



E-ISSN: 2320-7078

P-ISSN: 2349-6800

www.entomoljournal.com

JEZS 2020; 8(2): 652-660

© 2020 JEZS

Received: 11-01-2020

Accepted: 15-02-2020

Ajit Tiwari

Bioinformatics Centre, College of
Fisheries, CAU (I),
Lembucherra, Tripura, India

AD Upadhyay

Bioinformatics Centre, College of
Fisheries, CAU (I),
Lembucherra, Tripura, India

Himanshu Priyadarshi

Department of Fish Genetics and
Reproduction (FGR), College of
Fisheries, CAU (I),
Lembucherra, Tripura, India

Rumpi Ghosh

Bioinformatics Centre, College of
Fisheries, CAU (I),
Lembucherra, Tripura, India

Suresh Yambem

Department of Fish Genetics and
Reproduction (FGR), College of
Fisheries, CAU (I),
Lembucherra, Tripura, India

Corresponding Author:**Ajit Tiwari**

Bioinformatics Centre, College of
Fisheries, CAU (I),
Lembucherra, Tripura, India

Structural and functional analysis of ferritin heavy chain subunit in freshwater fishes

Ajit Tiwari, AD Upadhyay, Himanshu Priyadarshi, Rumpi Ghosh and Suresh Yambem

Abstract

Iron excess in form of ferrous (Fe^{2+}) in soil and water is one of the major issues in aquaculture in Tripura state of India. Excess iron in water causes iron toxicity in fishes, so it becomes important to identify the structural and functional role of iron related protein in fish species. Ferritin heavy chain subunit has di-Fe binding site in fourth helix that interacts with oxygen and is responsible for ferroxidase activity in fishes. In present study, structural and functional property of ferritin heavy chain protein in fishes was analyzed. The structural and functional analyses of ferritin in the fish species belonging to different orders were performed on ExPasy ProtParam, GOR IV, Swiss Model, and STRING 10.0. Physicochemical characterization of ferritin heavy chain subunit of freshwater fishes reveals that the protein is acidic, unstable and hydrophilic with maximum percentage of alpha helix followed by random coil and extended strand. The tertiary structure of ferritin heavy chain subunit was predicted by the Swiss Model and verified by Ramachandran plot. In protein-protein interaction analysis via STRING 10.0 tool, 10 enriched pathways of KEGG, 15 UNIPROT, 13 PFAM Domains common to InterPro Scan and 7 SMART Protein Domains were identified in 35 species of freshwater fish.

Keywords: Ferritin heavy chain subunit, ferroxidase, protein-protein interaction, structural analysis.

Introduction

In North-East India especially Tripura, the water bodies are embodied with high iron content. Excess iron in an aquatic ecosystem reacts as a catalyst in the Fenton reaction producing free radicals which are toxic to fishes [1]. Ferritin is an iron storage protein which is important in cellular homeostasis and cellular defense against oxidative stress produced by reactive oxygen species (ROS) [2, 3, 4]. The molecular weight of a ferritin protein is 450 kDa, made up of 24 protein subunits in eukaryotes systemize in such a way to form a hollow shell with an 80Å diameter cavity, stores 4,500 Fe (III) atoms in the form of biomineral [5, 6, 7]. In mammals, ferritin molecules of heavy (H) and light (L) chain subunits having molecular masses of 21 and 19 kDa. Heavy chains are important for Fe (II) oxidation, whereas Light chains assist in iron nucleation, mineralization, and long-term storage [7]. The Heavy and Light chain subunits co-assemble in different ratios to form a protein shell of 24 subunits capable of acquiring iron atoms. The Heavy chain is responsible for the ferroxidase activity of the protein found abundant in tissues of kidney and heart, whereas Light chain ferritins are more abundant in the spleen and liver [8].

Currently, only biophysical experimental techniques such as X-ray crystallography and nuclear magnetic resonance are able to provide precise protein tertiary structures. There are 17,473,872,940 protein sequences in the latest release of UNiVersal PROTein resource Knowledge Base (UniProtKB)/Translated European Molecular Biology Laboratory (TrEMBL) as of 22nd April 2014, whereas the Protein Data Bank (PDB) contained only 99,624 protein structures till then [9]. However, solving the protein structure via X-ray crystallography and Nuclear Magnetic Resonance (NMR) are costly and time consuming. Therefore, the computational based structural prediction of proteins is highly required to fill the gap between the protein sequences available in public domain databases and their experimentally solved structures. The primary structure prediction, secondary structure prediction, tertiary structure prediction and protein-protein interaction of ferritin heavy chain subunit for freshwater fish species were analyzed in current study.

Materials and Methods

Sequence retrieval

The amino acid sequence of Ferritin heavy subunit of 35 fishes was retrieved from National Center for Biotechnology Information (NCBI) with their accession number XP_015239430.1, CCH79_00005047, XP_007553939.1, XP_008403364.1, XP_014854331.1, XP_005812594.1, XP_018954752.1, NP_571660.1, AVO65123.1, XP_026071941.1, XP_016403866.1, XP_016135903.1, XP_020329170.1, XP_021442122.1, AAK08117.1, XP_024274891.1, ABY21333.1, NP_001117129.1, XP_004543325.1, XP_003445743.1, XP_005937262.1, XP_026233815.1, XP_028999730.1, XP_028442399.1, XP_020569048.2, XP_024150877.1, XP_007237415.1, XP_017573625.1, NP_001187267.1, XP_027008045.1, XP_003969725.1, XP_028811730.1, XP_023675493.1, XP_018619415.1, XP_026175970.1.

Structural and functional analyses

Several online web services and software were used for the analyses of ferritin heavy chain (fth) in fishes.

Physicochemical properties

Analyses were carried out online at the EXPASY website (<http://expasy.org/tools>). This tool was used in present study to investigate the physico chemical parameter of ferritin heavy chain subunit of fresh water fishes. The parameters calculated by this tool are molecular weight, theoretical pI (Isoelectric point), % amino acid composition, total number of negative and positive residue, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY) [10].

Secondary structure prediction

The secondary structure of the ferritin heavy subunit was analyzed by GORIV methodology [11].

Tertiary structure prediction

Template search for the query protein and % identity, homology modeling, three-dimensional structure of ferritin heavy subunit was predicted using Swiss model server [12, 13]. The predicted structure of the ferritin heavy subunit was validated through Ramachandran plot by utilizing rampage server [14].

Protein-protein interactions

STRING 10.0 was used to predict the protein/genes interaction from stored database (<http://string-db.org/>) [15]. The information encoded in database from different sources, including experimental repositories, computational prediction methods, and public text groups were retrieved using the tool. The eight number of input and 50 number of interactors shown by matching proteins to *Danio rerio* which is stored in a database.

