In silico 3-D structure prediction of cytochrome b protein of sisorid catfish *Glyptothorax ngapang*

Mahender Singh[1]*, Ashish Gupta[2] and W S Lakra[1]

[1]National Bureau of Fish Genetic Resources (NBFGR), Lucknow 226 002, India
[2]Amity Institute of Biotechnology, Amity University, Noida Campus, Noida 201 303, India

Samples of cold water catfish *Glyptothorax ngapang* were collected from Iril River at Serou, Manipur. Full length cytochrome b (cyt b) was amplified with primer pair L14724 and H15915 in 14 samples. PCR products were sequenced using automated ABI 3730 sequencer and submitted to NCBI. The nucleotide sequences were converted to amino acid sequences by defining the reading frame. Before prediction, it was confirmed that no 3-D structure was available for this protein in fish. A comparative modeling method was used for the prediction of the structure of cyt b protein. For the modeling, template protein 1BCC chain C, having identity 88%, E value 6e-175 and alignment score 1571, was obtained by Geno3D server. By comparing the template protein, a rough model was constructed for the target protein using MODELLER, a program for comparative modeling. The structure of cyt b protein of *G. ngapang* was found to resemble to a high resolution structure of 1BCC chain C of cytochrome bc1 complex from chicken. From Ramachandran plot analysis, it was observed that the percentage of residues falling into the most favoured regions was 93.1%.

Keywords: Chicken, cytochrome bc1 complex, *Glyptothorax ngapang*, modeling, Ramchandran plot, target

Introduction

Fish is the simplest and most diverse vertebrate. It is also considered as forefather of terrestrial vertebrates. Unlike other vertebrates, it respires through gills all over the life. As water contains only 15 mg/L oxygen, only the $1/30^{th}$ found in the air, there is need for highly efficient system in aquatic organisms for oxygen absorption. Fish gills can extract up to 95% of the oxygen in the water and are the most efficient among water-breathing organisms. Despite this, catfishes have also developed the process of aerial respiration to overcome the dry season, when river channels and lakes dry up and become crowded by trapped fish. The catfish can survive this situation by gasping oxygen from the air. While most catfishes gasp at the surface for air, others have taken the adaptation of aerial respiration to an extreme, e.g., albino walking catfish has pectoral spines and actually walk out of the water. In this way catfishes may be useful model for studying the transition from aquatic to aerial respiration. The order Siluriformes includes 37 recognized families of catfishes, widely distributed and diverse in freshwaters[1]. Family Sisoridae is an exclusively Asian family of bottom dwelling catfishes. *Glyptothorax* Blyth 1860 is the most diverse and the most widely distributed genus of this family with 89 nominal species found in Asia Minor (in the Tigris and Euphrates river drainages), eastwards to East Asia (to the Yangtze River drainage) and southwards to Southeast Asia[2].

The cellular respiration in all eukaryotes involves electron transport chain and proton translocation across the mitochondrial inner membrane, and ATP generation in the presence of oxygen. Eukaryotic cytochrome b (cyt b) is the only cytochrome coded by mitochondrial DNA. It binds non-covalently to two hemes, known as b562, the high potential form (bH), and b566, the low potential form (bL). Four conserved histidine residues are ligands of the iron atoms of these two hemes. It is the main subunit of trans-membrane cytochrome bc1 complex, also known as the respiratory chain complex III (EC 1.10.2.2) or ubiquinol-cytochrome c reductase, and plays key role in cellular respiration. Respiratory chain complex is minimally composed of three subunits: cyt b, cytochrome c_1 (cyt c_1) with one

