Journal of Entomology and Zoology Studies

Available online at www.entomoljournal.com

E-ISSN: 2320-7078
P-ISSN: 2349-6800
www.entomoljournal.com
JEZS 2020; 8(2): 540-550

Received: 19-01-2020
Accepted: 23-02-2020

Rumpi Ghosh
College of Fisheries, Lembucherra, Tripura (W), India

Anil Datt Upadhyay
College of Fisheries, Lembucherra, Tripura (W), India

Ajit Kumar Roy
College of Fisheries, Lembucherra, Tripura (W), India

Ajit Tiwari
College of Fisheries, Lembucherra, Tripura (W), India

Corresponding Author:
Rumpi Ghosh
College of Fisheries, Lembucherra, Tripura (W), India

Structural and functional analysis of cytochrome b protein of Indian major carps *Labeo rohita*

Rumpi Ghosh, Anil Datt Upadhyay, Ajit Kumar Roy and Ajit Tiwari

Abstract
Mitochondrial Cytochrome *b* is a monomeric, monoheme protein which is found in eukaryotic cells that facilitate electron transport during cellular respiration. In this present study cytochrome b protein of *Labeo rohita* was selected for a detailed computational investigation to exploit its physicochemical characteristics, structural properties including 3D model, model quality analysis, phylogenetic assessment and functional analysis using several available standard Bioinformatics tools. Physiochemical characterization was performed for identification of protein by computing theoretical isoelectric point (pI), molecular weight, extinction coefficient, instability index and aliphatic index etc. Secondary structure assessment of cytochrome b protein of *Labeo rohita* used PSIPRED and revealed a greater percentage of residues as the alpha helix. This protein having an average molecular weight about 42781.76 Da was found thermostable and alkaline and belonging to a metal enzyme. After prediction, the structure has been validated through Ramachandran plot and Rampage validation tools, this homology modelling based structure will provide an in site to its functional aspects. And then we used different standard server for active site prediction and found an active site. After that, we get the protein-protein interaction site through the string server. And finally submit this predicted 3D model in PDBM server and we get an accession no. PMID: PM0081835.

Keywords: EXPASY, cytochrome b, template library, PSIPRED, active site and dog scorer, etc.

Introduction
The name 'cytochrome' was introduced by Kellin [24] in 1925 to describe a group of intracellular hemeproteins that undergo oxidation-reduction and upon reduction, exhibit intense absorption bands between 510 and 615 nm. As currently used, the name appears to include all intracellular hemeproteins except haemoglobin, myoglobin, peroxidases, catalase, tryptophan 2, 3-dioxygenase, heme-thiolate proteins (P-450) and nitrite and sulphite reductases. Cytochrome b is a protein found in the mitochondria of eukaryotic cells. It functions as part of the electron transport chain and is the main subunit of transmembrane cytochrome bc1 and b6f complexes [2, 3]. In-plant chloroplasts and cyanobacteria, there is an analogous protein, cytochrome b6, a component of the plastoquinone-plastocyanin reductase, also known as the b6f complex. These complexes are involved in electron transport, the pumping of protons to create a proton-motive force (PMF). This proton gradient is used for the generation of ATP. These complexes play a vital role in cells [8]. Cytochrome b/b6 is an integral membrane protein of approximately 400 amino acid residues that probably has 8 transmembrane segments [9]. In plants and cyanobacteria, cytochrome b6 consists of two subunits encoded by the pet B and pet D genes. Cytochrome b/b6 non-covalently binds two heme groups, known as b562 and b566. The first scientific papers to employ the polymers chain reaction identified primer regions of the cytochrome b locus b that were broadly conserved across vertebrate taxa. The cyt b used in intra & interspecific molecular systematic studies [1]. Cytochrome is a protein that can transfer electrons with a chemical group called a heme group. The heme groups of cytochrome are similar to those of haemoglobin. Both have the same basic ring structure called a porphyrin ring is an iron atom (Fe). Four nitrogen's surround the iron atom. Many carbons and hydrogen's surround and connect to the four nitrogen's. The whole structure forms what's known as a heme group. The cytochrome b locus has in consequence been extensively used in intra & interspecific molecular systematic studies. Cytochrome *b*. Cytochrome with protoheme (the iron chelate of protoporphyrin IX), [23] as a prosthetic group. It contains one mole each of heme and FMN as prosthetic groups and acts as an L-lactate dehydrogenase. It is also called

