J. Aqua. **22** (2014)

IN-SILICO ANALYSIS AND HOMOLOGY MODELING OF CYTOCHROME C OXIDASE SUB UNIT I FROM INDIAN MAJOR CARP, *LABEO ROHITA*

Sukanta Kumar Pradhan[1]* and Lakshman Sahoo[2]

[1]*Department of Bioinformatics, Orissa University of Agriculture Technology (OUAT), BBSR-751003, Odisha, India*

[2]*Fish Genetics and Biotechnology Division, ICAR-Central Institute of Freshwater Aquaculture (CIFA), Kausalyaganga, BBSR-751002, Odisha, India*

**Corresponding Author: ksukantapradhan@gmail.com*

Labeo rohita is the popular table fish in the whole Indian sub-continent. Cytochrome c oxidase sub unit I (COX/COI) is the largest protein coding gene in the metazoan mitochondrial genome and COI is the most frequently studied region of the teleost mitochondrial genome. It is the terminal member of the respiratory chain catalyzing the reduction of dioxygen to water by ferrocytochrome C. In this study, COI of *L. rohita* (56.86KDa) has been characterized, and the 3-D structure was predicted by using Modeller version 9.12 by homology modeling. The secondary structure has been predicted using PSIPRED and TMHMM server, and the predicted model has been validated in RAMPAGE server. The protein statistics was carried out by using Protopram tool in Expasy server. The predicted 3-D model showed that 99.8% of residues have Φ and Ψ angle in the favored and allowed regions. It was found to be a transmembrane protein and consisted of an N-terminal signal peptide. The isoelectric focusing point, instability index, aliphatic index, grand average hydropathicity, extinction coefficient, absorbance and number of negatively charged and positively charged amino acids were; 6.05, 26.52, 109.81, 0.749, 121180, 2.142, 25 and 16 respectively, revealing the structure and properties of COI protein. The COX/CO-I protein sequences of *Catla catla, Cirrhinus mrigala, Labeo rohita, C. cirrhosus* and *Danio rerio* were also taken for the divergence study and mutation analysis.

INTRODUCTION

The Cytochrome c oxidase (COX) region of mitochondrial DNA is the most studied region of the fish mitochondrial genome. COX is one of the largest protein coding genes of metazoan mitochondrial genome. COX (E.C.1.9.3.1) is the terminal member of the respiratory chain catalyzing the reduction of dioxygen to water by ferrocytochrome C. It is the last enzyme in the respiratory electron transport chain of mitochondria located in the mitochondrial membrane. It receives an electron from each of four cytochrome c molecules, and transfers them to one oxygen molecule, converting molecular oxygen to two molecules of water. In the process, it binds four protons from the inner aqueous phase to make water, and in addition translocates four protons across the membrane, helping to establish a transmembrane difference of proton electrochemical potential that the ATP synthase then uses to synthesize ATP.

Functional characterization of a protein sequence is one of the most frequent problems in biology. The accurate three-dimensional (3-D) structure of the protein is necessary in order to understand its function. In the absence of an experimentally determined protein structure, predicted 3D models give the insights of its functions. Comparative modeling predicts the 3-D structure of a given protein sequence (target) based primarily on its alignment to one or more proteins of known structure (templates).