_____
*Author for correspondence:
Tel: +91-522-2442441; Fax: +91-522-2442441
E-mail: mahender@scientist.com; msverma@nbfgr.res.in

covalently bound heme c and a high potential Rieske iron-sulphur protein (ISP) containing a single $[Fe_2S_2]$ cluster. The general function of the complex is electron transfer between two mobile redox carriers, ubiquinol (QH_2) and cyt c. The electron transfer is coupled with the translocation of protons across the membrane matrix to inter membrane space (in mitochondria); thus, generating proton motive force in the form of an electrochemical proton potential, which can drive ATP synthesis[3]. The path of electron transfer from QH_2 to cyt c through the cytochrome bc1 complex is the protonmotive Q cycle. The Q cycle model postulates two separate ubiquinone binding sites in cyt b, called Q_o (quinol oxidising site) and Q_i (quinone reducing site). Q_o is located near the positive side of the membrane (inter membrane space) and Q_i is located near the negative side of the membrane (matrix). Electron transport starts with ubiquinol binding to cyt b, 2Fe/2S center and b_L heme; each pull an electron off the bound ubiquinol and two hydrogens are released into the intermembrane space. The 2Fe/2S center transfers its electron to membrane-bound cytochrome c1 and the b_L heme transfers its electron to the b_H heme. Cyt c1 transfers its electron to a water-soluble cytochrome c and the b_H heme transfers its electron to a nearby ubiquinone, turning the ubiquinone into a ubisemiquinone. Cyt c diffuses and the fully oxidized ubiquinone is released. Then another ubiquinol binds to cyt b and follow the same steps, except that b_H heme transfers its electron as well as two hydrogens from the matrix to the nearby ubisemiquinone, turning it into a ubiquinol[4]. The fully oxidized ubiquinone and ubiquinol are released. The membrane spanning region of each cyt bc_1 complex monomer consists of 13 transmembrane helices, 8 of which belong to cyt b, an integral membrane protein of approximately 400 amino acid residues[5].

A small fraction of electrons leaves the electron transport chain before reaching complex IV. Premature electron leakage to oxygen results in the formation of superoxide. The relevance of this, otherwise minor side reaction, is that superoxide and other reactive oxygen species are highly toxic. Mutation in cyt b may result in overproduction of superoxide, deficit bioenergetic pathways producing less ATPs or resistance to particular chemical/drugs, etc. Single point mutations in cyt b of *Plasdmodium falciparum* and *P. berghei* are associated with resistance to the anti-malarial drug atovaquone[6].

Mutation in cyt b results in exercise intolerance in human patients, they can't perform hard work. Mutation in fishes may also be having similar effects and reduces its fitness.

Comparative modeling (CM), also known homology modeling, is a class of methods for constructing an atomic-resolution model of a protein from its amino acid sequence, i.e., "query sequence" or "target"[7]. CM can produce high-quality structural models when the target and template are closely related, which has inspired the formation of a structural genomics consortium dedicated to the production of representative experimental structures for all classes of protein folds[8]. CM is becoming a useful technique in the field of bioinformatics because the knowledge of the 3-D structure of a protein would be an invaluable aid to understand the structural and functional detail of a particular protein[9]. Since there was no structure reported/predicted for this protein, this study is to predict 3-D structure of this respiratory chain protein of fish and to establish some important facts for this protein.

Materials and Methods
Sample Collection
In total, 14 specimens of *Glyptothorax ngapang* were collected from Serou (latitude/longitude/altitude: 24.16/093.52/2521) Manipur, India. The species identification was done based on morphometric and meristic characters by Professor W Vishwanath and Dr M Linthoingambi (original authors of the *G. ngapang*)[10]. Muscle and fin tissues were preserved in 95%, v/v ethanol and the vouchers were kept in 10%, v/v formaldehyde. Digital photographs were taken for the specimens under study. All these materials are kept at NBFGR, Lucknow.

DNA Extraction
Approximately 50 mg of tissue was used for DNA isolation following standard phenol/chloroform method[11] with partial modifications. Estimation of the DNA concentration was done on 0.7% agarose gels in submarine gel casting units (Wealtec, USA). The qualitative and quantitative estimation was done by observing the bands in ultraviolet light on UV transilluminator (GE Healthcare). The DNA was diluted to get a final concentration of approximately 50 ng/µL for PCR amplification.