flavocytochrome *b*. The cytochrome is part of the membrane-bound NADPH-oxidoreductase, which reduces O_2 to H_2O_2 and O_2 Cytochrome *b* with protoheme the iron chelate of protoporphyrin IX, see [23] as a prosthetic group but which lack a covalent bond between the porphyrin and the protein. The cytochrome bc1 complex is an energy-transducing, electron-transfer enzyme located in the inner mitochondrial membrane of oxygen-utilizing eukaryotic cells, where it participates in cell respiratory chain. A functionally similar but structurally simpler version of the bc_1 complex is located in the plasma membrane of many, but not all, bacteria, where it takes part in respiration, denitrification, nitrogen fixation, and cyclic photosynthetic electron transfer, depending on the species. In all of these organisms, the bc_1complex oxidizes a membrane-localized quinol and reduces a water-soluble, *c*-type cytochrome and links this redox reaction to translocation of protons across the membrane in which the bc_1 complex resides. The bc_1 complexes from mitochondria of several species have been crystallized and the mechanism of the enzyme is generally well understood, although some questions remain outstanding. Fish species of carp family is *Labeo rohita* or rui or roho which are basically found in rivers in South Asia. It is a Great omnivore and widely used in aquaculture. This roho is a large, silver-colored cyprinid fish with a well-formed head. Adults can reach a maximum weight of 45 kg (99 lb) and a maximum height of 2 m (6.6 ft).This species is omnivore because of specific food preferences at different stages of life, During the early stages of life it eats zooplankton when as it grows eats phytoplankton, and in adult is a herbivorous column feeder, eating mainly phytoplankton and submerged vegetation, it has modified thin gill rakers. It suggests that it feeds by blocking water. Rohu reaches sexual maturity between two and five years old. They are usually spawned in the middle of heavy rainy season, up to the middle of floodplains. The rohu season is usually encountered by south western animals. Spawn can be collected in rivers and raised by tanks and ponds [33]. Rohu is very commonly eaten in Bangladesh, Nepal, Pakistan and the Indian states of Tripura, Nagaland, Bihar, Odisha, Assam, West Bengal, Andhra Pradesh and Uttar Pradesh [34]. Fish species of carp family is *Labeo rohita* or rui or roho which are basically found in rivers in South Asia. Labeo genus is a carp in the family of Cyprinidae. They are habitat in freshwater areas in tropical and subtropical regions of Africa and Asia. It has common labeos in the lower Labeoninae and is often included in the Cyprininae as a Labehlon tribe. If the Labeoninae are adopted separately, the name Labehlonon is a tribal name in this sub-labeos. If the Labehlono are considered a tribe of the Cyprininae, the labeos are placed in a subordinate area of the Labeoina. Labeos are apparently similar to the "freshwater shark" of the genus. Epalzeorhynchos, also part of the Labeoninae (or Labehlon), but are not closely related. Labeos are larger, and have a more rounded body, as they swim much more freely than benthic ones like the Epalzeorhynchos. It is a Great omnivore and widely used in aquaculture. This roho is a large, silver-colored cyprinid fish with a well-formed head. Adults can reach a maximum weight of 45 kg (99 lb) and a maximum height of 2 m (6.6 ft).This species is omnivore because of specific food preferences at different stages of life, During the early stages of life it eats zooplankton when as it grows eats phytoplankton, and in adult is a herbivorous column feeder, eating mainly phytoplankton and submerged vegetation, it has modified thin gill rakers. It suggests that it feeds by blocking water. Rohu reaches sexual maturity between two and five years old. They are usually spawned in the middle of heavy rainy season, up to the middle of floodplains. The rohu season is usually encountered by south western animals. Spawn can be collected in rivers and raised by tanks and ponds [33]. The three Indian major carp species used in carp polyculture systems Rohu is very commonly eaten in Bangladesh, Nepal, Pakistan and the Indian states of Tripura, Nagaland, Bihar, Odisha, Assam, West Bengal, Andhra Pradesh and Uttar Pradesh [34]. Body almost equally, on average, body with cycloid scales, head without scales; palate compressed well, processing across the mouth, without lateral lobe; eye dorsolateral in position, is invisible to the outside of the head; mouth small and inferior; thick and convex lips with a depressed internal mass on each lip, lobate or whole; a couple of tiny barbels hidden in a separate storage area; no teeth in the jaws; pharyngeal teeth in three rows; the upper jaw that extends to the front of the eye; light (uncontrolled) dorsal fin rays of three or four rays, of dorsal fin rays 12 to 14; dorsal fin inserted midway between snout tip and base of caudal fin; lateral pectoral and pelvic wings; pectoral Fin with no osseous spine; caudal Fin closely fenced; lower lips often associated with isthmus with a narrow or wide bridge; scale before chart 12-16; lateral line distinct, complete and running along the medial line of the caudal peduncle; lateral line scales 40 to 44; line of lateral transverse scale — six or six and a half lines between lateral line and pelvic fin base; snout not truncate, except for the lateral lobe; bright colour on the back, lining on the flanks and voice.

Materials and Methods

Retrieval of sequences: For the analysis of the amino acid sequence of mitochondrion cytochrome b of *Labeo rohita* was downloaded from NCBI (AAR26377), the sequence length reported to be 380 amino acids residues.

Physical and chemical parameters calculation: For calculating of physical and chemical parameters Expasy's prot param server was used. For Physico-chemical characterization, theoretical isoelectric point (pI), Molecular Weight, Total number of positive and negative residues and Extinction Coefficient was determined; this method given by Gill and Hippel and integrated into Expasy's prot param server. Instability index, Grand average hydropathicity and aliphatic index were also computed by the Expasy's prot param [12, 10, 7].

Secondary structure prediction method (PSIPRED): The secondary structure prediction of COII was done by using PSIPRED (http://www.ebi.ac.uk/pdbsum/).PSI-blast-based structural prediction is a technique used to investigate protein composition. It is a server-side program, which contains a website that acts as an interface - an endpoint, which can predict the formation of a protein base (beta sheets, alpha helixes and coils) in the main sequence. PSIPRED is available as a web service and software. It allows for modification but emphasizes the provision of freeware by restricting the distribution of software and its effects [7, 8].

Predicted protein model evaluation and submission: Predicted protein model of cytochrome b of *Achromobacter piechaudii* was evaluated and verified from both QMEAN and SAVES server (http://nihserver.mbi.ucla.edu/SAVES). Ramachandran plot, verify 3D, and ERRAT [28] were evaluated at SAVES. The model is specified (PDB) format

was submitted to the Protein Model Database (PMDB-Http://bioinformatics.cineca.it/PMDB/).

Geno 3d server: This server was used to generate protein 3D model. The strategy used in Geno3D compares protein composition with edge measurements (distances and dihedral) satisfaction. Geno3D is most frequently used for homology or comparative protein structure modelling; the user provides sequence to be modelled (query) which is compared using PSI-BLAST method [11] against a protein sequence database issue from PDB.

Active site prediction of cytochrome b: Protein-Protein interactions are important from the aspect of the cellular function and determining how these proteins interact with their ligands and other small molecules [26]. The active site of cytochrome protein in *Labeo rohita* was predicted using Dog Site Scorer server /active site prediction server based on grid-based function prediction method [27]. DoGSiteScorer (Volkamer *et al.*, 2012) provides the functionality to detect potential binding Pockets and sub pockets of a protein of interest. Subsequently, it analyzes the geometric and Physico-chemical properties of these pockets and estimates druggability with the aid of a support vector machine (SVM). DoGSiteScorer has been evaluated on a large dataset containing 1069 structures and shows prediction accuracies of 88%. Thus, the method provides valuable information for target assessment and can now be accessed through a web server. DoGSiteScorer grid-based method uses a Difference of Gaussian filter to detect potential binding pockets - solely based on the 3D structure of the protein - and splits them into sub pockets. Global properties, describing the size, shape and chemical features of the predicted (sub) pockets are calculated. Per default, simple druggability. The score is provided for each (sub) pocket, based on a linear combination of the three descriptors describing volume, hydrophobicity and enclosure. Furthermore, a subset of meaningful descriptors is incorporated in a support vector machine (libsvm) to predict the (sub) pocket druggability score (values are between zero and one). The higher the score the more druggable the pocket is estimated to be. 1 [25].

Protein-protein interaction study: To know the interaction about Cytochrome b of *Labeo rohita* with other closely related proteins STRING v 10.0 (http://string-db.org/) server [31] was used. Cytochrome b of protein from *Labeo rohita* was selected as a query sequence and functional protein association network was generated. Moreover, the query sequence was also analysed to determine the family which the protein belongs.

Results and Discussion

Expasy's Prot param tool: Expasy's prot param tool was used to predict the physicochemical properties of the protein model. The Expasy's Prot param result was summarized in Table (1) for cytochrome b. [1]. The number of amino acid residues 380aa. In cytochrome b the number of amino acid residues is 380, the molecular weight of the protein 42781.76 Da, theoretical pI 6.86 which also indicate that the protein is negatively charged and precipitated in acidic medium, the total number of negatively charged residues 18, and the total number of positively charged residues 17. The extinction coefficient 92485 absorbed at 280 nm in water. Estimated half-life 30, considered human, the instability index is related to the half-life of the protein yeast and Escherichia coli cells and it was 30h>20h>and 10h, respectively. Instability index 34.89, this classifies that the protein is stable. Aliphatic index 120.89, that indicate this protein has a wide range of temperature stability. Grand average hydropathycity (GRAVY) 0.739, which also indicate greater hydrophobicity and the protein solubility in water. The result was summarized in table 1.