Result & Discussion

Structural and functional property of Ferritin heavy chain subunit (fth) Physicochemical properties

The primary structure analysis was done by using ProtParam server, where the amino acid composition, atomic composition, instability, molecular weight, isoelectric point, aliphatic index, and grand average of hydropathicity (GRAVY) calculated (Table 1). The range of theoretical isoelectric points and molecular weight are calculated as 5.45 (*Betta splendens*) to 5.78 (*Xiphophorus maculatus*) and

12739.41 (*Gambusia affinis*) to 20887.40 (*Takifugu rubripes*), respectively. Values of GRAVY ranged from -0.860 (*Oryzias melastigma*) to -0.679 (*Betta splendens*). The GRAVY value for a protein is calculated as the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence [16]. The score of protein indicate the nature of the protein. It is found hydrophilic at score below 0 (globular), while above at scores 0 is hydrophobic (membranous). The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine). It can be used as a measure of thermostability of proteins [17]. According to Table 1, the least aliphatic index was found to be 65.08 in *Mylopharyngodon piceus* while the highest was found to be 77.74 in *Betta splendens*. Computed values of instability index of ferritin heavy chain subunit were between 54.83 (*Tachysurus fulvidraco*) and 38.05 (*Oncorhynchus nerka*). The instability index in fth for *Oncorhynchus nerka* was 38.05, *Salmo salar* with 38.31 and *Mastacembelus armatus* with 39.55 were below 40, known to be stable protein. Based on the obtained result, high aliphatic index in fth indicated the structural stability of this protein. The instability index is used as measure of stability, unstability and in vivo half-life of a protein [18], therefore proteins which have in vivo half-life of less than 5 hours have been shown to have an instability index of more than 40, predicting unstability, whereas those with half-life of more than 16 hours have an instability index of less than 40 predicting stability [19]. Ferritin heavy chain subunit in freshwater fishes shows that the protein is acidic, hydrophilic and unstable except three fish species viz *Oncorhynchus nerka*, *Salmo salar* and *Mastacembelus armatus*.

Secondary structure prediction

In secondary structure analysis, alpha helices were more dominant among random coils and extended strands. The secondary structure of ferritin heavy chain subunit proteins was predicted by GOR4 server based on hierarchical neural network analysis which calculates the percentage of alpha helix, extended strand and random coil. The value of alpha helix, extended strands and random coils vary in fresh water fish species. Alpha helix ranged from 67.61 (*Oncorhynchus nerka*) to 44.14 (*Gambusia affinis*), extended strands from 15.82 (*Sinocyclocheilus rhinocerosus*) to 3.95 (*Pygocentrus nattereri*) and coiled region 43.24 (*Gambusia affinis*) to 27.12 (*Perca flavescens*) (Table 3). The residues like Asp, Ala, Glu, Gln, Leu, Lys, Ser, Gly and Tyr represented in Table 2 found maximum while Pro, Trp, Pyl, Met and Cys in minimum. It has been reported in previous study that the residues like Ala, Glu, Leu, Lys and Met are good helix formers while Pro, Gly, Tyr, and Ser are very poor helix formers of α -helices [20]. α -helices were central to all earlier attempt of researchers to predict secondary structure from amino acid sequence (e.g., [21, 22, 23, 24, 25] and they are still some characteristics that can be predicted with greatest accuracy [26, 27]. However, both theoretical and experimental studies have pointed out that compact states have a preference for secondary structure [28, 29, 30]. For instance, random coil may be involved in the binding of a ligand, with consequent changes in the conformation and activity of the protein. In Ferritin heavy chain subunit of freshwater fishes, alpha helices were found abundant followed by random coil and extended strand.

Tertiary structure prediction

Homology modeling predicted the 3D structure of the Ferritin

heavy subunit of different fishes. The structural conformation was generated from Swiss model server aligning the query sequence to the template sequence (Fig. 1). 3 D structural model were generated with the maximum % identity and QMQE value of the template. For homology modeling, identified template showed similarity in range from 84.80% (*Cyprinodon variegatus*) to 79.53% (*Cyprinus carpio*) sequence along with Q mean-score of 1.17 (*Gambusia affinis*) to -0.01 (*Cyprinodon variegatus*) and GMQE of 0.95 (*Gambusia affinis* and *Mastacembelus armatus*) to 0.91 (*Carassius auratus*) from the Swiss model database (Table 3). The parameter $C\beta$, All Atom, Solvation and Torsion of the model quality estimate were recorded along with QMEAN, GMQE score in range from 2.88 (*Oryzias latipes*) to 1.36 (*Cyprinus carpio*) in $C\beta$, 4.34 (*Gambusia affinis*) to 2.62 (*Xiphophorus maculatus*) in All Atom, 2.75 (*Gambusia affinis*) to 1.49 (*Mylopharyngodon piceus*) in Solvation, -0.16 (*Gambusia affinis*) to -1.16 (*Oncorhynchus kisutch*) in Torsion from the swiss model database The selected template contains a structure of mouse h-chain modified ferritin by X-ray diffraction technique (2.24Å). It was verified by swissmodel/Workspace possessing pdb code 3wnw.1.A from 84.80% (*Cyprinodon variegatus*) to 79.53 (*Cyprinus carpio*) sequence identities with the query sequence. The quality of the model was estimated by Qualitative Model Energy Analysis (QMEAN) score. The general understanding of ferritin structure is always based on the human ferritin subunit [31], frog ferritin [32] and the *E. coli* ferritin [33].

The combination of term QME-score, $C\beta$, All atom, torsion and solvation helped to estimate the model quality and selection. $C\beta$ atom is a mean force of two distance dependent interaction potentials which are the basis on optimization of a simplified energy function of a peptide backbone. This type of energy is the backbone of contact between function of the local interactions within all amino acid residues and groups of peptide. Ferritin heavy chain subunit in 35 freshwater fish species has all positive interaction. All atom types are applied to appraise long range interactions. A torsion angle potential over three consecutive amino acids is most important tool for the estimation of the local geometry of a model. In Swiss server 45 degree for the centre residue, 90 degree for the 2 adjacent residues is considered for evaluation of torsion angle energy. The negative energy indicates that a staggered conformation due to angle reduce to 60 degree [34]. A solvation potential is used to estimate the burial status of the residues, the agreement of the predicted and the calculated

secondary structures, and solvent accessibility prediction. The solvation energy of ferritin heavy chain ranged from 1.49 (*Mylopharyngodon piceus*) to 2.75 (*Gambusia affinis*) in freshwater fishes. Higher the solvation energy of globular proteins means their side chains are more hydrophilic and amphiphilic in nature. [34, 35, 36]. According to Ramachandran plots analyzed for over 35 structures at 2.0 Å resolution ranged from 96.20 (*Oncorhynchus kisutch*) to 98.57 (*Pygocentrus nattereri*) into the most favored region. It has been reported in previous studies that the good quality model can be obtained when greater than 90% of residues fall into the most favored region [37].