PCR Amplification

The complete cyt b gene was amplified using the universal set of primer pairs LI4724 (TGACTTGAARAACCAYCGTTG) and H15915 (CTCCGATCTCCGGATTACAAGAC)[12]. The 50 µL PCR reaction volume for cyt b region included 5.0 µL PCR buffer (10×), 1.0 µL dNTPs (2.5 mM each), 2.0 µL MgCl$_2$ (25 mM), each primer 0.5 µL (10 µM), *Taq* polymerase 0.5 µL (3 U/µL) and template DNA 2.0 µL (50 ng/µL). Amplifications were performed in MJ PTC-200 (BIORAD). The thermal regime for cyt b consists of initial denaturation of 3 min at 94°C, 35 cycles of 45 sec at 94°C, 50 sec at 50°C, 80 sec at 72°C, followed by final extension of 10 min at 72°C and held at 4°C. PCR products were visualized on 1.2% agarose gels and documented using gel documentation system (Biovis).

DNA Sequencing

Products with concentration range between 50 to 100 ng/µL were selected for sequencing. Sequencing was performed following the dideoxynucleotide chain termination method[13], using automated ABI 3730 sequencer. Products were labelled using the BigDye Terminator V.3.1 cycle sequencing kit (Applied Biosystems, Inc). Reagent quantity for one sequencing PCR reaction cocktail was Terminator ready reaction mix (2.5×) 8.0 µL, BigDye sequencing buffer (5×) 4.0 µL, PCR product (50 ng/µL) 1 µL, primer (3 µM) 1.0 µL (forward in one PCR tube and reverse in other tube), deionised water 6.0 µL, total volume 20 µL. Cycle sequencing PCR conditions were 96°C for 1 min and then 25 cycles of 96°C for 10 sec, 50°C for 5 sec, 60°C for 4 min. DNA sequencing was performed with forward primer and reverse primer to get forward and reverse strand sequences.

Sequence Analysis

Region of DNA sequences having Phred value above 20 (99% accuracy) were considered as good quality and selected for analysis. Reverse sequences were reversed and complimented, so as to make it parallel to forward sequence and then both are aligned. Ambiguities were referenced against the sequencing electropherograms or trace files and corrected accordingly. Full length consensus sequences were made from forward and reverse strands for all samples and aligned using ClustalW[14]. The consensus sequences were blasted in NCBI GenBank (http://www.ncbi.nlm.nih.gov) for the nearest similar sequence matches[15]. 16 sequences were submitted to GenBank and accession numbers are given with sample code in brackets: FJ349109 (8205A), FJ349110 (8412A), FJ349118 (8412I), FJ349119 (8412J), FJ349120 (8412K), FJ349123 (8412N), FJ349130 (8412V), FJ349133 (8412Y), FJ349134 (8412Z), FJ349137 (8412AC), FJ349138 (8412AD), FJ349140 (8412AF), FJ349141 (8412AL), FJ349142 (8412AO), FJ349143 (8412AP), FJ349144 (8412AQ). All sequences were aligned with Clustal W, a module integrated in software MEGA4[16]. All the nucleotide sequences were converted to amino acid sequences by defining the reading frame. Thus, this protein consists of 395 amino acids. The most common amino acid sequence was chosen for predicting the 3-D structure. The amino acid sequence of this target protein was submitted to NCBI (acc.no. ACR27009.1). It was ascertained that structure of cyt b in fishes is unavailable and, therefore, the work of structure prediction was undertaken.

Physical and Chemical Parameter Calculation

For calculation of physical and chemical parameters Expasy's ProtParam server was used. For physico-chemical characterization, theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues and extinction coefficient, the method given by Gill and Hippel[17] and integrated in Expasy's ProtParam, was used. Instability index[18], grand average hydropathy[19] and aliphatic index[20] were also computed using the Expasy's ProtParam.