Secondary structure: The secondary structure of mitochondrion Cytochrome b is predicted by using PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred) as present in figure 1.The secondary structure of Cytochrome b revealed the presence of 16 coil Regions and 8 Beta-Strand, a total of 6 alpha-helix regions is highlighted. On the other hand, in *Labeo rohita* in the same protein there is found coil Regions 12 Beta-Strand 1 and total alpha-helix regions 9.

Template searching: For homology modelling template was searched through Swiss model template library and template 1be3.1.C was selected (Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex). The experimental structures used for the construction of the model where cytochrome BC1 complex from Bovine (1be3.1.C) which had 76.78 % identity with target protein which was uses as a template and build a comparative homology modelling. Detailed output presented in figure 3. In this experiment figure A. represents only 3d structure for cytochrome b and B. represent for 1be3.1 template 3D structure and C is validated 3D modelled for template 1be.1 by Swiss model.

3D structure prediction: Swiss model template library Studies suggest that the accuracy of homology model is solely dependent on the template and when the sequence identity is less than 30% the reliability of model decreases [6]. Tertiary or 3D structure prediction was done using the Swiss model automated mode for homology modelling. Swiss model server searched for the solved templates with similar sequences, Best templates were aligned with the target amino acids sequence. Templates with best E-value, percentage similarities and a maximum number of query sequence covered were selected for homology modelling [9]. Predicted the structure of cytochrome b protein is given by A, B and C according to alphabetical order.

Table 1: cytochrome b primary result by expasy prot param server tools.

Composition	Cytochrome b
Total number of negatively charged residues	18
Total number of positively charged residues (Arg + Lys)	17
carbon	2043
Hydrogen	3081
Nitrogen	473
oxygen	500
Sulphur	16
Theoretical pI	6.86
Molecular weight	42781.76
Extinction coefficient	92485
Instability index	34.89
Amino acid residues	380
Grand average hydropathycity (GRAVY)	0.739

Table 2: Result of Ramachandran plot

Composition	number	percentage
Number of residues in most favoured region	1677	(82.7%)
Number of residues in the allowed region	280	(13.8%)
Number of residues of outlier region	70	(3.5%)
Number of residues of disallowed region	0	(0.0%)

Table 3: Phosphorylation site with different kinases

Amino acid	site	Kinase	*Score*
Y	5	EFGR	0.499
T	6	PKG	0.499
T	18	GSK3	0.453
T	24	GSK3	0.454
T	44	GSK3	0.455
S	48	PKC	0.872
T	52	PKC	0.684
S	70	GSK3	0.485
S	73	cdc2	0.502
T	74	PKC	0.507
S	80	PKA	0.578
T	88	cdc2	0.463
T	92	CaM-II	0.466
Y	100	INSR	0.385
T	101	cdc2	0.507
S	102	unsp	0.823
S	105	GSK3	0.462
T	106	PKC	0.617
S	109	Unsp	0.567
S	110	cdc2	0.468
T	112	PKC	0.529
Y	120	EFGR	0.518
S	133	cdc2	0.479
Y	140	INSR	0.505
Y	148	INSR	0.445
Y	149	INSR	0.494
S	151	cdc2	0.509
Y	152	INSR	0.415
T	154	INSR	0.399
T	157	cdc2	0.479
Y	171	PKG	0.455
S	176	INSR	0.408
T	184	CaM-II	0.466
T	189	PKC	0.484
S	192	Cdc2	0.474
Y	196	PKG	0.462
S	200	INSR	0.456
T	214	PKA	0.470
T	219	PKC	0.504
T	221	PKC	0.536
T	235	CaM-II	0.692
T	238	CaM-II	0.470
S	248	GSK3	0.896
S	250	PKC	0.434
S	258	Unsp	0.459
Y	264	PKG	0.527
T	268	INSR	0.493
Y	270	unsp	0.515
T	271	INSR	0.394
S	285	CaM-II	0.397
T	291	p38MAPK	0.491
T	302	p38MAPK	0.515
Y	309	cdk5	0.394
Y	318	INSR	0.397
S	323	INSR	0.491
S	328	PKA	0.515
T	342	Cdc2	0.895
S	354	PKC	0.515

T	355	Cdc2	0.895
T	361	Unsp	0.552
T	366	PKA	0.966
T	366	GSK3	0.540
S	372	PKC	0.442
Y	381	CaM-II	0.561
	400	SRC	0.472
	403		0.423

Ramachandran plot assessment

Protein model validity: A Rampage use for Ramachandran plot assessment. (also known as a Ramachandran diagram or a [φ,ψ] plot), is a way to visualize energetically allowed regions for backbone dihedral angles ψ against φ of amino acid residues in protein structure. The geometrical and structural consistency of the predicted model was evaluated by different approaches. The φ and ψ distributions of Ramachandran plot analysis using rampage. Before understanding the Ramachandran plot it is very important to understand the structure of a peptide bond which is rigid and planar. Simply a Ramachandran plot is a plot to visualize energetically allowed regions for a polypeptide backbone torsion angles psi (ψ) against (phi) φ. From the figure 6. it can be seen that out of 380 aa sequences 82.7% residues fall in the most favoured region, residues in the allowed region 13.8%, Residues of outlier region 3.5% whereas no one residues were fallen in the disallowed region. It also indicates that the degree of accuracy in the predicted structure. The Ramachandran plot assessment details in figure 6. [22]. Ramachandran plot is a way to visualize backbone dihedral angles ψ against φ of amino acid residues in protein structure. A Ramachandran plot can be used in two somewhat different ways. One is to show in theory which values, or conformations, of the ψ and φ angles, are possible for an amino-acid residue in a protein. A second is to show the empirical distribution of data points observed in a single structure:

Active site prediction of cytochrome b protein of *Labeo rohita*: The potential binding site and sub pockets are detected through Dog scorer site**.** Dog Site Scorer is a newly developed automatic tool combining pocket prediction, characterization and druggability estimation. DoG Site Scorer can be customized to work on the pocket and sub pocket level. Also, the drug ability estimation for pockets can be switched on (Fig. 7). DoGSiteScorer web server provides an easy to use interface to predict pockets and sub pockets of a protein structure of interest. Furthermore, key properties characterizing the pocket and drug ability estimations are supplied. Protoss is a fully automated hydrogen prediction tool for protein–ligands complexes. It adds missing hydrogen atoms to protein structures (PDB format) and detects reasonable protonation states, tautomers, and hydrogen coordinates of both protein and ligands molecules. Protoss investigates hydrogen bonds, metal interactions and repulsive atom contacts for all possible states and calculates an optimal hydrogen bonding network within these degrees of freedom**.** The active site of amino acid residues of cytochrome b analysis through the active site prediction server was found to be Threonine, alanine, histidine, phenylalanine, leucine, Isoleucine, asparagine, tryptophan, Tyrosine, glutamic acid, glutamine, cysteine, valine, serine, glycine, Arginine, glutamic acids and proline.