Functional interaction network analysis

In order to predict the interacting proteins, *Danio rerio* ferritin heavy chain protein was applied to the STRING 10.0 tool as a model of all freshwater fish samples. The Database was utilized to determine the interaction of known and predicted protein. The interactions occurred in the form of direct and indirect contacts. According to the results, 8 input with same functional role but different % identity given 50 functional partners in the network analyses, listed in Table 4. 10 enriched pathways of KEGG, 15 UNIPROT, 13 PFAM Domains common to InterPro Scan and 7 SMART Protein Domains were obtained by the protein– protein interaction analysis, which are listed in Table 5 with their characteristics. The K-means algorithm was used for protein clustering in ten different colours. Inter-cluster edges are represented by dashed lines. Small nodes: protein of unknown 3D structure; Large nodes: some 3 D structure is known or predicted; Extended lines: gene co-occurrence (Fig 2). Network nodes are proteins, and the edges represent the predicted functional associations. These networks can utilize functional genomics data; can be an important component for the understanding of cellular processes at system-level; can suggest new directions for future experimental research and provide cross predictions for efficient interaction mapping [38]. It is conceptually hierarchical and their interaction between proteins depict and narrow down a protein's function [39, 40, 41, 42].

In this investigation, functional partners and pathways of ferritin heavy chain subunit in freshwater fishes are represented in Table 4 and Table 5 respectively. The protein associated with iron related protein which somehow can be used to understand the numerous biological reactions within a cell. Hence, it can be a powerful tool for the prediction of the function of protein.

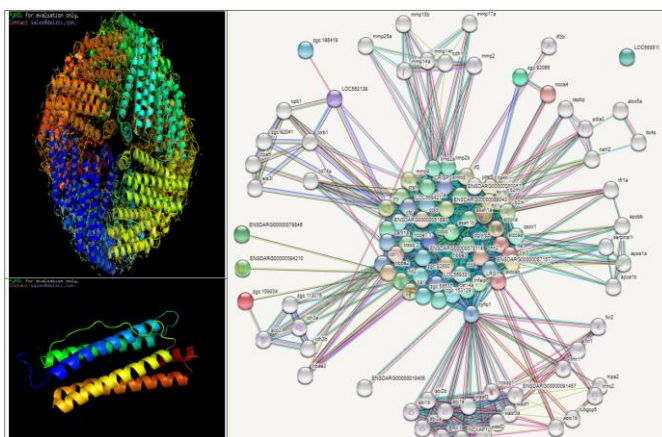


Fig 1: Comparative/Homology modelling of 3 D structure of Ferritin heavy chain subunit of *Danio rerio* (accession number: NP_571660.1) with template 3wnw.1.A via Swiss Model and Phyre 2.

Fig 2: The interactive network view of predicted protein-protein interactions using STRING 10.0 tool

Table 1: Physicochemical properties of the ferritin heavy chain subunit in freshwater fish species.

Physicochemical parameters	M. Wt.	No. of atoms	pI	Total -ve residue	Total +ve residue	Atomic composition					Estimated half-life	Instability index:	Aliphatic index	GRAVY
						C	H	N	O	S				
Fish species														
<i>Cyprinodon variegates</i>	20743.17	2854	5.56	29	21	907	1396	260	282	9	30 hrs	41.70	72.77	-0.807
<i>Gambusia affinis</i>	12739.41	1772	5.68	18	14	555	879	159	173	6	30 hrs	42.86	80.81	-0.710
<i>Poecilia Formosa</i>	20736.19	2853	5.65	29	21	908	1395	261	280	9	30 hrs	42.35	72.77	-0.782
<i>Poecilia reticulata</i>	20743.22	2858	5.70	29	22	908	1400	260	281	9	30 hrs	46.45	72.20	-0.797
<i>Poecilia maxicana</i>	20736.23	2857	5.78	29	22	909	1399	261	279	9	30 hrs	45.95	72.77	-0.785
<i>Xiphophorus maculatus</i>	20679.18	2850	5.78	28	21	907	1396	260	278	9	30 hrs	42.10	73.33	-0.747
<i>Cyprinus carpio</i>	20798.33	2857	5.58	29	21	911	1397	257	281	11	30 hrs	53.19	66.16	-0.786
<i>Danio rerio</i>	20719.23	2846	5.67	29	21	905	1392	258	280	11	30 hrs	52.58	66.16	-0.786
<i>Mylopharyngodon piceus</i>	20795.34	2851	5.55	29	20	911	1392	256	280	12	30 hrs	51.01	65.08	-0.777
<i>Carassius auratus</i>	20775.27	2856	5.55	29	20	913	1396	256	281	10	30 hrs	50.28	68.93	-0.770
<i>Sinocyclocheilus rhinoceros</i>	20736.26	2847	5.55	29	20	910	1391	255	280	11	30 hrs	54.34	66.72	-0.749
<i>Sinocyclocheilus grahami</i>	20759.29	2853	5.49	29	21	909	1396	256	281	11	30 hrs	50.76	66.72	-0.773
<i>Oncorhynchus kisutch</i>	20578.07	2836	5.56	29	21	905	1389	255	278	9	30 hrs	40.89	69.83	-0.801
<i>Oncorhynchus mykiss</i>	20689.21	2853	5.58	30	22	910	1398	258	278	9	30 hrs	40.46	69.83	-0.838
<i>Oncorhynchus nerka</i>	20748.32	2866	5.58	31	23	913	1407	259	278	9	30 hrs	38.05	73.18	-0.798
<i>Oncorhynchus tshawytscha</i>	20562.07	2835	5.56	29	21	905	1389	255	277	9	30 hrs	40.41	70.40	-0.786
<i>Oncorhynchus masou formosanus</i>	20552.03	2832	5.56	29	21	903	1387	255	278	9	30 hrs	41.81	70.40	-0.781
<i>Salmo salar</i>	20730.17	2854	5.57	29	21	908	1397	257	283	9	30 hrs	38.31	68.87	-0.821
<i>Maylandia zebra</i>	20805.22	2852	5.62	29	20	909	1390	262	281	10	30 hrs	46.02	70.00	-0.811
<i>Oreochromis niloticus</i>	20806.25	2855	5.62	29	20	910	1393	261	281	10	30 hrs	45.71	70.56	-0.793
<i>Haplochromis burtoni</i>	20736.11	2840	5.50	29	19	906	1383	259	282	10	30 hrs	47.54	70.00	-0.790
<i>Anabas testudineus</i>	20825.28	2866	5.56	29	21	911	1402	262	282	9	30 hrs	45.79	72.77	-0.797
<i>Betta splendens</i>	20669.18	2851	5.45	28	20	907	1398	258	279	9	30 hrs	42.47	77.74	-0.679
<i>Perca flavescens</i>	20757.30	2860	5.56	29	21	910	1402	258	280	10	30 hrs	40.49	73.90	-0.731
<i>Oryzias latipes</i>	20880.39	2865	5.54	30	21	912	1399	263	280	11	30 hrs	49.37	71.13	-0.823
<i>Oryzias melastigma</i>	20887.39	2866	5.57	30	22	910	1400	264	281	11	30 hrs	52.94	68.93	-0.860
<i>Astyanax mexicanus</i>	20772.26	2863	5.61	29	22	909	1403	261	281	9	30 hrs	51.56	71.13	-0.799
<i>Pygocentrus nattereri</i>	20840.30	2868	5.70	29	22	910	1403	265	281	9	30 hrs	50.26	71.13	-0.812
<i>Ictalurus punctatus</i>	20788.23	2855	5.56	29	21	907	1395	261	282	10	30 hrs	47.94	68.93	-0.813
<i>Tachysurus fulvidraco</i>	20791.25	2851	5.56	29	21	905	1392	262	281	11	30 hrs	54.83	67.29	-0.797
<i>Takifugu rubripes</i>	20887.40	2861	5.66	29	21	909	1396	264	280	12	30 hrs	47.75	66.72	-0.805
<i>Denticeps clupeoides</i>	20724.23	2854	5.51	29	22	905	1399	259	281	10	30 hrs	49.97	69.49	-0.792
<i>Paramomyrops kingsleyae</i>	20856.37	2863	5.54	30	21	911	1399	261	281	11	30 hrs	51.17	67.29	-0.821
<i>Scleropages formosus</i>	20846.34	2874	5.61	30	23	911	1409	263	282	9	30 hrs	50.89	71.13	-0.815
<i>Mastacembelus armatus</i>	20811.24	2865	5.56	29	21	914	1400	262	281	8	30 hrs	39.55	71.69	-0.811