Template Finding and Alignment of Template and Target Sequences

Best template is the basic key to generate high quality structural model. To find pertinent template for the modeling of target protein, template was searched through Geno3d server[21] and it was verified through PDBSum server[22]. From the comparative modeling, PDB 1BCC Chain C of cytochrome bc1 complex from chicken was selected as a template protein. The final alignment between template and target were generated by using align2d.py (Python script file). Default parameters were applied and the aligned sequences were inspected.

Secondary Structure Prediction

SOPMA[23] was used to calculate the secondary structural features of the cyt b sequence. The secondary structure was predicted by using default

parameters, viz., window width 17, similarity threshold 8 and number of states 4.

Model Generation and Refinement

By the software Modeller9v8, 100 models were constructed from the sequence alignment between cyt b and template protein[24]. The constructed rough models were solvated and subjected to constraint energy minimization with a harmonic constraint of 100 kJ/mol/Å^2 applied to all protein atoms, using steepest descent and conjugate gradient technique to eliminate bad contacts between protein atoms and structural water molecules. Computations were carried out *in vacuo* with the GROMOS96 43B1[25]. Parameter setting in Swiss PDB Viewer (SPDBV) was according to Guex and Peitsch[26].

Evaluation of Refined Model

In final step of comparative modeling, the refined structure of the model was subjected to a series of tests for its internal consistency and reliability. To evaluate backbone confirmation, Ramachandran plot (phi/psi) was obtained from Procheck analysis[27]. The Swiss PDB Viewer energy minimization test was applied to check for energy criterion in comparison with the potential of mean force derived from a large set of known protein structure. Packing quality of refined structure was investigated by the WHATIF[28]. This model was submitted in Protein Model DataBase (PMDB) with the accession number PM0076414.

Results and Discussion

The alignment score was 306235.8750 as calculated with the align2d.py. Based on the result obtained from mGenTHREADER program[29] and Geno3D server, the x-ray structure of the chain C, pdb|1BCC|C, i.e., a high resolution structure of cytochrome bc1 complex from chicken was selected as template. The Blast[14] of target protein with template protein 1BCC gave 6e-175 E-value, 76% identity and 1571 alignment score. Alignment between template and target protein along with amino acid detail is given in Fig. 1.

The secondary structure of cytochrome b protein was predicted by SOPMA (Self Optimized Prediction Method with Alignment). The secondary structure shows whether amino acid lies in a helix, strand, coil or β-turn. It was found that maximum residue lies in α-helix region (45.82%), 31.90% lies in random coil region, 17.97% in extended strand region and remaining 4.30% lies in β-turn region. α-Helix represents a way of saturating all the hydrogen bond donors and acceptors in the peptide backbone. Parameters showing the properties of proteins in nature are computed *in silico* for this protein using Expasy's ProtParam tool. Molecular weight of the protein was found to be 44384.4, theoretical isoelectric point (pI) 7.36 and the instability index (II) 32.14. Instability index less than 40 (as in our case) shows that predicted protein structure is stable. The aliphatic index (AI) which is defined as the relative volume of a protein occupied by aliphatic side chains (alanine, valine, isoleucine, and leucine) was 121.95, also a positive indicator for structural stability. Grand average of hydropathicity (GRAVY)[30] was 0.698, positive value indicates that protein is hydrophobic. Extinction coefficient (in units of M^{-1} cm^{-1}, at 280 nm measured in water) was 90090 when all Cys residues were in the form of cystine (oxidized) and 89840 when all Cys residues were reduced to cysteine. This is because cysteine does not absorb light at more than 260 nm, but Cystine does at 280 nm. Total number of negatively and positively charged amino acids was 18 in both the cases.