Phosphorylation site Prediction: Phosphorylation site has been located in the NetPhosK 3.1L Server. Different kinases like EGFR, PKG, CAM II, INSR, GSK3, PKC, cdc2, p38MAPK, SRC, cdk5, PKA, and UNSP are involved in the Phosphorylation of protein. The highest score was predicted for the site 403 tyrosine and the highest score is 0.966 having Threonine residue. After analysis, the % composition of amino acids has been obtained and is shown. The latter score is just barely above the threshold (0.500) and indicates that the confidence for this site being a true Phosphorylation site is quite low that shown in figure.8

String: In this view, the colour saturation of the edges represents the confidence score of a functional association. In this experiment number of nodes 11 and number of edges 52, the edges represent the predicted functional associations and average nodes degree 9.45, the average node degree is somehow many interactions that a protein has on the average in the network. Average local clustering coefficient 0.952, the clustering coefficient is a measure of how connected the nodes in the network. And PPI (Protein-protein Interaction) enrichment p-value <1.0e-16, a small PPI enrichment p-value indicate that the nodes are not random and that the observed number of edges is significant. The network view summarizes the network of predicted associations for a particular group of proteins. Network nodes represent proteins splices isoforms or post-translational modifications are collapsed, i.e. each node represents all the proteins produced by a single protein-coding gene locus. Highly connected networks have high values. Node coloured modes query proteins and first shell of interactors and white nodes second shell of interactors. The expected number of edges is to be expected if the nodes were to be selected at random. An edge drowns with up to 7 differently coloured lines – these lines represent the existence of the seven types of evidence used in predicting the associations. Redline indicates the presence of fusion evidence, Greenline –neighbourhood evidence, Blueline –Co-occurrence evidence, Yellow line – text mining evidence, Light blue line-database evidence, Blackline- co-expression evidence. Edges represent protein-protein associations are meant to be specific and meaningful, i.e. proteins jointly contribute to a shared function; this does not necessarily mean they are physically binding to each other. This experiment result shown in the figure.10 and 11.

Interaction records: Each interaction record contains two or more proteins which have been experimentally shown to bind, or to be in the same protein complex shown figure. 8and 9. Records are imported from several primary interaction databases and made non-redundant so that each record is only shown once. Ambiguous Proteins: -when interaction records are transferred from other organisms, ambiguities can happen: If the orthology/homology situation is unclear, a source protein in the foreign organism may be mapped to several proteins in your organism of interest.

Functional analysis

Functional analysis revealed ten potential interacting partners

of pet B in the protein interaction network as resolved by STRING analysis (fig: 8 & 9). The query protein pet B appeared to contain. The closest interacting protein having the shortest node was found CoxB, a preprotein translocase subunit while the distant interacting protein was found to be Cco B.

Protein model database: Finally the generated model for cytochrome b was successfully submitted in the Protein Model Database (PMDB) having the PMID: PM0081835 (Figure.10) [30].