Indian aquaculture has demonstrated a six and half fold growth over the last two decades, with freshwater aquaculture contributing over 95 percent of India's total aquaculture production. Currently, India is placed second in the world with 4.65 million tonnes over the world aquaculture production of 59.87 million tonnes (FAO, 2012). Production from freshwater fishes in India has constantly been dominated by Indian major carps such as rohu *(Labeo rohita*), catla (*Catla catla*) and mrigal (*Cirrhinus mrigala*). *L. rohita,* commonly known as rohu, is a popular table fish in India, Bangladesh, Pakistan and Myanmar (Talwar and Jhingran, 1991). Being an Indo Gangetic riverine species, rohu is the natural inhabitant of northern and central India. Subsequently, it has been transplanted into almost all river systems including the freshwaters of Andaman, where its population has successfully established (FAO, 2006). This fish is well known for its good taste, high protein content and high nutritive value and contributes more than 80% of the total aquaculture production of the Indian major carps (Talwar and Jhingran, 1991). The breeding and aquaculture practices of this particular species have been well established. However over the years the wild population of this important aquaculture species has declined due to overfishing, destruction of natural habitat, climate change, disease, water pollution, introduction of exotic species, poisoning, dynamite and destructive fishing, these factors not only destroyed the breeding and feeding grounds but also had enormous effect on reducing the biodiversity.

Fish mitochondrial DNA (mt DNA) all have a similar genomic organization (Lee *et al.,* 2001; Kim *et al.,* 2005; Nohara *et al.,* 2005) and are similar to other vertebrates including humans. Many parts of mt DNA such as those coding for protein genes or regulatory part as the control region are used as genetic markers for measurement of intra species and inter species diversity (Hebert *et al.,* 2003). This quality is because of an increased mutation rate for mtDNA, relative to nuclear DNA, which result in an accumulation of many base substitutions over a long period of time, providing tools for taxonomic, evolutionary and phylogentic research (Wallace, 1992; Nei and Kumar, 2000; Avise, 2001; Kartavtsev *et al.,* 2007). Cytochrome C oxidase I region (COX/CO-I) of mitochondrial DNA is the most studied regions in DNA barcoding (Kranthi *et al.,* 2006). COX/CO (E.C. 1.9.3.1.) is the terminal member of the respiratory chain catalyzing the reduction of dioxygen to water ferrocytochrome C. The mt genome of animals represent a better target for analysis of the nuclear genome because of its lack of introns and its limited exposure to recombination and its haploid mode of inheritance, COX likely possess a greater range in phylogenetic signal than any other mitochondrial gene (Hebert *et al.,* 2003). The present study encompasses *in silico* prediction of 3D model of COX/COI of *L. rohita* in order to understand its function and its role in species identification.

MATERIALS AND METHODS

The *Labeo rohita* COX/CO-I protein sequence was obtained from the National Centre for Biotechnology Information (NCBI) protein database (http://www.ncbi.nih.gov) (Accession number: JN412817 protein ID: AEO21147.1). Bovine heart cytochrome c oxidase (PDB 1V54A) was taken as the template structure for the comparative modeling which showed 88 % identity with COX/CO-I protein. The 3D structures of the COX/CO-I protein (516 amino acids) *Labeo rohita* were constructed by generating profiles using MODELLER The Model COX.B99990001.pdb was selected by taking the low DOPE score value (Sali and Blundell, 1993) (http://salilab.org/modeller/). The model was validated with RAMPAGE server (http://mordred.bioc.cam.ac.uk/-rapper/rampage.php). The secondary structure COX was predicted by using PSIPRED server (http://www.bioinf.cs.ucl.ac.uk/psipred/), transmemberane protein helix probability curve was analyzed by using TMHMM server V.2.0 (http://www/cbs.dtu.dk/services/TMHMM.2.0). The energy minimization was carried out with QMEAN server (https://swissmodel.expasy.org/qmean/).

The COX/CO-I protein sequences of *C. catla, C. mrigala, L. rohita, C. cirrhosus* and *D. rerio* were taken for the divergence study and mutation analysis. The divergence study was carried out in MEGA 5.05 (http://www.megasoftware.net/) using Neighbour Joining method and validated using Bootstrapped method using 1000 replications. The mutational analysis was carried out using MUSCLE (http://www.ebi.ac.uk/Tools/msa/muscle/).