With the help of Modeller 9v8, 100 models of target protein were generated. Stereochemical parameters of the protein were calculated by using the programme PROCHECK. The main chain parameter like Ramachandran plot quality, peptide bond planarity, C-alpha chirality, overall G factor and the bad contact per 100 residues were within the limits. These parameters were even in better range for side chain. Best model was selected based on comparison of all stereochemical parameters. This was subjected to refinement by using energy minimization process. The model obtained after refinement was evaluated for quality of the Ramachandran plot (Fig. 2).

By using the software PROCHECK, it was found that out of 395 amino acids, 93.2% (368) fall in most favoured region, 6.6% fall in additionally allowed region and 0.3% fall in generously allowed region. Ramachandran plot, for glycine, pre-proline and proline, showed that all these three fall under allowed regions. The packing quality of predicted model was found in normal range as calculated with the Whatif server. With the help of VERIFY_3D[31] and ERRAT[32], tools available at Structural Analysis and Verification Server (SAVES), it was shown that 62.12% of the residues had an averaged 3D-1D score > 0.2 and overall quality factor 87.306.

```
gi|237872932|gb|ACR27009.1|    HYENLLMVTRKAHPLLKIINDALIDLPAPSNISVWWNFGSLLLLCLMTQI  50
gi|3659969|pdb|1BCC|C          ----MAPNIRKSHPLLKMINNSLIDLPAPSNISAWWNFGSLLAVCLMTQI  46
                                  :    **:*****:**::***********,.*******  :******

gi|237872932|gb|ACR27009.1|    LTGLFLAMHYTSDISTAFSSVAHICRDVNYGWIIRNLHANGASFFFICIY 100
gi|3659969|pdb|1BCC|C          LTGLLLAMHYTADTSLAFSSVAHTCRNVQYGWLIRNLHANGASFFFICIF  96
                               ****:*****:* * ****** **:*:***:*****************:

gi|237872932|gb|ACR27009.1|    LHIGRGLYYGSYLYKETWNIGVVLLLLVMMTAFVGYVLPWGQMSFWGATV 150
gi|3659969|pdb|1BCC|C          LHIGRGLYYGSYLYKETWNTGVILLLTLMATAFVGYVLPWGQMSFWGATV 146
                               ******************* **:*** :* ********************

gi|237872932|gb|ACR27009.1|    ITNLLSAVPYMGDALVQWIWGGFSVDNATLTRFFAFHFLLPFAIVAATVL 200
gi|3659969|pdb|1BCC|C          ITNLFSAIPYIGHTLVEWAWGGFSVDNPTLTRFFALHFLLPFAIAGITII 196
                               ****:**:**:*.:**:* *******,*******:*******.. *::

gi|237872932|gb|ACR27009.1|    HALFFHETGSNNPVGLNSDADKISFHPYFSYKDVLGFIVLITALASLALF 250
gi|3659969|pdb|1BCC|C          HLTFLHESGSNNPLGISSDSDKIPFHPYYSFKDILGLTLMLTPFLTLALF 246
                               * *:**:*****:*:.**:*.**:***:*:.:**.: :****

gi|237872932|gb|ACR27009.1|    SPNLLGDPENFTPANPLVTPPHIKPEWYFLFAYAILRSIPNKLGGVLALL 300
gi|3659969|pdb|1BCC|C          SPNLLGDPENFTPANPLVTPPHIKPEWYFLFAYAILRSIPNKLGGVLALA 296
                               *************************************************

gi|237872932|gb|ACR27009.1|    FSILVLMVVPLLHTSKQQGTTFRPLTQLLFWTLVADVFILTWIGGMPVEH 350
gi|3659969|pdb|1BCC|C          ASVLILFLIPFLHKSKQRTMTFRPLSQTLFWLLVANLLILTWIGSQPVEH 346
                               *:*:*::*:*.***: *****:* *** ***::;****** ****

gi|237872932|gb|ACR27009.1|    PFIIIGQIASLLYFSLFLILYPLAGWLENKLLNFWCPSSLVLKHR 395
gi|3659969|pdb|1BCC|C          PFIIIGQMASLSYFTILLILFPTIGTLENKMLNY----------- 380
                               *******:*** **:::***:* * ****:**:
```