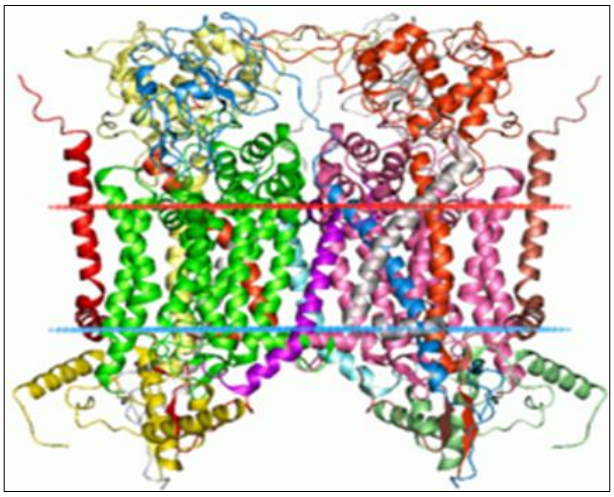


Fig 1: Cytochrome b

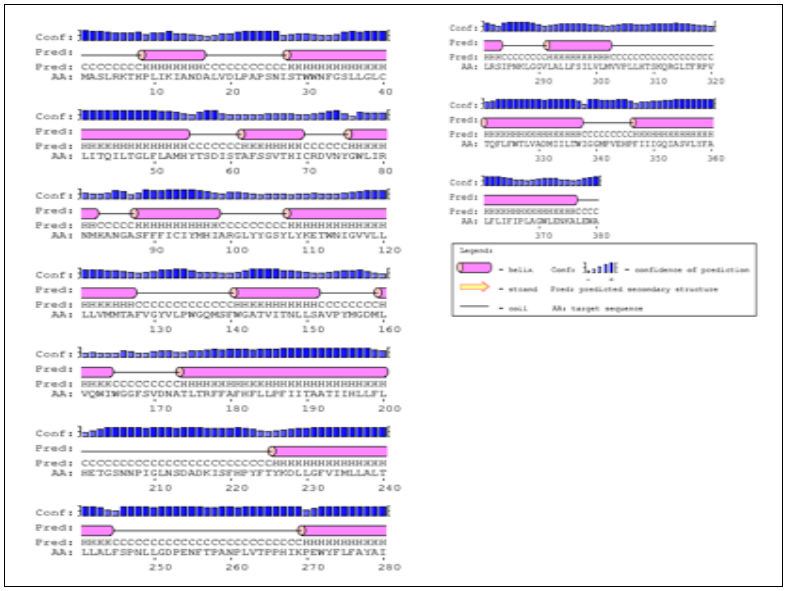


Fig 2: Secondary structure cytochrome b of *Labeo rohita* through PSIPRED

Template identification

☐ ✓	1be3.1.C	CYTOCHROME BC1 COMPLEX		76.78	X-ray, 3.0Å	hetero-oligomer	2 x HEM, 1 x FES, 1 x HEC
☐	1l0n.1.C	Cytochrome B		76.78	X-ray, 2.6Å	hetero-oligomer	3 x HEM, 1 x FES
☐	2a06.1.M	Cytochrome b, heme protein, mitochondrial		76.78	X-ray, 2.1Å	hetero-oligomer	4 x AZI, 4 x CDL, 2 x UQ, 6 x BHG, 2 x FES, 2 x HEC, 4 x HEM, 5 x PEE, 2 x SMA
☐	4d6t.1.C	CYTOCHROME B		76.78	X-ray, 3.6Å	hetero-oligomer	2 x CDL, 2 x HEM, 1 x HEC, 1 x 4X9, 2 x PEE
☐	1l0l.1.C	Cytochrome B		76.78	X-ray, 2.3Å	hetero-oligomer	3 x HEM, 1 x FES, 1 x FMX
☐	5gup.48.A	Cytochrome b		75.20	EM	monomer	None
☐	5gup.59.A	Cytochrome b		75.20	EM	monomer	None
☐	5gup.48.A	Cytochrome b		75.13	EM	monomer	None
☐	5gup.59.A	Cytochrome b		75.13	EM	monomer	None
☐	5j4z.53.A	Cytochrome b		74.01	EM, 5.8Å	monomer	None
☐	3h1j.1.C	Cytochrome b		74.34	X-ray, 3.0Å	hetero-oligomer	4 x CDL, 2 x UQ, 2 x PLC, 2 x FES, 2 x HEC, 4 x HEM, 6 x PEE, 2 x SMA
☐	3bcc.1.C	UBIQUINOL CYTOCHROME C OXIDOREDUCTASE		74.34	X-ray, 3.7Å	hetero-oligomer	2 x AMY, 6 x HEM, 2 x SIG, 2 x FES
☐	3h1j.1.C	Cytochrome b		74.54	X-ray, 3.0Å	hetero-oligomer	4 x CDL, 2 x UQ, 2 x PLC, 2 x FES, 2 x HEC, 4 x HEM, 6 x PEE, 2 x SMA

Fig 3: Template identification from the Swiss model template library.

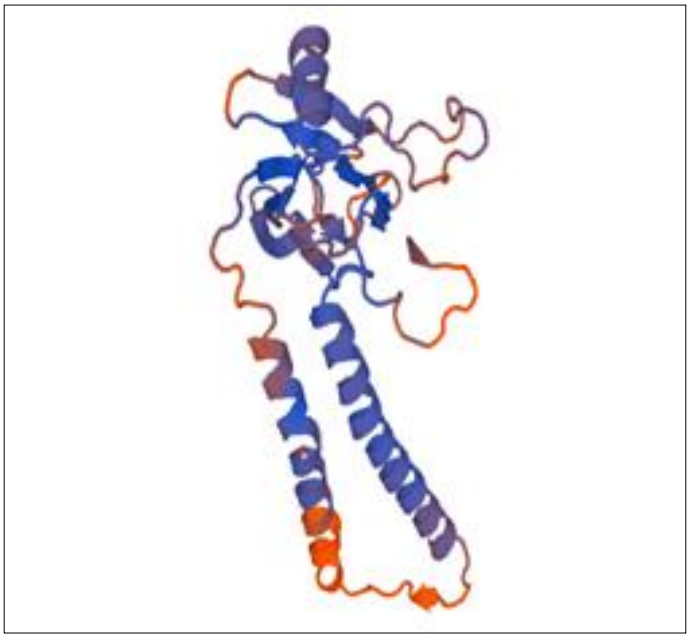


Fig 4: The Modelled tertiary structure of cytochrome b oxidase subunit 2 using Swiss model server.

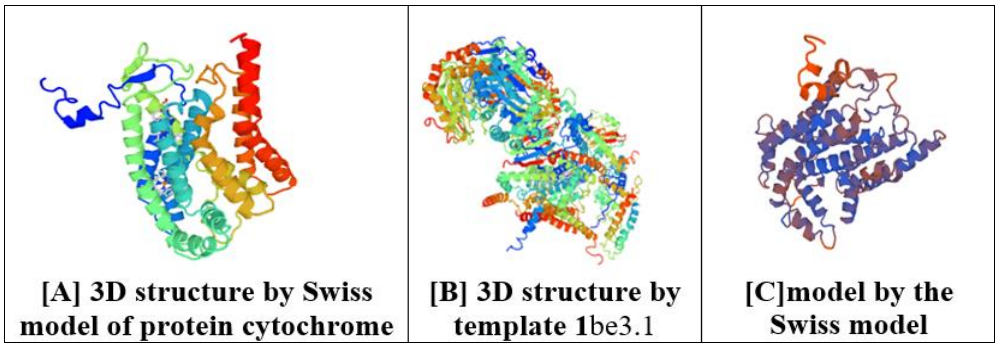


Fig 5: Predicted 3D Structure of Cytochrome b protein are given by A, B and C according to alphabetical order.