RESULTS

In this study, the three dimensional model of *Labeo rohita* Cytochrome c oxidase was constructed by comparing the structure of the bovine heart Cytochrome c oxidase using Modeller version 9.12 (Fig. 1a, b). Results of the Ramachandran plot indicated that most of residues had Φ and Ψ angle in the favored (98.2%) and allowed regions (1.6%) (Fig. 2a, b) the bond angle, bond length and torsion angles were in the range of value expected for a naturally folded protein (Fig. 4-6). The secondary structure of COX results revealed that the helix regions were H1 (Gly16-Leu41), H2 (Ile57-Phe68), H3 (Leu105-Gly117), H4 (Asp144-Leu150), H5 (His151-Ile169), H6 (Trp186-Leu195), H7 (Ile254-Val258), H8 (Gly272-Ile280), H9 (Lys319-Leu327), H10 (Leu339-Ile345), H11 (Met383-Pro398), H12 (Ser407-Phe418), H13 (Asn451-Ser454) and H14 (Gly457-Ala478), and strand regions were S1 (Thr127-Val128), S2 (Val243-Ile247), S3 (Met310-Ala313), S4 (Ile323-Lys333), S5 (Ile365-Leu367), S6 (Thr370-His376) and S7 (Phe510-Gln512). The COX domain is divided into helix and strand regions. The trans-membrane helix probability curve was presented in the Figure 3 and found that it is a trans-membrane protein containing N-terminal signal peptide and 12 transmembrane helices. The COX protein of *L. rohita* consisted of 516 amino acids, total number of negatively charged amino acid (Asp + Glu) and positively charged amino acids (Arg + Lys) were found to be 25 and 16 respectively. Instability index (II), aliphatic index, grand average hydropathicity (GRAVY), extinction coefficient, absorbance (assuming cysteine both in reduced form and in cysteine form) and theoretical PI were 26.52, 109.81, 0.749, 121180, 2.142 and 6.05 respectively. Molecular weight of COX was 56.86 Kda and consisted of 2696 carbon, 4030 hydrogen, 626 nitrogen, 676

oxygen and 26 sulphur atoms. The amino acid composition (Table 1) revealed that the frequency of leucine was highest at the same time cysteine occurred least frequently.

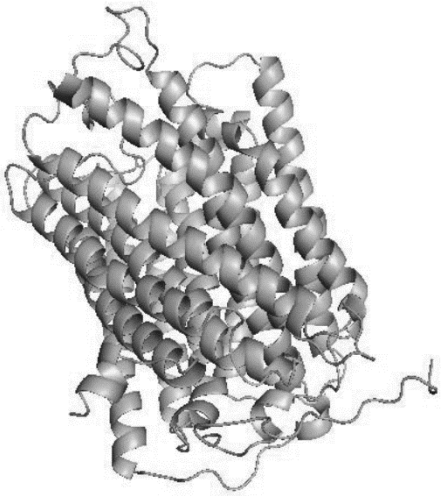

Fig.1a. Three dimensional structure of COX JN412817 before energy minimization

Fig.1b. Three dimensional structure of COX JN412817 after energy minimization.