Table 2: Amino acid composition in percent form of ferritin heavy chain subunit in freshwater fish species.

Amino acid %	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	Pyl	Sec
Fish species																						
<i>Cyprinodon variegates</i>	5.6	5.6	5.1	7.9	2.3	7.3	8.5	5.1	5.1	3.4	11.3	6.2	2.8	4.0	1.7	6.8	2.3	1.1	4.5	3.4	0.0	0.0
<i>Gambusia affinis</i>	3.6	4.5	4.5	8.1	2.7	6.3	8.1	7.2	4.5	3.6	13.5	8.1	2.7	2.7	2.7	6.3	3.6	1.8	1.8	3.6	0.0	0.0
<i>Poecilia Formosa</i>	5.6	5.6	5.1	7.9	2.3	6.8	8.5	5.1	5.6	3.4	11.3	6.2	2.8	4.5	1.7	6.8	2.3	1.1	4.0	3.4	0.0	0.0
<i>Poecilia reticulata</i>	5.1	5.6	5.1	7.9	2.3	6.8	8.5	5.1	5.1	4.0	10.7	6.8	2.8	4.5	1.7	7.3	2.3	1.1	4.0	3.4	0.0	0.0
<i>Poecilia maxicana</i>	5.6	5.6	5.1	7.9	2.3	6.2	8.5	5.1	5.6	3.4	11.3	6.8	2.8	4.5	1.7	6.8	2.3	1.1	4.0	3.4	0.0	0.0
<i>Xiphophorus maculatus</i>	5.6	5.6	4.5	7.3	2.3	6.8	8.5	5.6	5.6	4.0	11.3	6.2	2.8	4.5	1.7	6.8	2.8	1.1	4.0	2.8	0.0	0.0

<i>Cyprinus carpio</i>	6.2	5.1	5.1	5.6	1.7	1.7	10.7	4.5	5.1	2.8	9.6	6.8	4.5	5.1	1.7	7.9	1.7	1.1	4.0	4.0	0.0	0.0
<i>Danio rerio</i>	6.8	5.1	5.1	5.6	1.7	6.8	10.7	4.5	5.6	2.3	9.6	6.8	4.5	5.1	1.7	7.3	1.7	0.6	4.0	4.5	0.0	0.0
<i>Mylopharyngodon piceus</i>	6.8	4.5	5.1	5.6	1.7	6.8	10.7	4.5	5.6	2.8	9.6	6.8	5.1	5.1	1.7	7.3	1.7	1.1	4.0	3.4	0.0	0.0
<i>Carassius auratus</i>	6.2	4.5	5.1	5.6	1.7	6.8	10.7	4.5	5.6	2.8	10.7	6.8	4.0	5.1	1.7	7.9	1.7	1.1	4.0	3.4	0.0	0.0
<i>Sinocyclocheilus rhinoceros</i>	6.8	4.5	4.5	5.6	1.7	6.8	10.7	4.5	5.6	2.8	9.6	6.8	4.5	5.1	1.7	7.9	1.7	1.1	4.0	4.0	0.0	0.0
<i>Sinocyclocheilus grahami</i>	6.8	5.1	5.6	5.6	1.7	6.8	10.7	4.5	4.5	2.8	9.6	6.8	4.5	5.1	1.7	7.3	1.7	1.1	4.0	4.0	0.0	0.0
<i>Oncorhynchus kisutch</i>	6.2	4.5	5.7	7.4	1.7	6.8	9.1	4.5	5.1	2.8	9.7	7.4	3.4	4.5	2.8	5.7	2.3	1.1	4.0	5.1	0.0	0.0
<i>Oncorhynchus mykiss</i>	6.2	5.1	5.7	6.8	1.7	6.8	10.2	4.5	5.1	2.8	9.7	7.4	3.4	4.5	2.8	5.1	1.7	1.1	4.0	5.1	0.0	0.0
<i>Oncorhynchus nerka</i>	6.8	5.7	5.1	8.0	1.7	6.2	9.7	4.0	5.1	3.4	10.2	7.4	3.4	4.5	2.3	5.1	1.7	1.1	4.0	4.5	0.0	0.0
<i>Oncorhynchus tshawytscha</i>	6.8	4.5	5.7	7.4	1.7	6.8	9.1	4.5	5.1	2.8	9.7	7.4	3.4	4.5	2.8	5.1	2.3	1.1	4.0	5.1	0.0	0.0
<i>Oncorhynchus masou formosanus</i>	6.8	4.5	5.7	7.4	1.7	6.8	9.1	4.5	5.1	2.8	9.7	7.4	3.4	4.5	2.3	5.7	2.3	1.1	4.0	5.1	0.0	0.0
<i>Salmo salar</i>	5.6	4.5	5.6	6.8	1.7	7.3	9.6	4.5	5.1	2.8	9.6	7.3	3.4	4.5	1.7	6.8	2.8	1.1	4.0	5.1	0.0	0.0
<i>Maylandia zebra</i>	6.2	5.1	5.6	7.3	2.3	7.3	9.0	4.0	6.2	2.3	10.7	6.2	3.4	4.5	1.7	7.3	1.1	1.1	4.0	4.5	0.0	0.0
<i>Oreochromis niloticus</i>	6.2	5.1	5.1	7.3	2.3	7.3	9.0	4.0	6.2	2.8	10.7	6.2	3.4	4.5	1.7	7.3	1.7	1.1	4.0	4.0	0.0	0.0
<i>Haplochromis burtoni</i>	6.2	4.5	5.6	7.3	2.3	7.3	9.0	4.0	6.2	2.3	10.7	6.2	3.4	4.5	1.7	7.9	1.1	1.1	4.0	4.5	0.0	0.0
<i>Anabas testudineus</i>	5.6	6.2	5.1	7.3	2.3	7.3	9.0	4.0	5.1	3.4	11.3	5.6	2.8	4.5	2.3	7.9	1.7	1.1	4.0	3.4	0.0	0.0
<i>Betta splendens</i>	6.8	5.6	5.6	7.9	2.8	7.3	7.9	4.0	4.5	2.8	12.4	5.6	2.3	4.5	2.3	6.2	2.3	1.1	4.0	4.0	0.0	0.0
<i>Perca flavescens</i>	7.3	5.1	5.6	7.3	2.3	6.8	9.0	3.4	5.1	2.8	11.3	6.8	3.4	4.5	1.7	6.8	1.7	1.1	4.0	4.0	0.0	0.0
<i>Oryzias latipes</i>	5.6	5.6	6.2	7.3	2.8	6.8	9.6	5.1	5.6	3.4	11.3	6.2	3.4	4.5	1.7	5.1	1.7	1.1	4.0	2.8	0.0	0.0
<i>Oryzias melastigma</i>	5.6	6.2	6.2	7.9	2.3	6.8	9.0	5.1	5.1	3.4	10.7	6.2	4.0	4.5	1.7	5.6	1.7	1.1	4.0	2.8	0.0	0.0
<i>Astyanax mexicanus</i>	6.8	6.2	5.1	6.8	1.7	7.3	9.6	4.0	4.5	3.4	10.2	6.2	3.4	4.5	2.3	7.9	1.1	1.1	4.0	4.0	0.0	0.0
<i>Pygocentrus nattereri</i>	6.8	6.8	5.1	6.8	1.7	7.3	9.6	4.0	5.1	3.4	10.2	5.6	3.4	4.5	1.7	7.9	1.1	1.1	4.0	4.0	0.0	0.0
<i>Ictalurus punctatus</i>	6.8	5.6	5.6	7.3	1.7	7.3	9.0	4.0	5.1	3.4	9.6	6.2	4.0	4.5	1.7	7.3	1.7	1.1	4.0	4.0	0.0	0.0
<i>Tachysurus fulvidraco</i>	7.3	6.2	5.1	7.3	1.7	7.3	9.0	4.0	5.1	3.4	9.0	5.6	4.5	4.5	1.7	7.3	1.7	1.1	4.0	4.0	0.0	0.0
<i>Takifugu rubripes</i>	6.2	6.2	5.1	7.3	2.3	7.3	9.0	4.0	5.6	2.8	10.2	5.6	4.5	5.1	1.7	7.3	1.7	1.1	3.4	3.4	0.0	0.0
<i>Denticeps clupeioides</i>	7.9	6.2	5.1	6.8	1.7	7.3	9.6	4.0	4.0	2.8	9.6	6.2	4.0	4.0	2.3	6.8	1.7	1.1	4.5	4.5	0.0	0.0
<i>Paramormyrops kingsleyae</i>	6.8	5.6	5.1	6.8	1.7	6.8	10.2	4.5	5.6	3.4	9.6	6.2	4.5	4.5	1.7	5.6	2.8	1.1	4.0	3.4	0.0	0.0
<i>Scleropages formosus</i>	6.8	6.8	5.1	7.3	1.7	6.8	9.6	4.0	4.5	3.4	10.2	6.2	3.4	4.5	1.7	6.8	2.3	1.1	4.0	4.0	0.0	0.0
<i>Mastacembelus armatus</i>	6.8	5.6	6.2	7.9	1.7	7.3	8.5	4.5	5.1	3.4	10.7	6.2	2.8	5.1	1.7	5.1	2.8	1.1	4.0	3.4	0.0	0.0