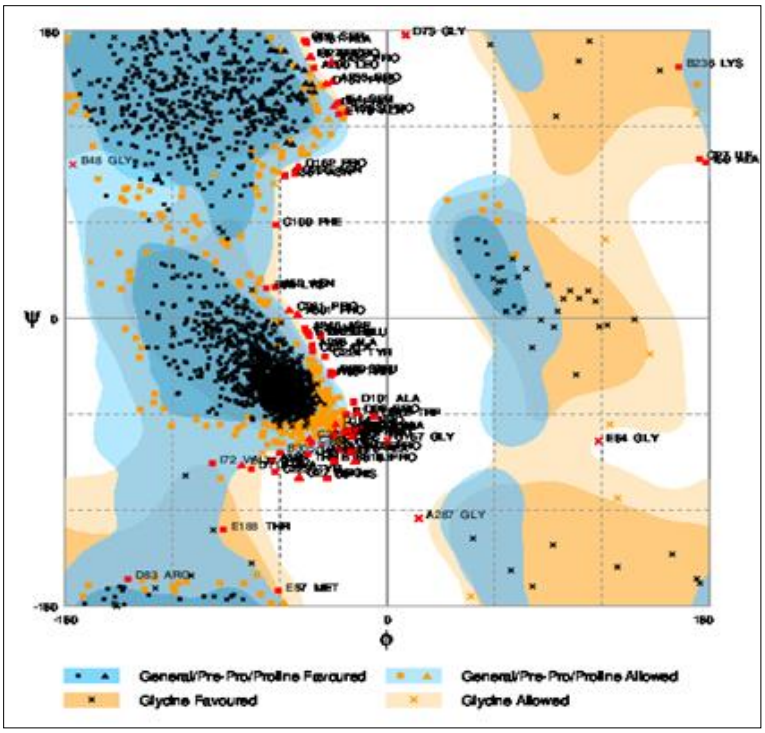


Fig 6: Ramachandran plot for cytochrome b in *Labeo rohita*

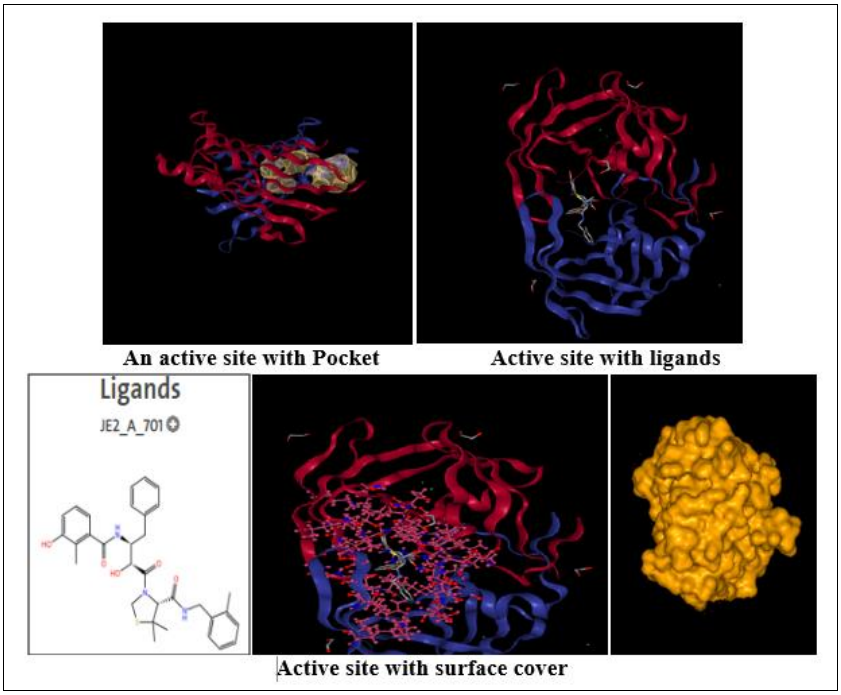


Fig 7: The active site with a surface by dog scorer

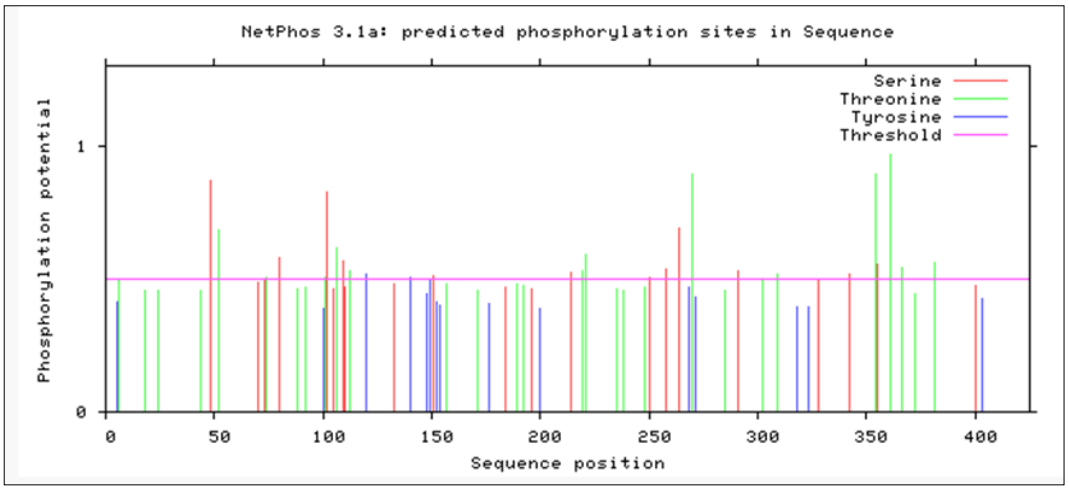


Fig 8: Phosphorylation site prediction

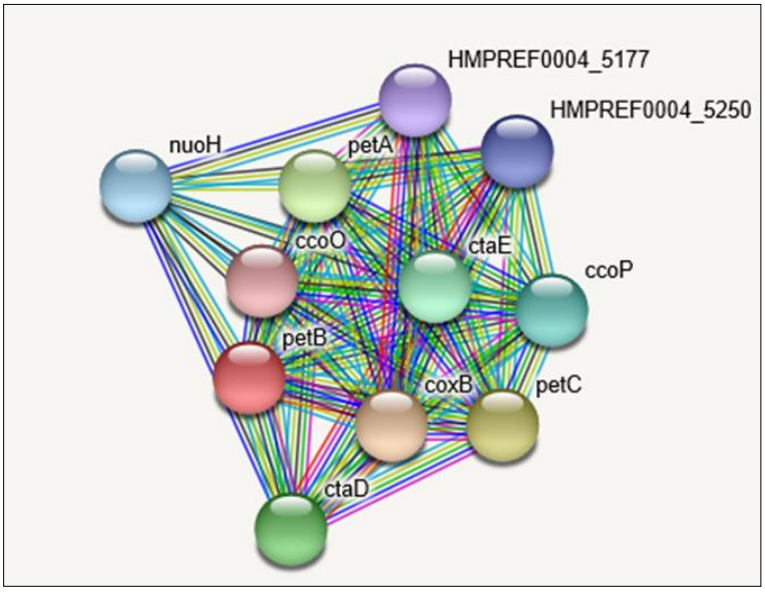
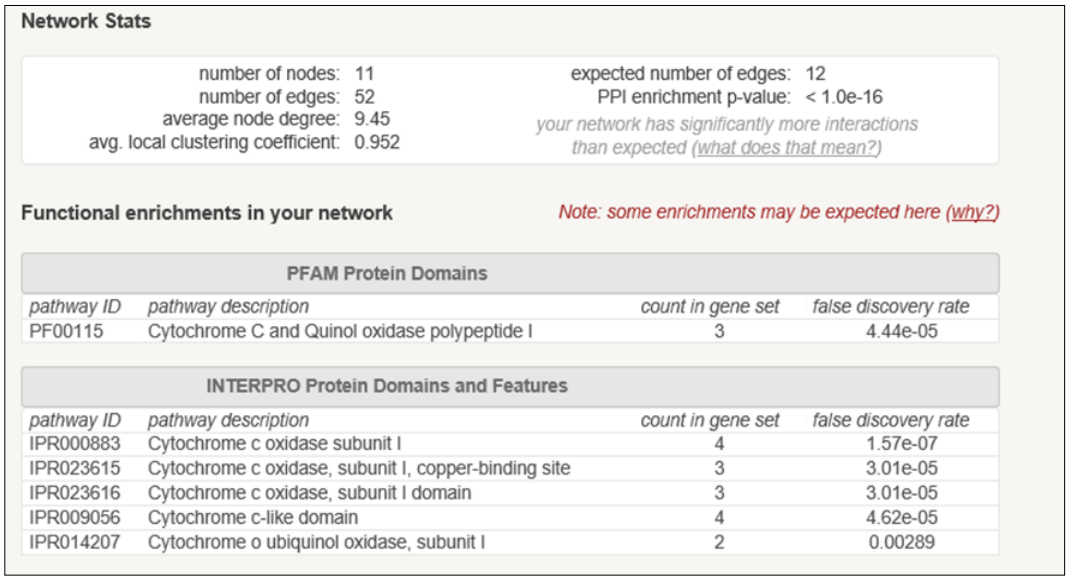


Fig 9: Protein-protein interaction network visualized by STRING.