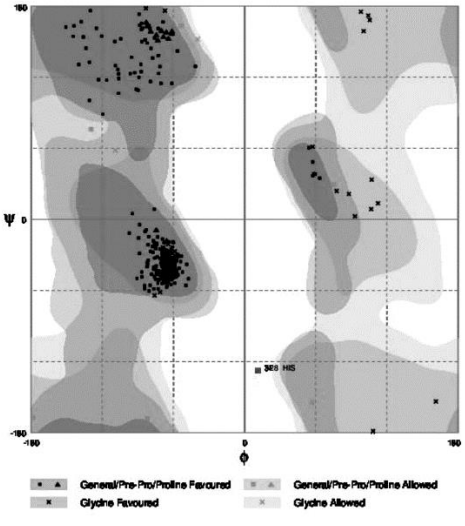

Fig. 2a. Ramachandran Plot of COX JN412817 pdb model

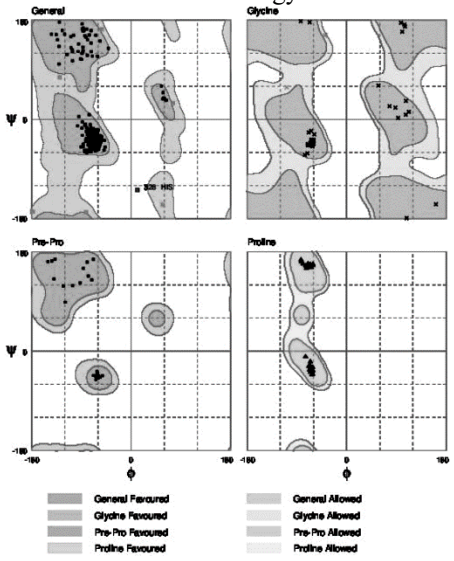

Fig. 2b. Ramachandran Plot of COX JN412817 pdb model Glycine and Proline allowed regions

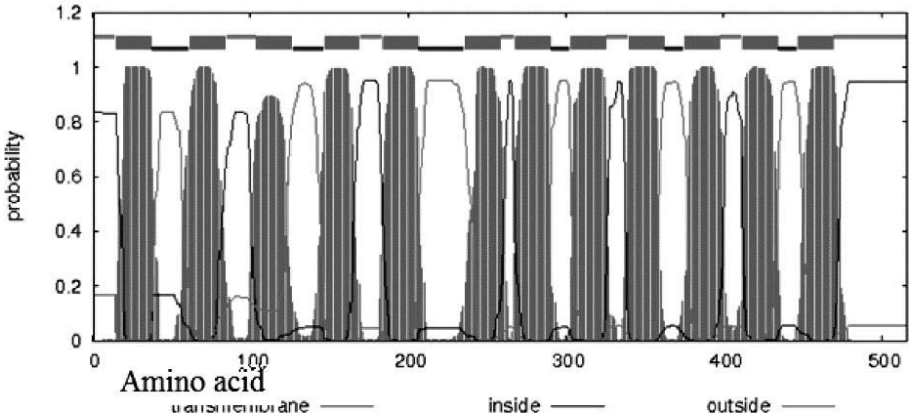

Fig. 3. Transmembrane protein helix probability curve of COX JN412817.

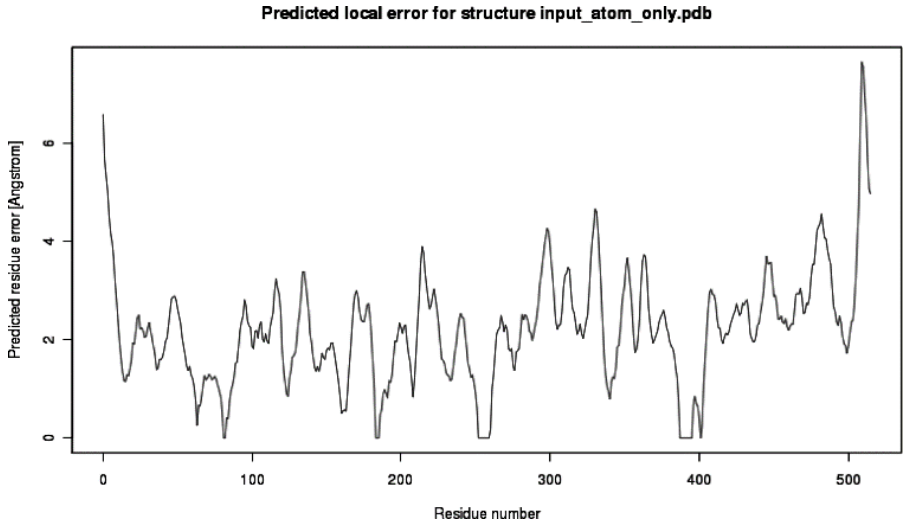

Fig. 4. Energy Plot of COX JN412817.

