A Comparative Study on Retinol Binding Protein in Two Freshwater Fishes, Danio rerio and Cyprinus carpio

Chittaranjan Baruah^{1*}, Papari Devi² and DK Sharma³

¹Department of Zoology, Darrang College, Tezpur, Assam, India ²Department of Zoology, Gauhati University, Guwahati, Assam, India ³School of Biological Science, USTM, Meghalaya, India

Research Article

Received date: 25/05/2016 Accepted date: 24/12/2016 Published date: 29/12/2016

*For Correspondence

Chittaranjan Baruah, Department of Zoology, Darrang College, Tezpur-784 001, Assam, India, Tel: 091-99542940800

E-mail: chittaranjan_21@yahoo.co.in

Keywords: Comparative modelling, RBP, Vitamin A, Phylogeny

ABSTRACT

Retinol-binding proteins (RBP) are a family of proteins with diverse functions. They are carrier proteins that bind retinol. An in silico study has been performed for 3D structure prediction and molecular phylogenetic analysis of RBP from two freshwater fishes, Danio rerio and Cyprinus carpio. The analyses were performed using sequence data of rbp gene and retinol binding protein, extracted from Gen Bank and UNIPROKB/SWISSPROT databases respectively. The evolutionary analyses of rbp gene have been conducted in MEGA5 by Maximum Likelihood and Maximum Parsimony methods. The present study represents the application of comparative modeling for 3D structure prediction of RBP and molecular phylogenetic analyses of rbp gene in MEGA5, using two different methods namely Maximum Parsimony and Maximum Likelihood. The predicted structures were deposited to Protein Model Database (PMDB) after statistical verification. The study shows that rbp gene of Cyprinus carpio is the closest ortholog of the rbp gene of Denio rerio. The structures of RBP can be helpful in structural biology for further investigations on active sites, molecular mechanism of function and structure based phylogeny. The structure will be helpful in the further understanding of the RBP-retinol complex interacts with transthyretin in freshwater fishes. The phylogenetic tree revealed the evolutionary relationship of RBP family.

INTRODUCTION

Freshwater fish having both retinol and dehydroretinol showed the capabilities of cleaving the provitamin A-status carotenoids into both forms of vitamin A, through either central or terminal cleavage ^[1-13]. Retinol-binding protein (RBP) which delivers the retinol from the liver to the target tissues in vertebrates, acts as a transport protein for plasma and cellular retinoids ^[14] and transporter of small hydrophobic molecules like lipids, bilins, steroids etc., belongs to the lipocalin family. Documentation of piscine RBP in fishes has already been made ^[15,16]. The fish RBP is a monomeric protein of 21 KDa provided with single retinol binding site ^[17]. Moreover, the transport of RBP is associated with transthyretin (TTR), which, in fact, prevents the filtration of RBP through the kidney glomerulus.

The structure of RBP is highly conserved ^[18] and there has been considerable similarity among the RBPs of various species, yet full knowledge on the RBP is not yet known ^[19]. It is established that RBP molecule binds to a TTR tetramer and the interection domain with TTR in human has been shown ^[20,21]. In fact the mechanism of retention of RBP even without TTR^[22] is not known in fishes, yet Folli ^[23] recorded the distinct binding site and structural properties of TTR in fish.

Lipocalins are a group of proteins with a wide range of functions with high level of sequence and closer similarity in their folding status (Flower 1995) The RBP attains at a stable condition at low pH ^[17]. RBP consists of a single globular domain protein ~40 Ű made up of an N-terminal coil, a beeta-sheet core, an alpha-helix and C-terminus coil. It is an eight stranded up and down beeta-barrel core where the retinol is sandwiched ^[5].