Table 3: Secondary and Tertiary structure prediction analytical data of ferritin heavy chain subunit in freshwater fish species.

Secondary and Tertiary structure analytical data	Alpha helix %	Extended strand %	Random coil %	Ligands		GMQE	QMean	C β	All Atom	Solvation	Torsion	Template identity %	Favoured %
				Fe ion	Mg ion								
<i>Fish species</i>													
<i>Cyprinodon variegatus</i>	54.80	11.86	33.33	6	26	0.94	-0.01	2.47	2.89	1.80	-1.08	84.80	97.78
<i>Gambusia affinis</i>	44.14	12.61	43.24	6	6	0.95	1.17	1.68	4.34	2.75	-0.16	84.11	98.02
<i>Poecilia formosa</i>	54.80	10.73	34.46	6	24	0.94	0.49	2.79	2.72	2.00	-0.69	83.63	97.49
<i>Poecilia reticulata</i>	54.54	10.73	35.03	6	24	0.93	0.26	2.24	2.69	1.83	-0.77	81.87	96.84
<i>Poecilia maxicana</i>	54.80	10.73	34.46	6	24	0.93	0.51	2.62	2.67	1.95	-0.63	83.04	97.14
<i>Xiphophorus maculatus</i>	55.37	10.73	33.90	6	24	0.93	-0.08	2.24	2.62	1.81	-1.10	83.04	96.99
<i>Cyprinus carpio</i>	55.37	15.82	28.81	24	24	0.92	0.14	1.36	2.69	1.61	-0.69	79.53	98.08
<i>Danio rerio</i>	57.06	15.82	27.12	24	24	0.93	0.32	1.57	2.81	1.97	-0.66	80.12	98.13
<i>Mylopharyngodon piceus</i>	57.06	15.82	27.12	24	24	0.92	0.14	1.39	2.73	1.49	-0.66	80.12	98.13
<i>Carassius auratus</i>	56.50	13.56	29.94	-	24	0.91	0.06	1.73	2.69	1.56	-0.81	80.12	98.13
<i>Sinocyclocheilus rhinoceros</i>	56.50	15.82	27.68	-	24	0.92	0.01	1.59	2.75	1.55	-0.83	80.12	98.08
<i>Sinocyclocheilus grahami</i>	56.50	15.25	28.25	-	24	0.92	0.32	1.29	2.78	1.70	-0.54	80.12	97.88
<i>Oncorhynchus kisutch</i>	61.36	7.95	30.68	6	28	0.93	-0.15	2.03	2.95	1.84	-1.16	81.29	96.20
<i>Oncorhynchus mykiss</i>	61.36	7.39	31.25	6	24	0.93	0.68	2.46	3.21	1.86	-0.43	80.47	97.16
<i>Oncorhynchus nerka</i>	67.61	3.98	28.41	6	24	0.92	0.00	2.10	2.93	1.62	-0.95	80.12	96.40