Network Stats

number of nodes: 11
number of edges: 52
average node degree: 9.45
avg. local clustering coefficient: 0.952

expected number of edges: 12
PPI enrichment p-value: < 1.0e-16
your network has significantly more interactions than expected (what does that mean?)

Functional enrichments in your network

Note: some enrichments may be expected here (why?)

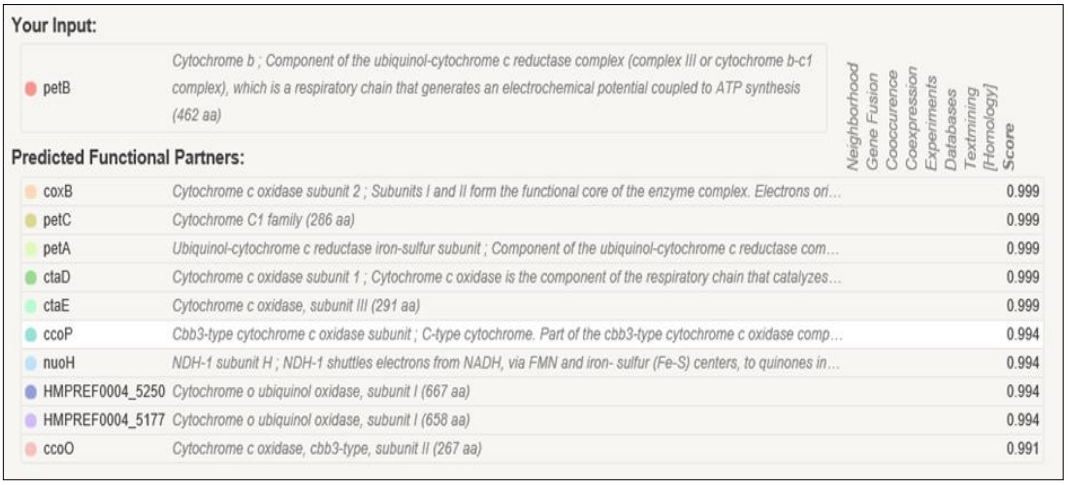
PFAM Protein Domains

pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
PF00115	Cytochrome C and Quinol oxidase polypeptide I	3	4.44e-05

INTERPRO Protein Domains and Features

pathway ID	*pathway description*	*count in gene set*	*false discovery rate*
IPR000883	Cytochrome c oxidase subunit I	4	1.57e-07
IPR023615	Cytochrome c oxidase, subunit I, copper-binding site	3	3.01e-05
IPR023616	Cytochrome c oxidase, subunit I domain	3	3.01e-05
IPR009056	Cytochrome c-like domain	4	4.62e-05
IPR014207	Cytochrome o ubiquinol oxidase, subunit I	2	0.00289

Fig 10: Functional network stats of protein through the string. Functional enrichments in your network



Your Input:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
petB	*Cytochrome b ; Component of the ubiquinol-cytochrome c reductase complex (complex III or cytochrome b-c1 complex), which is a respiratory chain that generates an electrochemical potential coupled to ATP synthesis (462 aa)*									

Predicted Functional Partners:

coxB	*Cytochrome c oxidase subunit 2 ; Subunits I and II form the functional core of the enzyme complex. Electrons ori…*									0.999
petC	*Cytochrome C1 family (286 aa)*									0.999
petA	*Ubiquinol-cytochrome c reductase iron-sulfur subunit ; Component of the ubiquinol-cytochrome c reductase com…*									0.999
ctaD	*Cytochrome c oxidase subunit 1 ; Cytochrome c oxidase is the component of the respiratory chain that catalyzes…*									0.999
ctaE	*Cytochrome c oxidase, subunit III (291 aa)*									0.999
ccoP	*Cbb3-type cytochrome c oxidase subunit ; C-type cytochrome. Part of the cbb3-type cytochrome c oxidase comp…*									0.994
nuoH	*NDH-1 subunit H ; NDH-1 shuttles electrons from NADH, via FMN and iron- sulfur (Fe-S) centers, to quinones in…*									0.994
HMPREF0004_5250	*Cytochrome o ubiquinol oxidase, subunit I (667 aa)*									0.994
HMPREF0004_5177	*Cytochrome o ubiquinol oxidase, subunit I (658 aa)*									0.994
ccoO	*Cytochrome c oxidase, cbb3-type, subunit II (267 aa)*									0.991

Fig 11: String server of predicted interacting protein with the query protein.

Fig 12: Protein model database

Conclusions

After Primary structure analysis and physicochemical characterization we found that the protein is negatively charged and precipitated in acidic medium, Instability index 34.89, this classifies that the protein is stable. Aliphatic index 120.89 and (GRAVY) 0.739, that also indicate this protein has a wide range of temperature stability and hydrophobic nature. Structural analysis was performed after building the model using the Swiss model server. 3D structure information of cytochrome protein will help us to know the role of protein and interaction of their domains with their ligands. 3D structure predictions of protein generate a large amount of

data creating a gap between available sequences and solved structure. Based on the finding it could be concluded that further characterization of cytochrome oxidase protein is novel and will be important for evaluating how the regulation of this protein. Ramachandran plot is important to determine protein structure and the position of amino acids (aa). If the % of aa is more than 88 % in the allowed region the protein model is good for insilico studies. It tells about allowed and disallowed regions of the aa that reflects the stability of protein structure. This work might be valuable contribution in the field of Bioinformatics research and may help other researchers to get an idea about the protein-protein interaction essential for agricultural industries, including dairy industry and bioavailability of phosphorus in soil environment. Hence, this work will also help for detection and identification of such type of proteins in vivo or in silico. However other modelling techniques and Insilico studies would be needed to affirm the claims obtained in the study.

Acknowledgement
Authors are thankful to the Dean, College of Fisheries, Central Agriculture University, Lembucherra, and Agartala for encouragement and support. This work has been carried out under DBT funded BIF centre to the College, hence the financial assistance provided by the DBT, GOI, New Delhi for the centre as well as for this study is duly acknowledged.