Although, there is availability of sequence information for Retinol Binding Protein (RBP), yet there is no structural information available. Therefore, the biochemistry and molecular mechanism of their functions in freshwater fish is still not very well understood due to lack of their species specific structural information. Therefore, an attempt was made to predict the 3D folding pattern and sequence analysis of RBP. The computational model of RBP can be helpful in structural biology for revealing functional

information and characterization. It is said that a model of a protein predicted using in-silico method is around 95% similar to its natural structure ^[24]. The key to a successful homology modelling is not usually the server or software used to predict the 3D model, the skill of designing a good alignment to a template structure is far more critical ^[20].

MATERIALS AND METHODS

Sequence Analyses

The study was extended to data mining and sequence analyses of RBP from Danio rerio and Cyprinus carpio (UNIPROKB/ SWISSPROT Accession numbers Q9PT95 and Q9DET6 respectively). The statistical analyses were performed using CLC Genomics Workbench (CLC Bio, Hyderabad). The important calculations for the amino acid composition, atomic composition, Theoretical pl, molecular weight, Formula, Extinction coefficients, half-life, Instability index, Aliphatic index, Hydrophobicity, charge vs. pH etc. were carried out under sequence analysis.

Molecular Phylogenetic Analysis of RBP Gene and RBP

The sequences for the rbp gene have been separately aligned using ClustalW 1.6 ^[18] integrated in software MEGA5, using default parameters. The rbp dataset were subjected to phylogenetic analyses. Evolutionary analyses were conducted in MEGA5. The evolutionary history of rbp gene and RBP protein were inferred by using Maximum Likelihood estimates. Nucleotide and amino acid substitution models that best fit each dataset and the model parameters were estimated using Akaike information criterion implemented in the program MODELTEST version 3.7 ^[25]. The evolutionary analysis rbp gene involved 11 nucleotide sequences and the phylogeny analysis of RBP involved 8 amino acid sequences.

Three-dimensional Structure Prediction

WU BLAST ^[26] and FASTA ^[27-29] searches were performed independently with PDB ^[30,31] for obtaining a suitable template. The significance of the BLAST results was assessed by expect values (e-value) generated by BLAST family of search algorithm. The target-template alignment ^[32] was carried out using ClustalW version 2.1 and Modeller 9.10 ^[33] programmes. Comparative (Homology) modelling was conducted by the Modeller version 9.10 ^[34]. The final 3D structures with all coordinates for RBP for both the targets were obtained by optimization of a molecular probability density function (pdf) of Modeller ^[35]. The molecular pdf for homology modelling was optimized with the variable target function procedure in Cartesian space that employed the method of conjugate gradients and molecular dynamics with simulated annealing ^[36].

The 3D structures for RBP were evaluated ^[37] by ERRAT ^[38] and ProCheck ^[39] programmes. After fruitful verification, the coordinate files were successfully deposited to PMDB ^[40]. All the graphic presentations of the 3D structures were prepared using Chimera ^[41] and RasMol programs ^[42].

RESULTS

Sequence Analyses

The rbp gene (cDNA) sequence of the present study was 891 and 1006 nucleotide long in Cyprinus carpio and Danio rerio with molecular weights of 288.258 kDa and 325.597 kDa respectively. The melting temperature ranged between 81.85 to 83.86 at 0.1M salt concentration (**Table 1**). The nucleotide sequence analysis based on the homologous rbp mRNA (cDNA) sequence showed the domination of A:T in the rbp gene (**Table 1**; **Figure 1**). The frequency of AT in Cyprinus carpio and Danio rerio were 0.538 and 0.586 respectively (**Table 1**).

Information	Danio rerio	Cyprinus carpio	
Length	1006 nuc	891 nuc	
Molecular weight	325.597 kDa	288.258 kDa	
Frequencies of A+T	0.586	0.538	
Frequencies of C+G	0.414	0.462	
Melting temperature (°C) at [salt]=0.1M	81.85	83.86	

Table 1. Nucleotide sequence statistics.