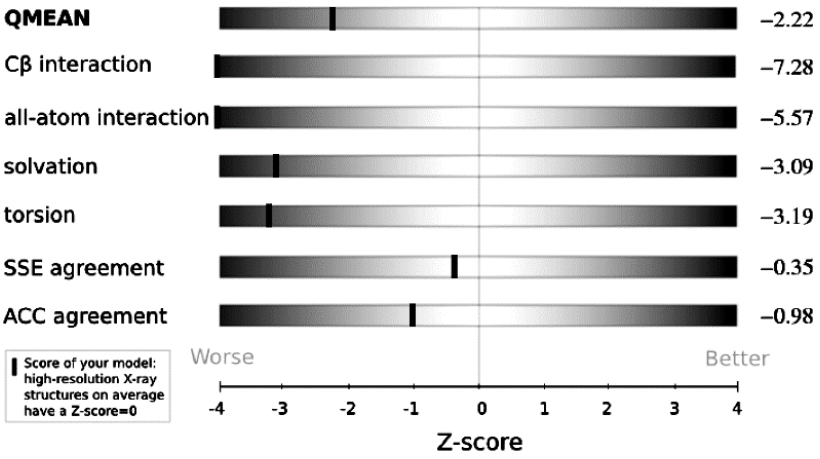

Fig. 5. Z-score of COX JN412817

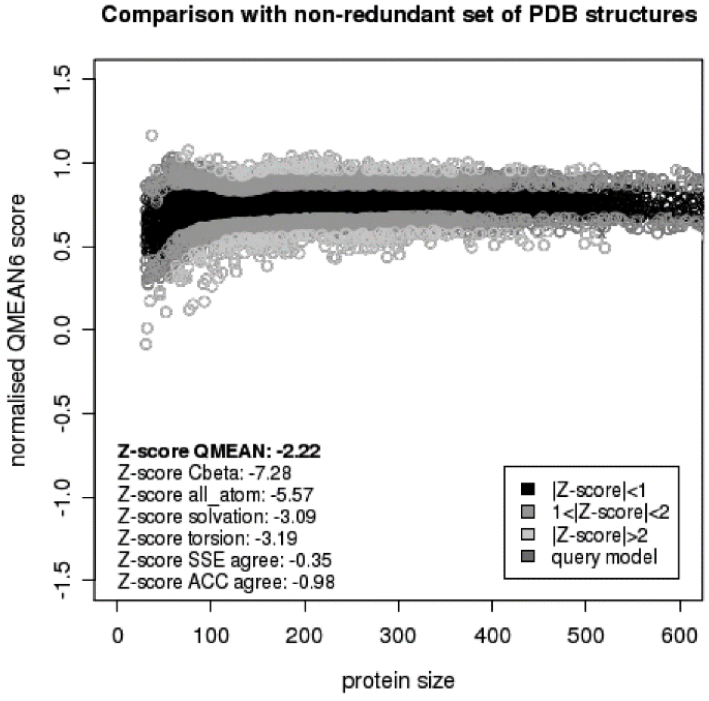

Fig. 6. Comparison of Z-score COX JN412817 with other structures of PDB

Table 1. Amino acid composition of COX JN412817.

Name of the Amino Acid	No of Amino acid present	Frequency
Ala	46	8.90%
Arg	8	1.60%
Asn	15	2.90%
Asp	14	2.70%
Cys	1	0.20%
Gln	8	1.60%
Glu	11	2.10%
Gly	46	8.90%
His	19	3.70%
Ile	39	7.60%
Leu	64	12.40%
Lys	8	1.60%
Met	25	4.80%
Phe	40	7.80%
Pro	28	5.40%
Ser	32	6.20%
Thr	35	6.80%
Trp	17	3.30%
Tyr	19	3.70%
Val	41	7.90%
Pyl	0	0.00%
Sec	0	0.00%