<i>Oncorhynchus tshawytscha</i>	61.36	7.95	30.68	6	24	0.93	0.61	2.31	3.12	1.87	-0.47	81.07	97.26
<i>Oncorhynchus masou formosanus</i>	61.36	7.95	30.68	6	24	0.94	0.37	2.59	3.10	1.83	-0.74	80.70	97.93
<i>Salmo salar</i>	61.02	7.34	31.64	6	28	0.92	0.12	2.40	3.06	1.94	-0.98	80.12	97.98
<i>Maylandia zebra</i>	57.06	12.43	30.51	6	24	0.94	0.54	2.75	2.90	1.57	-0.51	83.04	97.04
<i>Oreochromis niloticus</i>	57.06	12.43	30.51	6	28	0.94	0.14	2.46	2.94	1.70	-0.89	84.21	97.63
<i>Haplochromis burtoni</i>	57.06	12.43	30.51	6	28	0.93	0.19	2.74	3.06	1.69	-0.89	82.46	97.44
<i>Anabas testudineus</i>	57.63	7.91	34.46	6	24	0.94	0.46	2.47	2.98	1.73	-0.59	83.63	98.03
<i>Betta splendens</i>	51.41	11.86	36.72	6	24	0.94	0.36	2.26	2.75	1.56	-0.59	81.87	96.55
<i>Perca flavescens</i>	66.10	6.78	27.12	6	24	0.94	0.48	2.69	3.06	1.62	-0.58	82.46	97.19
<i>Oryzias latipes</i>	56.50	10.73	32.77	6	24	0.93	0.26	2.88	3.10	2.16	-1.00	82.46	97.39
<i>Oryzias melastigma</i>	57.63	9.60	32.77	6	26	0.94	0.31	2.83	3.05	2.09	-0.92	83.04	97.44
<i>Astyanax mexicanus</i>	60.45	9.04	30.51	6	24	0.94	0.64	2.57	2.90	1.68	-0.41	82.46	97.98
<i>Pygocentrus nattereri</i>	66.10	3.95	29.94	6	26	0.93	0.53	2.68	2.82	2.04	-0.65	81.40	98.57
<i>Ictalurus punctatus</i>	64.97	6.21	28.81	6	24	0.94	0.50	2.16	2.74	1.63	-0.47	83.04	97.83
<i>Tachysurus fulvidraco</i>	59.32	12.99	27.68	6	24	0.94	0.67	2.36	2.78	1.75	-0.37	81.29	97.73
<i>Takifugu rubripes</i>	57.63	11.30	31.07	6	24	0.94	0.52	2.50	2.85	1.74	-0.54	82.46	97.78
<i>Denticeps clupeioides</i>	65.54	6.21	28.25	6	26	0.93	0.44	2.66	3.07	1.98	-0.73	80.70	98.28
<i>Paramormyrops kingsleyae</i>	63.84	4.52	31.64	6	24	0.94	0.75	2.32	2.86	2.02	-0.38	81.87	97.98
<i>Scleropages formosus</i>	61.02	9.04	29.94	6	26	0.94	0.39	2.25	2.78	1.96	-0.69	81.87	98.47
<i>Mastacembelus armatus</i>	61.02	6.78	32.20	6	26	0.95	0.24	2.47	2.88	1.98	-0.88	84.21	98.08
Average	58.62	10.17	31.20	7.0	24.28	0.93	0.35	2.28	2.92	1.82	-0.64	81.91	97.62

Table 4: Characteristics of input protein ferritin heavy chain subunit and predicted functional partners in selected 35 freshwater fish species via STRING 10.0

Index	ID	Term	P-value
1.	fth1a, fth1b, zgc:109934, ENSDARG00000094210, ENSDARG00000079846, zgc:92066, LOC558816, zgc:198419	Ferritin; Stores iron in a soluble, non-toxic, readily available form. Important for iron homeostasis. Iron is taken up in the ferrous form and deposited as ferric hydroxides after oxidation (177 aa)	
Predicted Functional partners			
1.	ctssb.1	Cathepsin Sb, tandem duplicate 1; Belongs to the peptidase C1 family (330 aa)	0.939
2.	LOC562139	Similar to chymotrypsinogen B1; Zgc-136461; Belongs to the peptidase S1 family (263 aa)	0.932
3.	Tfa	Serotransferrin; Transferrins are iron binding transport proteins which bind Fe(3+) ion in association with the binding of an anion, usually bicarbonate.	0.922
4.	ncoa4	Nuclear receptor coactivator 4 (576 aa)	0.916
5.	Aldoa, aldoa	Aldolase a, fructose-bisphosphate, b (364 aa)	0.914, 0.913
6.	ctssb.2	Cathepsin Sb, tandem duplicate 2; Belongs to the peptidase C1 family (330 aa)	0.912
7.	cnm2	Calponin; Thin filament-associated protein that is implicated in the regulation and modulation of smooth muscle contraction. It is capable of binding to actin, calmodulin, troponin C and tropomyosin. The interaction of calponin with actin inhibits the actomyosin Mg-ATPase activity	0.910
8.	aldoca	Aldolase C, fructose-bisphosphate, a (364 aa)	0.910
9.	Ctsd	Ctsd protein; Cathepsin D; Belongs to the peptidase A1 family (398 aa)	0.908
10.	hbbe2, hbbe3	Hemoglobin beta embryonic-2; Hemoglobin beta embryonic-3; Belongs to the globin family (147 aa)	0.907
11.	lta4h	Leukotriene A4 hydrolase; Belongs to the peptidase M1 family (611 aa)	0.906
12.	cst3	Cystatin C (amyloid angiopathy and cerebral hemorrhage); Belongs to the cystatin family (134 aa)	0.906
13.	mmp9	Matrix metalloproteinase 9; Belongs to the peptidase M10A family (680 aa)	0.905
14.	ilf2	Interleukin enhancer-binding factor 2 homolog; Appears to function predominantly as a heterodimeric complex with ILF3. This complex may function to regulate transcription of undefined genes (By similarity)	0.905
15.	Try	Trypsinogen 2; Trypsin; Belongs to the peptidase S1 family (247 aa)	0.904
16.	ENSDARG00000089043	Protein tyrosine phosphatase, non-receptor type 6; Belongs to the protein-tyrosine phosphatase family. Non-receptor class 2 subfamily	0.904