References

1. Borek A, Ekiert R, Osyczka A. Molecular effects of mitochondrial mutations in cytochrome b of complex III and their impact on the levels of free radical production.
2. Howell N. Evolutionary conservation of protein regions in the proton motive cytochrome b and their possible roles in redox catalysis. J Mol. Evol. 1989; 29(2):157-69. Doi: 10.1007/BF02100114. PMID 2509716.
3. Esposti MD, De Vries S, Crimi M, Ghelli A, Patarnello T, Meyer A. Mitochondrial cytochrome b: evolution and structure of the protein. Biochim. Biophys. Acta. 1993; 1143(3):243-71. Doi:10.1016/0005-2728(93)90197 N. PMID 8329437.
4. Blankenship, Robert. Molecular Mechanisms of Photosynthesis. Blackwell Publishing, 2009, pp. 124-132.
5. Howell N. Evolutionary conservation of protein regions in the proton motive cytochrome b and their possible roles in redox catalysis. J Mol. Evol. 1989; 29(2):157-169. Doi:10.1007/BF02100114. PMID 2509716.
6. Kallberg Y *et al*. Evolutionary conservation of the ribosomal biogenesis factor Rbm19/Mrd1: implications for function. PLoS One. 2012; 7(9):e43786.
7. Ghosh R, Upadhayay AD, Roy AK, Singh M. *In silico* analysis and 3D structure prediction of mitochondrial RHO GTPase 2 protein of *Danio rerio* (Zebrafish) by homology modelling JJ Bioinformatics and proteomics. 2016; (1):006.
8. Bhandari M, RSSK, Kodavaty J, KRCV. *In silico* analysis and 3D structure prediction of human gastrin protein, IJARIE, ISSN: 2278-6244.
9. Aman Singh Bhalla, Roma Lal. *In silico* sequence analysis, Structure prediction and evaluation of fatty acid Synthase protein (FAS) of Gallus gallus (Chicken) by homology modelling. IJSR. ISSN: 2319-7064.
10. Amisha Singh, Radha Chaube. Bioinformatic analysis, structure modelling and active site prediction from catfish Heteropneustes fossilis IJRITCC: ISSN: 2321-8169.
11. Müller A, MacCallum RM, Sternberg MJE. Benchmarking PSI-BLAST in genome annotation. J Mol. Biol. 1999; 293:1257-1271.
12. Prakruthi Appaiah, Prasanna Basu. *In silico* designing of protein Rich in large Neutral amino acids using Bovine alpha-S1-casein for treatment of phenylketonuria, Journal of proteomics and bioinformatics. 2016; 9:11.
13. Achintya Mohan Goswami. Structural modelling and in silico analysis of non-synonymous single nucleotide polymorphisms of human 3β-hydroxysteroid dehydrogenase type 2Meta Gene. 2015; 5:162-172. Published online 2015 Aug 8. PMCID:PMC4539073
14. Jasnnaik Danish I, Hariharan L, Gupta Pramodkumar P. *In silico* 3D structure modelling and analysis of Galactoside 2- alpha -L- fucosyltransferase 1, Research and reviews: A Journal of bioinformatics, ISSN: 2393-8722, 1(3).
15. Larsen RD, Ernst LK, Nair RP. Molecular cloning, sequence, and expression of human GDP -L-fucose: beta -D- galactoside 2 alpha-L-fucosyltransferase cDNA that can form the H blood group antigen. proc. Natl. Acad. Sci. U.S.A. 1990; 87:6674-8.
16. Krane CM, Kishore BK. Aquaporins: the membrane water channels of the biological world, Biologist. 2003; 50(2):81-86.
17. Ishibashi K, Kuwahara M, Sasaki S. Molecular Biology of Aquaporins, Rev Physiol Biochem Pharmaco. 2003; 141:81-86.
18. Morten Källberg, Haipeng Wang, Sheng Wang, Jian Peng, Zhiyong Wang, Hui Lu *et al*. Template-based protein structure modelling using the Raptor X web server. Nature Protocols. 2012; 7:1511-1522.
19. Jianzhu MA, Sheng Wang, Feng Zhao, Jinbo Xu. Protein threading using context-specific alignment potential. Bioinformatics (Proceedings of ISMB). 2013; 29(13):i257-i265.
20. Jian Peng, Jinbo Xu. A multiple-template approach to protein threading. Proteins. 2011; 79(6):1930-9. Doi: 10.1002/prot.23016. Epub 2011 Apr 4.
21. Jian Peng, Jinbo Xu. Raptor X exploiting structure information for protein alignment by statistical inference. Proteins. 2011; 79(S10):161-171.
22. Please Cite. Lovell SC, Davis IW, Arendall WB, de Bakker PIW, Word JM *et al*. Structure validation by Calpha geometry: phi, psi and Cbeta deviation. Proteins: Structure, Function & Genetics. 2002; 50:437-450.
23. IUPAC-IUB Joint Commission on Biochemical Nomenclature (1986), Nomenclature of Tetrapyrroles (Recommendations 1986), Eur. J Biochem. 178, 277-328 (1988); Pure Appl. Chem. 1987; 59:779-832.
24. Keilin D. Proc. R. Soc. (Lond.) B. Biol. Sci. 1925; 98:312-339.
25. Volkamer A, Kuhn D, Grombacher T, Rippmann F, Rarey M. Combining global and local measures for structure-based druggability predictions. J Chem. Inf. Model. 2012; 52:360-372.
26. Burgoyne NJ, Jackson RM. Predicting protein interaction sites: binding hot-spots in protein-Protein and Protein-ligands interfaces, bioinfo. 2006; 22:1335-1342.
27. Volkamer A, Kuhn D. Rippmann M. Rarey DoG Site Scorer: A wed-server for automatic binding site prediction, analysis, and drugg ability assessment Bio info, 2012.
28. Macarthur MW, Laskowski RA, Thornton JM, Curr.

Opin. Struct. Biol. 1994; 4:737-737. http://dx.doi.org/10.1016/S0959-440X(94)90172-4.
29. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E *et al*. Nucleic Acids Res. 2003; 31(1):365-70.
30. Pramanik K, Ghosh PK, Soumyajit Ray, Sarkar A, Mitra S, Maiti TK. An in silico structural, functional and Phylogenetic analysis with three-dimensional protein modelling of alkaline phosphatase enzyme of Pseudomonas Aeruginosa, Journal of genetic engineering and biotechnology. 2017; 15:527-537.
31. Damian Szklarczyk, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller. Jaime Huerta-Cepas Milan Simonovic Alexander Roth Alberto SantosKalliopi P. Tsafou Michael Kuhn Peer Bork Lars J. Jensen Christian von Mering STRING v10: protein-protein interaction networks, integrated over the tree of life, Nucleic Acids Research, Volume 43, Issue D1, 28 January, 2015, D447-D452.
32. Volkamer A, Kuhn D, Rippmann F, Rarey M. Bioinformatics. 2012; 28(15):2074-5. Doi: 10.1093/bioinformatics/bts310. Epub 2012 May 23.
33. Froese Rainer, Pauly Daniel, eds. Labeo rohita in Fish Base. May version, 2013.
34. Rohu Fish Farming Information Guide - Agri Farming. Agrifarming.in. 26 August 2015. Retrieved 8 September, 2018.