Fig. 1—Alignment between template and target sequence
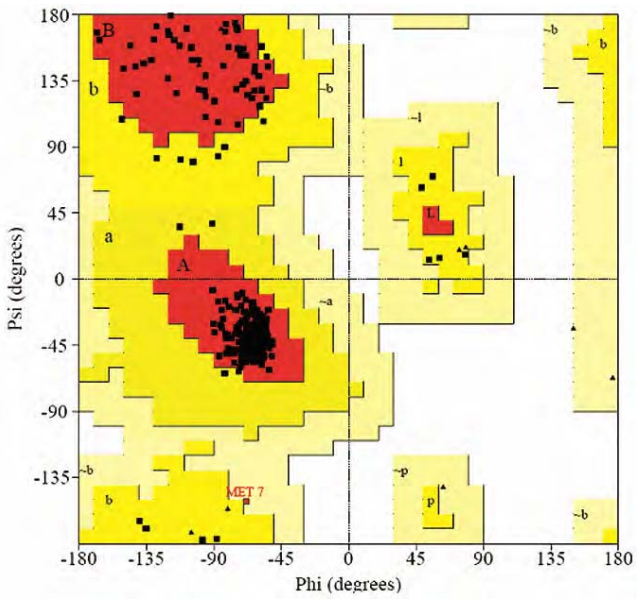

Fig. 2—Ramachandran plot of best model (model 4)

Ray diagram of 3-D structure of cyt b was generated by YASRA[33]. Cyt b is located across mitochondrial inner membrane (MIM) as shown in Fig. 3. The lower side is mitochondrial matrix and upper side is intermembrane space (IMS). Ray diagram shows 8 trans-membrane helices connected with loops towards IMS and matrix. Four highly conserved histidine residues are ligands of the iron atoms of two hemes. His84 and His183 are ligands for heme $b566/b_L$, and His98 and His197 are ligands for heme $b562/b_H$ as shown in Figs 1 and 3 (excluding initial 4 amino acids in target). Active site for quinol oxidation (Qo) is located towards intermembrane space and quinone reduction (Qi) active site is located towards matrix.

The wire diagram of template pdb 1BCC chain C and the predicted model PM0076414 (target) were superimposed to compare the 3-D structures of template and target proteins. As expected, Fig. 4 shows maximum residues of PM0076414 aligning with 1BCC chain C, showing conservation of active
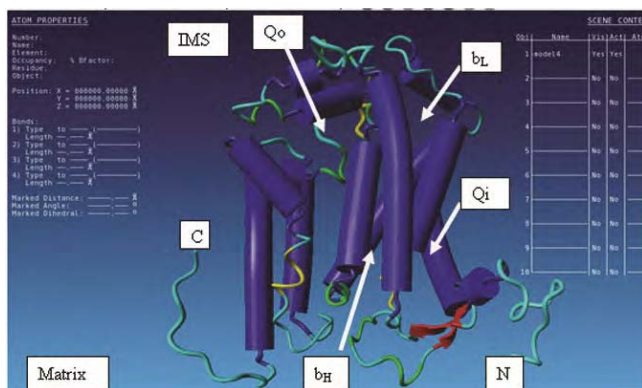
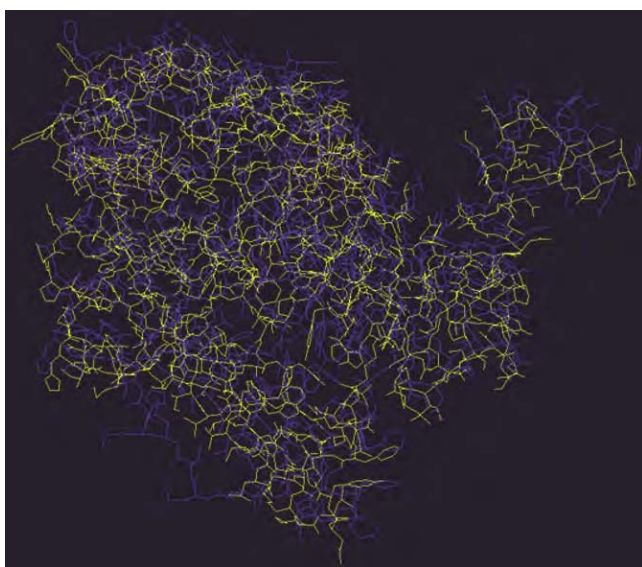Fig. 3—Ray diagram of 3-D structure of cytochrome b