17.	b2ml, b2m	Beta-2-microglobulin; Component of the class I major histocompatibility complex (MHC). Involved in the presentation of peptide antigens to the immune system	0.904
18.	ba1	Hemoglobin subunit beta-1; Involved in oxygen transport from gills to the various peripheral tissues; Belongs to the 19.globin family	0.904
20.	Ctsh	Cathepsin H; Belongs to the peptidase C1 family (330 aa)	0.904
21.	LRG1	Si-dkey-90m5.4; Leucine-rich alpha-2-glycoprotein 1 (327 aa)	0.904
22.	nit2	Omega-amidase NIT2; Has a omega-amidase activity. The role of omega-amidase is to remove potentially toxic intermediates by converting alpha- ketoglutaramate and alpha-ketosuccinamate to biologically useful alpha-ketoglutarate and oxaloacetate, respectively	0.904
24.	hbbe1.1	Novel protein similar to zebrafish embryonic 1 beta-globin (BE1); Hemoglobin beta embryonic-1.2; Belongs to the globin family	0.904
25.	ENSDARG00000089963	Hemoglobin beta embryonic-1.1; Belongs to the globin family (88 aa)	0.903
26.	asah1b, asah1a	N-acylsphingosine amidohydrolase (acid ceramidase) 1b (395 aa), (390 aa)	0.903
27.	pglyrp5	Peptidoglycan recognition protein 5 (238 aa)	0.903
28.	ENSDARG00000051890	Haptoglobin (308 aa)	0.902
29.	CRISP3	Cysteine-rich secretory protein 3; Belongs to the CRISP family (245 aa)	0.902
30.	timp2b	Tissue inhibitor of metalloproteinase 2b (265 aa)	0.902
31.	tnfaip6	Tumor necrosis factor, alpha-induced protein 6 (271 aa)	0.902
32.	timp2a	Tissue inhibitor of metalloproteinase 2a (220 aa)	0.902
33.	zgc:153129	Cystatin 14b, tandem duplicate 1; Zgc-153129 (101 aa)	0.902
34.	qsox1	Quiescin Q6 sulfhydryl oxidase 1 (778 aa)	0.902
35.	Cda	Cytidine deaminase; This enzyme scavenges exogenous and endogenous cytidine and 2'-deoxycytidine for UMP synthesis; Belongs to the cytidine and deoxycytidylate deaminase family	0.902
36.	zgc:92880	Hemoglobin, beta adult 2; Zgc-92880; Belongs to the globin family (147 aa)	0.902
37.	cst14a.1, zgc:56530	Cystatin 14a, tandem duplicate 1; Si-ch211-161h7.6 (100 aa), Cystatin 14a, tandem duplicate 2; Zgc-56530 (100 aa)	0.902
38.	ENSDARG00000078114, cfp	Complement factor properdin (437 aa)	0.901
39.	ptx3a	Pentraxin 3, long a (447 aa)	0.901
40.	ENSDARG00000087157	Calcium activated nucleotidase 1b (411 aa)	0.901
41.	cyfip1	Cytoplasmic FMR1-interacting protein 1 homolog; Involved in formation of membrane ruffles and lamellipodia protrusions and in axon outgrowth. Binds to F-actin but not to RNA (By similarity)	0.901
42.	cant1a	Calcium activated nucleotidase 1a (404 aa)	0.901
43.	idh1	Isocitrate dehydrogenase 1 (NADP+), soluble; Belongs to the isocitrate and isopropylmalate dehydrogenases family	0.901
44.	LOC569427, LOC569381	Si-ch211-173a9.7; Sci-ch211-173a9.6; Olfactomedin 4 (466 aa)	0.900
45.	armac8	Armadillo repeat containing 8 (674 aa)	0.900
46.	Qpct	Glutaminy-peptide cyclotransferase (365 aa)	0.900

Table 5: Observed KEGG pathways and PFAM protein domains in protein-protein interaction of ferritin heavy chain subunit in selected 35 freshwater fish species.

Index	ID	Term	Count in gene set	P-value
KEGG Pathway				
1.	dre04216	Ferroptosis	5 of 42	4.58e-07
2.	dre00030	Pentose phosphate pathway	3 of 33	0.00038
3.	dre00051	Fructose and mannose metabolism	3 of 43	0.00054
4.	dre04145	Phagosome	4 of 142	0.00055
5.	dre04142	Lysosome	4 of 136	0.00055
6.	dre00590	Arachidonic acid metabolism	3 of 48	0.00055
7.	dre04210	Apoptosis	4 of 159	0.00062
8.	dre00010	Glycolysis / Gluconeogenesis	3 of 74	0.00093
9.	dre01230	Biosynthesis of amino acids	3 of 82	0.0011
10.	dre01200	Carbon metabolism	3 of 125	0.0033
UniProt				
1.	KW-0409	Iron storage	8 of 11	2.35e-16
2.	KW-0408	Iron	12 of 261	1.42e-12
3.	KW-0645	Protease	11 of 321	2.55e-10
4.	KW-0732	Signal	21 of 3209	6.27e-07
5.	KW-0704	Schiff base	3 of 9	1.01e-05
6.	KW-0720	Serine protease	4 of 92	0.00020
7.	KW-0490	MHC I	2 of 3	0.00020
8.	KW-0324	Glycolysis	3 of 30	0.00020
9.	KW-0479	Metal-binding	15 of 2762	0.00037
10.	KW-0964	Secreted	6 of 400	0.00041
11.	KW-1015	Disulfide bond	10 of 1364	0.00061
12.	KW-0788	Thiol protease	3 of 85	0.0021
13.	KW-0561	Oxygen transport	2 of 19	0.0022
14.	KW-0456	Lyase	3 of 88	0.0022
15.	KW-0482	Metalloprotease	2 of 76	0.0260
PFAM protein Domains				
1.	PF00210	Ferritin-like domain	8 of 12	3.91e-16
2.	PF00274	Fructose-bisphosphate aldolase class-I	3 of 5	7.78e-06
3.	PF08246	Cathepsin propeptide inhibitor domain (I29)	3 of 27	0.00037
4.	PF00965	Tissue inhibitor of metalloproteinase	2 of 3	0.00037
5.	PF00112	Papain family cysteine protease	3 of 30	0.00037
6.	PF02244	Carboxypeptidase activation peptide	2 of 7	0.00096
7.	PF00089	Trypsin	4 of 177	0.0023
8.	PF04389	Peptidase family M28	2 of 15	0.0027
9.	PF00031	Cystatin domain	2 of 15	0.0027
10.	PF00246	Zinc carboxypeptidase	2 of 24	0.0047
11.	PF00042	Globin	2 of 23	0.0047
12.	PF09342	Domain of unknown function (DUF1986)	2 of 50	0.0168
13.	PF07654	Immunoglobulin C1-set domain	2 of 79	0.0366
SMART Domain				
1.	SM00848	Cathepsin propeptide inhibitor domain (I29)	3 of 28	0.00058
2.	SM00645	Papain family cysteine protease	3 of 30	0.00058
3.	SM00206	Tissue inhibitor of metalloproteinase family.	2 of 3	0.00058
4.	SM00020	Trypsin-like serine protease	4 of 141	0.00084
5.	SM00043	Cystatin-like domain	2 of 16	0.0023
6.	SM00631	Zn_pept	2 of 21	0.0032
7.	SM00407	Immunoglobulin C-Type	2 of 63	0.0213

Conclusion

Ferritins are regarded as acute-phase proteins that respond to stress and inflammation. Ferritin synthesis is stimulated with the increase of iron level in vertebrates. The increased ferritin expression during iron overload reflects the need to store excess iron, minimizing free iron toxicity which in turn can inhibit the bacterial infection in fish species. In this study ferritin heavy chain protein in freshwater fish's shows that the protein is hydrophilic, acidic, thermo stable, unstable except four fish species, homologous to mouse ferritin heavy chain. 10 enriched pathways of KEGG, 15 UNIPROT, 13 PFAM Domains and 7 SMART Protein Domains were obtained by the protein-protein interaction analysis. The investigation of ferritin heavy chain subunit in freshwater fishes revealed the

structural features and their association in detoxification and iron homeostasis. It will be helpful in unraveling the mechanisms of stress management against the iron toxicity in ponds used for aquaculture in North East Region.