The mutational analysis of COX/CO-I proteins of *C. catla, C. mrigala, L. rohita, C. cirrhosus* and *D. rerio* showed that at position second A - T in *D. rerio,* at position 29 V – A in *L. rohita* and *C. catla.* At position 46 S – A in *D. rerio,* position 135 N –K in *C. mrigala,* position 166 T-A in *L. rohita* and *C. catla,* position 175 A - T in *D. rerio,* position 187 S - A in *D. rerio*, position 262 S - A in *D. rerio,* position 331 S - A in *D. rerio*, position 347 L - Q in *C. mrigala,* position 359 A - S in *C. mrigala,* position 463 I- V in *D. rerio* and *C. cirrhosus,* position 477 A - T in *D. rerio,* position 489 M – A in *D. rerio,* position 505 Y – F in *D. rerio.*

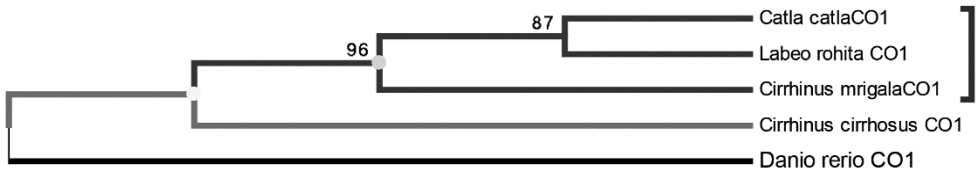

Fig. 7. NJ Phylogenetic Trees showing divergence

The computational phylogenetic analysis showed that *C. catla, C. mrigala, L. rohita* were grouped into one group but *C. mrigala* was more diverged from other two species (Fig. 7). It was also observed that the species *D. rerio* was more diverged from these three species than *C. cirrhosus.*

DISCUSSION

The complete mtDNA sequences of many aquatic species have been sequenced and several complete genome sequencing project were in the pipeline. The complete mitochondrial genomes of three Indian major carps have been sequenced recently (Bej *et al.,* 2012a, b; 2013). With the ever increasing genomic scale data, structural biology now faces demanding chore of characterizing the shapes and dynamics of the encoded proteins in order to facilitate the understanding their functions and mechanism of actions. Advances in the techniques of structure determination at high resolution level such as X-ray crystallography, nuclear magnetic resonance spectroscopy, have expedited the quality and the speed of structural studies (Zhang and Kim, 2003). Contemporary information clearly depicts that the known protein sequences ~1,000,00 (Boeckmann *et al.,* 2003) vastly outnumber the available protein structures ~20,000 (Berman *et al.,* 2002). At the same time domains in protein sequences are progressively developing units that can be grouped into small number of families of domains with similar sequences and structures evolving (Vitkup *et al.,* 2001). Therefore, the evolutionary relationships between protein sequences make it possible to use computational methods, such as threading (Dominqus *et al.,* 2000) and comparative protein structure modeling (Abhilash and Nandhini, 2003) to predict structures of protein sequences based on their similarity to available protein structures.

A combination of experimental structure determination procedures and computational modeling techniques were used in several structural genomics studies to deduce a sufficient number of suitably selected structures, with the intention that best part of other sequences can be placed within modeling distance of at least one known structure. To get the most out of proteins that could be modeled more accurately, a combined exertion toward structure deduction of new folds by X-ray crystallography and nuclear magnetic resonance spectroscopy is in progress, as visualized by structural genomics (Sali, 1998; Montelione and Anderson, 1999; Sanchez *et al.,* 2000; Vitkup *et al.,* 2001). Projection has been made that 90% of all globular and transmembrane proteins could be organized into roughly 16000 families containing protein domains having more

than 30% sequence identity to each other (Vitkup *et al.,* 2001). Structures of 4000 protein families out of 16000 families are already determined experimentally; the others present suitable targets for structural genomics. Comprehensive prospective of whole genome sequencing projects will only be recognized once all protein functions are assigned and understood. At this juncture comparative modeling will play an important bridging role in these efforts.