The primary structures of RBP are shown to comprise of 192 and 213 amino acid residues in D. rerio and C. carpio respectively. The amino acid alanine (A) and Aspertic acid (D) has been found predominantly rich in the RBP of these two fish species (**Figure 2A**). Sequence analysis of RBP revealed negative hydropathy on average (-0.228 to -0.235) (**Table 2; Figure 3**). The molecular weight of RBP in the fishes of the present study ranged from 21970.8 Da(in D. rerio) to 24522.8 Da (in C. carpio). The isoelectric point of the RBP ranged from 4.83 (in D. rerio) to 4.98 (in C. carpio) (**Table 2; Figure 2B**). Extinction coefficients for RBP are 42900 [Abs 0.1% (=1 g/I) 0.758] and 50015 [Abs 0.1% (=1 g/I) 0.748] in D. rerio and C. carpio respectively. The instability index (II) of RBP was computed to be 33.52 to 36.50. There were 27 to 29 negatively charged and 18 to 22 positively charged amino acid resides in the RBP sequence. The aliphatic index for RBP was computed in the range of

73.23 to 74.65 (**Table 2**). Multiple sequence alignment of the RBP protein showed that Danio rerio has a deletion of 'FLESNTT VKQDCALGTCWAQDCL' from 19th to 39th positions (**Figure 4**).







Figure 2A. Distribution of nucleotides in rbp gene of D. rerio and C. carpio.

Protein statistics	Danio rerio	Cyprinus carpio
Number of amino acids	192	213
SWISS PROT AC number	Q9PT95	Q9DET6
Molecular weight	21970.8 Da	24522.8 Da
Theoretical pl	4.83	4.98
Total number of negatively charged residues (Asp+Glu)	27	29
Total number of positively charged residues (Arg+Lys)	18	22
Formula	$C_{980}H_{1470}N_{258}O_{291}S_{14}$	$C_{1089}H_{1653}N_{289}O_{324}S_{17}$
Total number of atoms	3013	3372
Ext. coefficient (Abs 0.1% (=1 g/l) 0.758, assuming ALL Cys residues appear as half cystines)	42900	50015
Instability index	33.52	36.50
Aliphatic index	73.23	74.65
Grand average of hydropathicity (GRAVY)	-0.235	-0.228

Table 2. Protein statistics.

Phylogenetic Profile of rbp Gene and RBP

The evolutionary tree of rbp gene among eleven (11) fishes supports the fact that rbp gene in Danio rerio and Cypruns carpio has parallel evolution and they are represented as sister taxa with bootstrap support 100%, while Sulmo salar, Onchthyncus sp. and Anguilla Anguilla were clustered together as their successive sister group as observed in both in the MP and ML tree. Oreochromis niloticus and Tetradon nigroviridis represented as outgroup in the rbp gene phylogeny (Figures 5 and 6).

The evolutionary history of rbp gene was inferred by using the Maximum Likelihood method based on the Kimura 2-parameter model ^[43-49]. The bootstrap consensus tree inferred from 1000 replicates ^[50] is taken to represent the evolutionary history of the taxa analyzed ^[51-55]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000

e-ISSN:2321-6190 p-ISSN:2347-2294

replicates) are shown next to the branches ^[56]. Initial tree(s) for the heuristic search were obtained automatically as follows. When the number of common sites was <100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter=0.4797)]. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There was a total of 530 positions in the final dataset.











Figure 3. Plot of local Hydropathy for RBP (Kyte-Doolittle scale, Kyte and Doolittle). A. Danio rerio, B. Cyprinus carpioi.



Figure 4. Amino acid sequence alignment of RBP. '-' represent sequence not conserved. The sizes of the letter in the sequence logo represent the degree of conservation of respective amino acid in each alignment position.



Figure 5. Molecular phylogenetic analysis of rbp gene using Likelihood estimates. A. rbp gene, B. RBP protein. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). The scale bars represent the branch lengths measured in the number of substitutions per site over the whole sequence.