Fig. 4—Superimposition of predicted model PM0076414 (yellow colour) and PDB 1BCC chain C (in blue colour) in Swiss Pdb Viewer 4.0.1[26]

sites of proteins in fish and chicken. It shows the retention of structure and function in two species, despite their large evolutionary distance. Comparison of amino acid sequences of template with target sequences has shown 76.3% (290/380) identical amino acids, 14.2% (54/380) conserved amino acids, 2.6% (10/380) semi conserved (having similar properties) amino acids and only 6.8% (26/380) different amino acids. The overall result provides the evidence that the predicted three dimensional structure of cyt b (PMDB ID PM0076414) is quite acceptable[34].

In silico study of protein, nucleic acid and other biomolecules is helpful for the purpose of simulation in almost all research fields. With molecular modeling, it is so easy to design new drugs and molecules which accurately bind to targets. Comparative modeling is becoming a useful technique in the field of bioinformatics because the knowledge of three dimensional structure of a protein would be an invaluable aid to delineate the structural and functional details of a particular protein. The three dimensional structure of cyt b may be further used in characterizing the protein and to understand its mechanism.

Acknowledgement

References

1   Sullivan J P, Lundberg J G & Hardman M, A phylogenetic analysis of the major groups of catfishes (Teleostei: Siluriformes) using rag1 and rag2 nuclear gene sequences, *Mol Phylogen Evol*, 41 (2006) 636-662.

2   Ferraris, C J Jr, Checklist of catfishes, recent and fossil (Osteichthyes: Siluriformes), and catalogue of siluriform primary types, *Zootaxa*, 1418 (2007) 1-628.

3   Bechmann G, Schulte U & Weiss H, Mitochondrial ubiquinol-cytochrome c oxidoreductase, in *Molecular mechanisms in bioenergetics*, edited by L Ernster. (Elsevier, Amsterdam) 1992, 199-216.

4   Nicholls D G & Ferguson S J, *Bioenergetics* (Academic Press, San Diego, California) 2002, 114-117.

5   Xia D, Yu C A, Kim H, Xia J Z, Kachurin A M *et al*, Crystal structure of the cytochrome BC1 complex from bovine heart mitochondria, *Science*, 277 (1997) 60-66.

6   Siregar J E, Syafruddin D, Matsuoka H, Kita K & Marzuki S, Mutation underlying resistance of *Plasmodium berghei* to atovaquone in the quinone binding domain 2 (Qo(2)) of the cytochrome b gene, *Parasitol Int,* 57 (2008) 229-232.

7   Martí-Renom M A, Stuart A C, Fiser A, Sánchez R, Melo F *et al*, Comparative protein structure modeling of genes and genomes, *Annu Rev Biophys Biomol Struct*, 29 (2000) 291-325.

8   Williamson A R, Creating a structural genomics consortium, *Nat Struct Biol*, 7 (2000) 953.

9   Gupta A, Pal S K, Narayan J & Potukuchi S M, *In silico* structure prediction of p27SJ, a novel protein in St John's Wort, that suppresses expression of HIV-1 genome, *Biofrontiers*, l (2009) 32-35.