Acknowledgement

The author is thankful to the Dean, College of Fisheries, Lembucherra, Tripura, for encouragement and moral support. This work has been carried out under DBT funded Project "Establishment of Bioinformatics Infrastructure Facility for Biology Teaching through Bioinformatics" and hence financial support by the Department of Biotechnology, Ministry of Science and Technology, Government of India, New Delhi.

References

- Arosio P, Levi S. Cytosolic and mitochondrial ferritins in the regulation of cellular iron homeostasis and oxidative damage. *Biochim Biophys Acta*. 2010; 1800:783-92.
- Crichton RR, Charloteauxwauters M. Iron transport and storage. *Eur J Biochem*. 1987; 164:4850-506.
- Aisen P, Listowsky I. Iron transport and storage proteins. *Annu Rev Biochem*. 1980; 49:357-393.
- Anderson GJ, Frazer DM. Hepatic iron metabolism. *Semin Liver Dis*. 2005; 25:420-432.
- Aisen P, Listowsky I. Iron transport and storage proteins. *Annu Rev Biochem*. 1980; 49:357-393.
- Harrison PM, Arosio P. The ferritins: molecular properties, iron storage function and cellular regulation. *Biochim Biophys Acta*. 1996; 1275:161-203.
- Torti FM, Torti SV. Regulation of ferritin genes and protein. *Blood*. 2002; 99:3505-3516.
- Arosio P, Levi S. Cytosolic and mitochondrial ferritins in the regulation of cellular iron homeostasis and oxidative damage. *Biochim Biophys Acta*. 2010; 1800:783-792.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H *et al*. The protein data bank. *Nucleic Acids Research*. 2000; 28(1):235-242.
- Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD *et al*. Protein identification and analysis tools on the ExpASY server. In: Walker JM, editor. *The proteomics protocols handbook*. Humana Press, 2005 571-607.
- Sen TZ, Jerigen R, Kloczkowski A (2005). GOR V server for protein secondary structure prediction. *Bioinformatics*, 21(11), 2787-2788.24.
- Guex N, Peitsch MC, Schwede T (2009). Automated comparative protein structure modeling with SWISS-MODEL and SwissPdbViewer: a historical perspective. *Electrophoresis*, 1, S162-S173.
- Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, *et al*. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acid Research*, 42, W252-W258.
- Lovell c, Davis IW, Arendall WB, De Bakker PIW, Word JM, *et al*. (2003). Structure validation by C α geometry: phi, psi and C β deviation. *Proteins*, 50(3), 437-450.29.
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res*. 43:D447-D452.
- Kyte J, Doolittle RF. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol*. 157:105-132.
- Ikai A. (1980). Thermostability and aliphatic index of globular proteins. *J Biochem*. 88:1895-1898.
- Guruprasad K, Reddy BB, 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. *Protein Eng*. 4:155-161.
- Rogers S, Wells R, Rechsteiner M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science*. 234:364-368.
- Levitt M and J Greer (1977). Automatic identification of secondary structure in globular proteins. *J. Mol. Biol.* (1977) 114, 181-293
- Lippert, T., Rarey, M. (2009). Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *Journal of Cheminformatics*, 1, 13.
- Davies, D.R. (1964). A correlation between amino acid composition and protein structure *J. Mol. Biol.*, 9, 605-609.
- Guzzo, A. V. (1965). The influence of amino acid sequence on protein structure. *Biophys. J.* 5, 809-822.
- Prothero, J. W. (1966). Correlation between the distribution of amino acids and alpha helices. *Biophys. J.* 6, 367-370.
- Cook, D.A. (1967). The relation between amino acid sequence and protein conformation. *J. Mol. Biol.*, 29 (1967), pp. 167-171.
- Chou, P.Y. and Fasman, G. (1974). Conformational parameters of amino acids in helical, β -sheet, and random coil regions calculated from proteins. *Biochemistry*, 13: 211-222.
- Lim, V.I. (1974). Algorithm for prediction of alpha helical and Beta structural regions of globular proteins. *J. Mol. Biol.* 88:873-894
- Yee, D.P., Chan, H.S., Havel, T.F. & Dill, K.A. (1994). Does compactness induce secondary structure? A study of poly-alanine chains computed by distance geometry. *J. Mol. Biol.* 241, 557-573. 65.
- Dill, K.A., Bromberg, S., Yue, K., Fiebig, K.M., Yee, D.P., Thomas, P.D. & Chan, H.S. (1995). Principles of protein folding: A perspective from simple exact models. *Protein Sci.* 4, 561-602. 66.
- Pitsyn, O.B. (1995). Structures of folding intermediates. *Curr. Opin. Struct. Biol.* 5, 74-78.
- Lawson DM, Artymiuk PJ, Yewdall SJ, Smith JM, Livingstone JC, Treffry A, Luzzago A, Levi S, Arosio P, Cesareni G. (1991). Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts [J]. *Nature*, 349, 541-544.
- Ha Y, Shi D, Small W, Theil EC, Allewell NM. (1999). Crystal structure of bullfrog M ferritin at 2.8 Å resolution: analysis of subunit interactions and the binuclear metal center [J]. *Journal of Biological Inorganic Chemistry*, 4(3), 243-256.
- Macedo S, Romao CV, Mitchell E, Matias PM, Liu MY, Xavier AV, LeGall J, Teixeira M, Lindley P, Carrondo MA. (2003). The nature of the di-iron site in the bacterioferritin from *Desulfovibrio desulfuricans* [J]. *Natural Structural Biology*, 10(4), 285-290.
- Sohpal, Kumar, Vipin, Dey, Apurba and Singh, Amarपाल (2013) Z-score for homology modeling and structure prediction of globular protein of homosapiens & plant. *International journal of advance research in science and engineering*. IJARSE, Vol. No.2, ISSN-2319-8354(E).
- Benkert, P and Künzli, M (2009). QMEAN server for protein model quality estimation. *Nucleic Acids Res.* 37 (suppl 2), pp W510-W514.
- Arnold, K., Bordoli, L., Kopp, J. (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, *Bioinformatics Volume 22* (2), pp 195-201.2006.
- Benkert, P. and Biasini, M (2011). Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27 (3), pp 343-350.
- Schwartz AS, Yu J, Gardenour KR, Finley RL Jr, Ideker T. (2009). Costeffective strategies for completing the interactome. *Nat Methods*. 6:55-61.
- Bairoch A, Boeckmann B. (1994). The SWISS-PROT protein sequence data bank: current status. *Nucleic Acids Res.* 22:3578-3580.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT. (2000). Gene ontology: tool for the unification of biology. *Nat Genet.* 25:25-29.
- Ouzounis CA, Coulson RM, Enright AJ, Kunin V, Pereira-Leal JB. (2003). Classification schemes for protein structure and function. *Nat Rev Genet.* 4:508-519.
- Lee D, Redfern O, Orengo C. (2007). Predicting protein function from sequence and structure. *Nat Rev Mol Cell Biol.* 8:995-1005. as *Argulus* is identified.