Figure 6. The predicted final structures of RBP displayed by UCSF Chimera. a. Danio rerio and b. Cyprinus carpio.

e-ISSN:2321-6190 p-ISSN:2347-2294

The evolutionary history of RBP protein was inferred by using the Maximum Likelihood method based on the Whelan And Goldman model ^[57-60]. The tree with the highest log likelihood (-1108.0087) is shown. When the number of common sites was <100 or less than one fourth of the total number of sites, the maximum parsimony method was used; otherwise BIONJ method with MCL distance matrix was used. A discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter=0.6365)]. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 175 positions in the final dataset.

The Tertiary Structures of RBP

The resultant 3D structures of RBP are based on the coordinates from PDB Id 1qab, selected through WU BLAST and EBI's FASTA and target-template alignment results. The tertiary structures of RBP in Danio rerio has 4 sheets, 5 beta hairpins, 4 beta bulges, 14 strands, 1 helix, 24 beta turns and 3 gamma turns and the RBP structure of Cyprinus carpio has has 4 sheets, 7 beta hairpins, 5 beta bulges, 16 strand, 2 helices, 29 beta turns and 3 gamma turns. Procheck verification proved that the models are of good quality (88.8% to 96.3% amino acids in theoretically allowed region) as judged by Ramachandran Plot ^[61,62] The structures of RBP have been deposited to Protein Model Database ^[63]. The tertiary structures of RBP of Danio rerio and Cyprinus carpio have been assigned PMDB IDs PM0076072 and PM0078117 respectively.

DISCUSSION

Based on WU BLAST and EBI's FASTA results 1QAB (Chain E, human retinol binding protein) was the best template for modelling of target proteins. The computation has been carried out on the complete sequence for both the proteins. The sequence analysis showed that the molecular formula for RBP of Danio rerio and Cyprinus carpipio are computed as $C_{980}H_{1470}N_{258}O_{291}S_{14}$ and $C_{1089}H_{1653}N_{289}O_{324}S_1$ respectively. Sequence analysis of RBP revealed negative hydropathy on average (Table 2; Figure 3), which signifies the polar and hydrophilic in nature of the RBP. The instability index (II) for Danio rerio and Cyprinus carpio (33.52 to 36.50) classifies the RBP as stable. ProCheck analyses of the structure of the RBP of fish are in conformity with the human serum retinol binding protein (PDB ID 1RBP), having amino acid residues of 175.

The models presented here can serve as a guide for the allocation of amino acid residues involved in each fold, which is important for further investigations on the molecular mechanism of functions. The present study was performed for sequence analyses and prediction of 3D structure of RBP from Danio rerio and Cyprinus carpio using the Homology modelling. A series of molecular modelling and computational methods were combined to gain insight into the 3D structure. Further study, investigating the role of other factors in RBP biosynthesis in wet lab is going on in our laboratory, which could add important information to the overall understanding of the fish model.

ACKNOWLEDGEMENTS

The paper is dedicated with deep reverence to Late Professor UC Goswami, Former Professor and Head of Department of Zoology, Gauhati University, Assam, India who inspired the authors for molecular and in-silico analysis of Retinoids and Carotinoids.

REFERENCES

- 1. Barua AB and Goswami UC. Formation of vitamin A in freshwater fish. Biochem. J. 1977;164:133-136.
- 2. Goswami UC and Barua AB. Origin of retinol in freshwater fish. Curr. Sci. 1981;50:150-151.
- 3. Goswami UC and Barua AB. Distribution of retinol and dehydroretinol in freshwater fish. Indian J. Biochem. Biophys. 1981;18:383-385.
- 4. Goswami UC and Barua AB. Distribution of retinol and dehydroretinol in freshwater fish. Indian J. Biochem. Biophys. 1981;18:383-385.
- 5. Altschul SF. Amino acid substitution matrices from an information theoretic perspective. J. Mol. Biol. 1991-219:555-565.
- 6. Goswami UC and Bhattacharya S. Biosynthesis of 3-dehydroretinol : Metabolism of lutein (Beta-E-carotene, 3-3' diol) in Clarias batrachus and Ompok pabo. Biochem. Internat. 1982;5:545-552.
- 7. Goswami UC. Metabolism of cryptoxanthin in freshwater fish. Br. J.Nutr. 1984;52:575-581.
- 8. Goswami UC. Biosynthesis of vitamin A2 (Dehydroretinol) in freshwater Bagrid, Mystus. Comp. Physiol. Ecol. 1984;9:87-90.
- 9. Goswami UC. Occurence of diversed retinoid molecules in freshwater piscian diversity. Fish Research. 2006;4:157-158.
- 10. Berman H, et al. The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data.Nucleic Acids Res 35. 2007;D301-303.
- 11. Goswami UC. Metabolism and utilization and utilization of pigment molecules in designing feeds for freshwater ornamental fish and crustaceans. In Emerging trends in Zoology, edited by U.C.Srivastava and Santosh Kumar, NPH Publication, New Delhi. 2011; pp:379-394.