10 Vishwanath W & Linthoingambi I, Fishes of the genus *Glyptothorax* Blyth (Teleostei: Sisoridae) from Manipur, India, with description of three new species, *Zoos' Print J*, 22 (2007) 2617-2626.

11 Sambrook J, Fritsch E F & Maniatis T, *Molecular cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press, New York) 1989.

12 Irwin D M, Kocher T D & Wilson A C, Evolution of cytochrome b in mammals, *J Mol Evol*, 32 (1991) 128-144.

13 Sanger F, Nicklen S & Coulson A R, DNA sequencing with chain terminating inhibitors, *Proc Natl Acad Sci USA*, 74 (1977) 5463-5467.

14 Thompson J D, Higgins D G & Gibson T J, CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting position specific gap penalties and weight matrix choice, *Nucleic Acids Res*, 22 (1994) 4673-4680.

15 Altschul S F, Gish W, Miller W, Myers E W & Lipman D J, Basic local alignment search tool, *J Mol Biol*, 215 (1990) 403-410.

16 Tamura K, Dudley J, Nei M & Kumar S, MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0, *Mol Biol Evol*, 24 (2007) 1596-1599.

17 Gill S C & von Hippel P H, Calculation of protein extinction coefficients from amino acid sequence data, *Anal Biochem*, 189 (1989) 283.

18 Guruprasad K, Reddy B V & Pandit M W, 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 (1990) 155-161.

19 Kyte J & Doolittle R F, A simple method for displaying the hydropathic character of a protein, *J Mol Biol*, 157 (1982) 105-132.

20 Ikai A, Thermostability and aliphatic index of globular proteins, *J Biochem*, 88 (1980) 1895-1898.

21 Combet C, Jambon M, Deléage G & Geourjon C, Geno3D: Automatic comparative molecular modeling of protein, *Bioinformatics*, 18 (2002) 213-214.

22 Laskowski R A, PDBsum: Summaries and analyses of PDB structures, *Nucleic Acids Res*, 29 (2001) 221-222.

23 Geourjon C & Deleage G, SOPMA: Significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, *Comput Appl Biosci*, 11 (1995) 681-684.

24 Eswar N, John B, Mirkovic N, Fiser A, Ilyin V A *et al*, Tools for comparative protein structure modeling and analysis, *Nucleic Acids Res*, 31 (2003) 3375-3380.

25 Walter R P S, Hünenberger P H, Tironi I G, Mark A E, Billetor S R *et al*, The GROMOS biomolecular simulation program package, *J Phys Chem A*, 103 (1999) 3596-3607.

26 Guex N & Peitsch M C, SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis*, 18 (1997) 2714-2723.

27 Laskowski R A, MacArthur M W, Moss D S & Thorton J M, PROCHECK: A program to check the stereochemical quality of protein structures, *J Appl Cryst*, 26 (1993) 283-291.

28 Vriend G, WHAT IF: A molecular modeling and drug design program, *J Mol Graph*, 8 (1990) 52-56.

29 McGuffin L J & Jones D T, Improvement of the GenTHREADER method for genomic fold recognition, *Bioinformatics*, 19 (2003) 874-881.

30 Kyte J & Doolittle R F, A simple method for displaying the hydropathic character of a protein, *J Mol Biol*, 157 (1982) 105-132.

31 Lüthy R, Bowie J U & Eisenberg D, Assessment of protein models with three dimensional profiles, *Nature (Lond)*, 356 (1992) 83-85.

32 Colovos C & Yeates T O, Verification of protein structures: Patterns of nonbonded atomic interactions, *Protein Sci*, 2 (1993) 1511-1519.

33 Krieger E, Koraimann G & Vriend G, Increasing the precision of comparative models with YASARA NOVA—A self-parameterizing force field, *Proteins*, 47 (2002) 393-402.

34 Castrignanò T, De Meo P D, Cozzetto D, Talamo I G & Tramontano A, The PMDB protein model database, *Nucleic Acids Res*, 34 (2006) D306-309.