In the era of genomics it has been proposed that a single gene sequence would be sufficient to identify all / at least the vast majority of existing animal species, and proposed the use of mitochondrial DNA gene cytochrome c oxidase sub unit I (COI) as an universal molecular taxonomy systems for animals commonly acknowledged as DNA barcoding, protein coding COI gene are responsible well conserved proven to be robust molecular marker for elucidating inter species relationship. The structure of plant cytochrome P_{450s} was predicted by homology modeling (Rupsinghe and Schuler, 2006). They concluded that homology modeling signifies a dependable and relatively rapid alternate method for analyzing structure-function relationships and predicting substrates for many such $P_{450s.}$ Similarly, Azhaguraj and Selvanayagam, (2010) determined the structure of COI gene of *Channa punctata*. Though the COI gene of Indian major carps have been used for DNA bacoding (Mohanty *et al.,* 2013), the secondary and 3-D structure was totally ignored. This motivated us to investigate the mitochondrial COI gene of rohu, one of the important Indian major carp. The homology modeling carried out in the present investigation will give more insight on DNA barcoding. Homology modeling of rohu COI was performed with using tools like BLASTP, Swiss pdb Viewer, Modeler version 9.12, Rampage server (Ramchandran plot), PSIRED, QMEAN server, TMHMM and Expasy server. The present comparative computational approach would help the researchers by giving them a hand-in idea, so that they can predict the target proteins COX / COI of other freshwater carp species and at the same time it would be helpful to conclude the 2-D and 3-D structure and functions of COI and also be useful for DNA barcoding studies.

REFERENCES

Abhilash, M. and M. Nandhini, 2003. Homology modeling and In-silco analysis of hamaggutinin protein from HINI influ-enza A virus. *Int. J. Phar. Sci. Res.,* **1 (1):** 40-50.

Avise, C, 2001. Phylogeography. The History and Formation of Species. Harward University Press, Cambridge.

Azhaguraj, R. and M. Selvanayagam, 2010. Homology modeling and *In silco* analysis of COX from *Channa punctata* (Bloch). *Biores. Bull.,* **2:** 044-050.

Berman, H. M., T. Battistuz, T. N. Bhat, W. F. Bluhm, P. E. Bourne, K. Burkhardt, Z. Feng, G. L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padillia, V. Ravichabndran, B. Scneider, N. Thanki, H. Weissig, J. D. Westbrook, C. Zardecki, 2002. The Protein Data Bank. Acta. Crystallogr. *D. Biol. Crystallogr*., **58:** 899-907.

Bej D, L. Sahoo, S. P. Das, S. Swain, P. Jayasankar, P. C. Das and P. Das, 2012a. Complete mitochondrial genome sequence of *Catla catla* and its phylogenetic consideration. *Mol. Biol. Rep.,* **39(12):** 10347-10354.

Bej D, L. Sahoo, S. P. Das, S. Swain, P. Jayasankar, P. C. Das and P. Das, 2012b. Complete mitochondrial genome of *Labeo rohita*. *mtDNA* **23(6):** 441-443.

Bej D, L. Sahoo, S. P. Das, S. Swain, P. Jayasankar and P. Das, 2013. Complete mitochondrial genome sequence of *Cirrhinus mrigala* (Hamilton, 1822). *mtDNA*, **24(2):** 91-93.

Boeckmann, B, A. Bairoch, R. Apweiler, M. C. Blatter, A. Estericher, E. Gasteiger, M. J. Martin, K. O. Michoud, C. Donovan, I. Phan, S. Pilout and M. Schnedier, 2003. The SWISS PROT Protein knowledgebase and its supple-ment TrEMBL. *Nucleic Acids Res.*, **31:** 365-370.