- 12. Goswami J and Goswami UC. Carotenoids pigmentation in Polyacanthus fasciatus during spawning period. J. Pharmacy 2012;2:5-8.
- 13. Gasteiger E, et al. Protein Identification and Analysis Tools on the ExPASy Server; (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. 2005; pp:571-607.
- 14. Apweiler R, et al. UniProt: the Universal Protein knowledgebase. Nucleic Acids Res. 2004;32:D115-119.
- 15. Shidoji Y and Muto Y. Vitamin A transport in plasma of the non-mammalian vertebrates: isolation and partial characterization of piscine retinol-binding protein. J. Lipid Res. 1977;18:679-691.
- 16. Kyte J and Doolittle R. A simple method for displaying the hydropathic character of a protein. J. Mol.Biol. 1982;157:105-132.
- 17. Lüthy R, et al. Assessment of protein models with three-dimensional profiles. Nature. 1992;356:83-85.
- 18. Higgins D, et al. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence eighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673-4680.
- 19. Hooft RWW, et al. Errors in protein structures. Nature. 1996;381:272-272.
- 20. Murzin AG, et al. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 1995;247:536-540.
- 21. Zemla A, et al. Processing and analysis of CASP3 protein structure predictions Proteins, Suppl. 1999;3:22-29.
- 22. Martí-Renom MA, et al. Comparative protein structure prediction. John Wiley & Sons, Inc. Current Protocols in Protein Science. 2002;1:291-292.
- 23. Fiser A and Sali A. ModLoop: automated modelling of loops in protein structures Bioinformatics 2003;18:2500-2501.
- 24. Peicheng Du, et al. Have we seen all structures corresponding to short protein fragments in the Protein Data Bank? An update, Protein Engineering 2003;16:407-414.
- 25. Rahman MA, et al. Transgen. Res. 1998;7:357-369.
- 26. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389-3402.
- 27. Pearson WR and Lipman DJ. Improved Tools for Biological Sequence Comparison. PNAS 1988;85:2444-2448.
- 28. Pearson WR. Rapid and Sensitive Sequence Comparison with FASTP and FASTA. Methods in Enzymology. 1990;183:63-98.
- 29. Pearson WR. Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms. Genomics 1991;11:635-650.
- 30. Kouranov A. The RCSB PDB information portal for structural genomics. Nucleic Acids Res. 2006; 34:302-305.
- 31. Goswami U C. Vitamin A in fresh water fish, Presidential lecture in the section Animal, Vetenary and fishery Sciences, 94th Indian Science congress, Annamalai University, Annamalainagar. 2007;P1-26.
- 32. Lassmann T and Sonnhammer EL. Automatic assessment of alignment quality. Nucleic Acids Research. 2005;33:7120-7128.
- 33. Fiser A, et al. Modelling of loops in protein structures, Protein Science. 2000;9:1753-1773.
- 34. Marti-Renom MA, et al. Comparative protein structure modelling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. 2000;29:291-325.
- 35. Lassmann T and Erik LL. Sonnhammer Kalign, Kalignvu and Mumsa: web servers for multiple sequence alignment. Nucleic Acids Research. 2006;34:596-599.
- 36. Sali A and Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 1993;234:779-815.
- 37. Giorgetti A, et al. Evaluating the usefulness of protein structure models for molecular replacement. Bioinformatics. 2005;21:72-76.
- 38. Colovos C and Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. Protein Sci. 1993;2:1511-1519.
- 39. Laskowski RA, et al. From protein structure to biochemical function? J. Struct. Funct. Genomics, 2003;4:167-177.
- 40. Tiziana C, et al. The PMDB Protein Model Database CASPUR, Nucleic Acids Research 2006;34:306-309.
- 41. Pittersen EF, et al. UCSF Chimera-a visualization system for exploratory research and analysis.J.Comput.chem. 2004;25:1605-1612.
- 42. Sayle R and Milner-White. EJ Rasmol; biomilecular graphics for all. Trends Biochem. Sci. 1995; 20:374.

