC
Goldfish GBL	Cys21- Cys120
	Cys219- Cys233
Common carp MBL	Cys28- Cys221
	Cys235- Cys243
Zebra fish GBL	Cys32- Cys238
	Cys102- Cys246
	Cys154- Cys224
Oncorhynchus mykiss H2	Cys28- Cys240
	Cys76- Cys232
	Cys147- Cys217
Oncorhynchus mykiss H3	Cys26- Cys240
	Cys28- Cys232
	Cys76- Cys217
Salmo salar	Cys26- Cys230
	Cys215- Cys238
Sch-GBL	Cys31- Cys240
	Cys153- Cys226
Common carp GBL	Cys31- Cys229
	Cys156- Cys243

Table 5: Calculated secondary structure elements by SOPMA

	Goldfish GBL	*Oncorhynchus mykiss* H2	*Oncorhynchus mykiss* H3	*Salmo salar*	Common carp MBL	Zebra fish GBL	Sch-GBL	Common carp GBL
Alpha helix	26.02%	19.01%	21.49%	22.92%	31.02%	31.47%	29.92%	27.7%
3_{10} helix	0%	0%	0%	0%	0%	0%	0%	0%
Pi helix	0%	0%	0%	0%	0%	0%	0%	0%
Beta bridge	0%	0%	0%	0%	0%	0%	0%	0%
Extended strand	13.41%	17.77%	19.83%	19.17%	13.88%	11.95%	12.60%	13.28%
Beta turn	7.32%	4.13%	5.37%	5%	5.31%	3.98%	7.09%	6.64%
Bend region	0%	0%	0%	0%	0%	0%	0%	0%
Random coil	53.25%	59.09%	53.31%	52.92%	49.80%	52.59%	50.39%	52.34%
Ambiguous states	0%	0%	0%	0%	0%	0%	0%	0%
Other states	0%	0%	0%	0%	0%	0%	0%	0%

Guruprasad [6], suggested that a protein whose instability index is smaller than 40 is predicted as stable while a value above 40 predicts that the protein may be unstable. The instability index values for the MBL homologues were found to be ranging from 24.58 to 43.97. The results showed that except for MBL homologues from salmonids (*Salmo salar, Oncorhynchus mykiss*) all are stable ones.

The aliphatic index (AI) for the MBL homologues was found to be ranging between 69.31 to 80.44. The very high value of AI of all proteins under study suggests that these proteins are stable over a wide temperature range as the protein occupied by aliphatic side chains (A, V, I and L) is regarded as a positive factor for the increase of thermal stability of globular proteins. The grand average hydropathy (GRAVY) value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence. GRAVY indices of MBL homologues are ranging from -0.503 to -0.334. This low range of value indicates the possibility of better interaction with water.

Functional analysis of these proteins includes prediction of transmembrane region, disulfide bond and identification of important motifs. SOSUI distinguishes between membrane and soluble proteins from amino acid sequences and predicts the transmembrane helices for the former. The Transmembrane regions and their length were tabulated in Table 3. The server SOSUI classifies MBL homologue (Goldfish GBL) as soluble protein and other MBL homologues proteins as transmembrane proteins. SOSUI server has identified one transmembrane region in these proteins. The transmembrane regions are rich in hydrophobic amino acids. As disulphide bridges play an important role in determining the thermo stability of these proteins. CYS_REC was used to determine the Cysteine residues and disulphide bonds. Possible pairing and pattern with probability were presented in Table 4. Results showed that all proteins contain disulphide linkages.

The secondary structure of teleost MBL homologue proteins were predicted by SOPMA (Self Optimized Prediction Method with Alignment) which correctly predicts 69.5% of amino acids for a state description of the secondary structure prediction [5]. The secondary structure indicates whether a given amino acid lies in a helix, strand or coil. Secondary structure features as predicted using SOPMA were represented in Table 5. The results revealed that random coils dominated among secondary structure elements followed by alpha helix, extended strand and beta turns for all sequences. The secondary structure were predicted by using default parameters (Window width: 17, similarity threshold: 8 and number of states: 4).

CONCLUSION

In this study MBL homologues of teleosts were selected. Physicochemical characterization were performed by computing theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient, instability index, aliphatic index and grand average hydropathy (GRAVY). Functional analysis of these proteins was performed by SOSUI server. For these proteins disulphide linkages, motifs and profiles were predicted. Secondary structure analysis revealed that random coils dominated among secondary structure elements followed by alpha helix, extended strand and beta turns for all sequences. Computationally based characterization of the features of the proteins found or predicted in completely sequenced proteomes is an important task in the search for knowledge of protein function.

ACKNOWLEDGEMENT

Authors thank members of the Molecular Genetics laboratory DCFR, Bhimtal for helpful discussions. Authors also wish to acknowledge the encouragement and facilities provided by the Director of DCFR (ICAR), Bhimtal during the course of this investigation.

REFERENCES

1. Fletcher, G.L. and M.L. Rise, 2012. Aquaculture Biotechnology. Wiley-Blackwell, John Wiley & Sons Ltd, pp: 121.
2. Nelson, J.S., 1994. Fishes of the World. 3rd edition. Wiley, New York.
3. Nakao, M., M. Tsujikura, S. Ichiki, T.K. V.O. and T. Somamoto, 2011. The Complement system in teleost fish: progress of post-homolog-hunting researches. Dev. Comp. Immunol., 35: 1296-1308.
4. Lu, J., C. Teh, U. Kishore and K.B. Reid, 2002. Collectins and ficolins: sugar pattern recognition molecules of the mammalian innate immune system. Biochim. Biophys. Acta, 1572: 387-400.
5. Geourjon, C. and G. Deléage, 1995. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput. Appl. Biosci., 11: 681-684.
6. Guruprasad, K., B.V.P. Reddy and M.W. Pandit, 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.