Domingues, F. S, Koppensteiener, W. A. and M. J. Sippal, 2000. The role of protein structure in genomics. *Febs. Lett.,* **476:** 98-102.

FAO 2006. The State of World Fisheries and Aquaculture, Rome: Food and Agriculture Organization of the United Nations.

FAO, 2012. The State of World Fisheries and Aquaculture. Rome: Food and Agriculture Organization of the United Nations.

Hebert, P. D. N., A. Cywinska, S. L. Ball and J. R. Dewarrd, 2003. Biological identifications through DNA Barcodes. *Proc. R. Soc. Lond. Biol. Sci.,* **270** : 313-321.

Kartavtsev, K. P., T. J. Park, K. A Vinnikov., S. N. I. Sharina and J. S. Lee, 2007. Cytochrome b cyt-b gene sequence analysis in six flatfish species (Teleostei; Pleuronectide) with Phy-logenetic and Taxonomic insights. *Mar. Sci.*, **152:** 757-773.

Kim, I. C., S. O. Jung, Y. M. Lee, C. J. Lee, J. K. Park and J. S. Lee , 2005. The Complete mitochondrial genome of the ray fish *Raja porosa* (Chondrichtyes; Rajiidae). *DNA Seq.,* **16:**187-194.

Kranthi, S, K. R. Kranthi, A. A. Bharose, S. N. Syed, C. S. Dhawad and E. K. Patil, 2006. Cytochrome oxidase I se-quence of *Helicoverpa* (Noctuidae: Lepidoptera) species in India- Its utility as a molecular tool. *J. Bio-tech.,* **5:** 195-199.

Lee, J. S., M. Miya, Y. S. Lee, C. G. Kim, E. H. Park and N. M. Aokiy, 2001. The complete DNA sequence of the mitochondrial genome of the self fertilizing fish *Rivulus marmoratus* (Cyprinnodontifome Rivuilidae) and the first description of the duplication of control region in fish. *Gene,* **280:**1-7.

Montelione, G. T. and S. Anderson, 1999. Structural genomics keystone for a human proteome project. *Nat.Struct.Biol.*, **6:** 11-12.

Nei, M. and Kumar, S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York, p333.

Nohara, M, M. Nishida, M. Miya and T. Kawa, 2005. Evolution of the mitochondrial genome in cephalochordate as in-ferred from complete nucleotide sequence from two epi-genichty species. *J .Mol. Evol.,* **60:** 526-537.

Rupasinghe, I and M. A. Schuler, 2006. Homology modeling of plant cytochrome P450s. *Phytochem. Rev.,* **5:** 473 – 505.

Sali, A., 1998. 100, 000 protein structures for the biologist. *Nat. Struct. Biol.,* **5:** 1029-1032.

Sali, A. and T. L. Blundell, 1993. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.,* **234:** 779-815,

Sanchez, R, U. Pieper, F. Melo, N. Eswar, M. A. Marti-Renom, M. S. Madhusudhan, N. Mirkovic. and A. Sali, 2000. Protein structure modeling for structural genomics. *Nat. Struct. Biol.,* **7:** 986-990.

Talwar P. K. and A. G. Jhingran, 1991. Inland Fishes of India and Adjacent Countries, Volume 1, Oxford & IBH Publishing Co. Pvt. Ltd., New Delhi, India

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei and S. Kumar 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.,* **28(10):** 2731-2739.

Vikup, D., E. Melamud, J. Moult and C. Sander, 2001. Complete-ness in structural genomics. *Nat. Struct .Biol.,* **8:** 559-566.

Wallace, D. C., 1992. Disease of the mitochondrial DNA. *Annu. Rev. Biochem.,* **61:** 1175-1212.

Zhanng, C. and S. H. Kim, 2003. Overview of structural genomics from structure to function. *Curr. Opin. Chem. Biol.*, **7:** 28-32.