- 43. Boeckmann B, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003 Nucleic Acids Res. 2003;31:365-370.
- 44. Cornolis GR and Van Gijsegem F. Assembly and function of typeIII secretory systems. Annu.Rev.Microbiol. 2000;54:735-774.
- 45. Eisenberg D, et al. VERIFY3D: assessment of protein models with three-dimensional profiles Methods Enzymol. 1997;277:396-404.
- 46. Goswami UC. Metabolism of cryptoxanthin in freshwater fish. Br. J.Nutr. 1984;52:575-581.
- 47. Goswami UC. Biosynthesis of vitamin A2 (Dehydroretinol) in freshwater Bagrid, Mystus. Comp. Physiol. Ecol. 1984;9:87-90.
- 48. Fleming A. On a remarkable bacteriolytic element found in tissues and secretions. Proc Roy Soc Ser B. 1922;93:306-317.
- 49. Gough J, et al. Assignment of homology to genome sequences using a library of Hidden Markov Models that represent all proteins of known structure. J. Mol. Biol. 2001;313:903-919.
- 50. Hutchinson EG and Thornton JM. PROMOTIF A program to identify and analyze structural motifs in proteins. Protein Science. 1996;5:212-220.
- 51. Laskowski RA, et al. ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res. 2005;33:89-93.
- 52. Madera M, et al. The superfamily database in 2004: additions and improvements. Nucl. Acids Res. 2004;32:235-239.
- 53. Moult J, et al. Critical assessment of methods of protein structure prediction (CASP)-round V Proteins 2003;53:334-339.
- 54. Muto Y, et al. Isolation and characterization of serum retinol-binding protein. Methods of Enzymology. 1982;81:840-852.
- 55. Notredame C, et al. T-Coffee: A novel method for multiple sequence alignments. Journal of Molecular Biology. 2000;302:205-217.
- 56. O'Sullivan O, et al. 3Dcoffee Combining Protein Sequences and Structures within Multiple Sequence Alignments. Journal of Molecular Biology. 2004;340:385-395.
- 57. Peter RE, et al. Interactions of catecholamines and GnRH in regulation of gonadotropin secretion in teleost fish. Recent. Prog. Horm. Res. 1986;42:513-548.
- 58. Pontius J, et al. Deviations from standard atomic volumes as a quality measure for protein crystal structures.protein database searches with composition-based statistics and other refinements", Nucleic Acids Res. 1996;29:2994-3005.
- 59. Ramachandran and Sasisekharan. The Ramachandran plot (Ramachandran and Sasisekharan) Adv. Protein Chem. 1968;23:283-437.
- 60. Sharma N and Goswami UC. Functioning of Lycopene, a Non-provitamin: A Carotenoid in Mammalian System : A Review. Proc Zool Soc. 2012;1:4.
- 61. Varadarajan L. Silk in Northeastern and Eastern India: The Indigenous Tradition, Modern Asian Studies. 1988;22:561-570.
- 62. Vocadlo DJ, et al. Nature 2001;412:835.
- 63. Williams H, et al. Identification and analysis of additional copies of the PDGF-R- and colony stimulating factor 1 receptor genes in fugu. Gene 2002;295:255-264.