Interdiscip Sci Comput Life Sci (2011) 3: 110–120
DOI: 10.1007/s12539-011-0071-z

In silico Prediction and Characterization of 3D Structure and Binding Properties of Catalase from the Commercially Important Crab, *Scylla serrata*

Biswaranjan PAITAL[1], Sunil KUMAR[2*], Rohit FARMER[3]
Niraj Kanti TRIPATHY[4], Gagan Bihari Nityananda CHAINY[1,5,6]
[1](Department of Zoology, Utkal University, Bhubaneswar 751004, India)
[2](Bioinformatics Centre, Institute of Life Sciences, Nalco Square, Bhubaneswar 751023, India)
[3](Department of Computational Biology and Bioinformatics, Jacob School of Biotechnology and Bioengineering, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Allahabad 211007, India)
[4](Departments of Zoology, Berhampur University, Berhampur 760007, India)
[5](Department of Biotechnology, Utkal University, Bhubaneswar 751004, India)
[6](Present address of the author; Imgenex India Pvt. Ltd., E-5, Infocity, Bhubaneswar 751024, India)

Received 30 September 2010 / Revised 22 November 2010 / Accepted 24 November 2010

Abstract: The enzyme catalase breaks down H_2O_2, a potentially harmful oxidant, to H_2O and O_2. Besides oxidase activity, the enzyme also exhibits peroxidase activity. Therefore, it plays an important role in maintaining health and regulating pathophysiology of the organisms. However, 3D structure of this important enzyme in invertebrates particularly in crabs is not yet available. Therefore, an attempt has been made to predict the structure of the crab catalase and to envisage its catalytic interaction with H_2O_2. A three dimensional model of crab catalase was constructed using the NADPH binding site on Beef Liver catalase from *Bos taurus* (PDBID: 7CAT) as template by comparative modeling approach. Backbone conformation of the modeled structure by PROCHECK revealed that more than 98% of the residues fell in the allowed regions, ERRAT results confirmed good quality of modeled structure and VERIFY3D profile was satisfying. Molecular docking has been used to know the binding modes of hydrogen peroxide with the crab catalase protein. The receptor structures used for docking were derived from molecular dynamics (MD) simulations of homology modeled structure. The docking results showed that the three important determinant residues Arg68, Val70 and Arg108 in catalase were binding with H_2O_2 as they had strong hydrogen bonding contacts with the substrate. Our analysis provides insight into the structural properties of crab catalase and defines its active sites for binding with substrate. These data are important for further studies of catalase of invertebrates in general and that of crabs in particular.
Key words: homology modeling, molecular dynamics, docking, *Scylla serrata*, catalase, structural biology, aquaculture, hydrogen peroxide.

1 Introduction

Catalase is an important antioxidant enzyme that converts H_2O_2, a powerful and potentially harmful oxidizing agent into water and molecular oxygen. Therefore, the enzyme plays an important role in regulating pathophysiology of the organisms (Mates *et al.*, 1999) and in graying process of human hair because its low level leads to accumulation of H_2O_2 which in turn starts bleaching hair from their roots (Wood *et al.*, 2009). Previous studies on catalase of higher vertebrates indicated that the enzyme exists as tetramer protein without any isomeric form. Each monomer of catalase has more than 500 amino acids with a molecular weight of 60 or 75 kDa. The enzyme contains a haem group in its active site. It is accessible from the surface of the enzyme through a hydrophobic channel. Catalase of some organisms such as cow contains NADPH as cofactor which prevents the formation of inactive catalase (McDowall *et al.*, 2010) besides protecting the enzyme from oxidation by its own substrate, *i.e.* H_2O_2. Haem-containing catalase of some species is also bifunctional in nature exhibiting both catalase (EC 1.11.1.6) and peroxidase enzyme (EC 1.11.1.7) activities (Mutsuda *et al.*, 1996).

An important limiting factor to the rate of any enzyme catalyzed reaction in solution is the frequency

*Corresponding author.
E-mail: skybiotech@gmail.com

with which enzyme and substrate molecules collide with each other. As the active site of catalase is buried deep in its structure, availability of H_2O_2 for catalysis immediately cannot be expected upon its collision with the enzyme. Because for catalysis, H_2O_2 has to move further in the hydrophobic channel of the active site in order to reach at its haem group. The structure and amino acid composition of the active site and the body of catalase of different organisms may vary more or less with each other. Therefore, catalases of different organisms have different activity level (Switala *et al.*, 2002). It may be due to differences in both structure and involvement of amino acids of their active site. So, it is an important task in oxidative stress research to draw a correlation between the structural arrangement and involvement of amino acids on the active site of catalase and its activity in different organisms. From this view point, 3D structural knowledge of catalase becomes highly important. It is reported that ectotherms are always vulnerable to environmental stress (Winston *et al.*, 1991). Accordingly, both transcript and translation products of catalase vary with respect to different environmental stress (Bohacova *et al.*, 2006). Our working model, the crab *Scylla serrata* has a great commercial value throughout the world. It is a euryhyaline and ectothermal animal. Its catalase activity is also reported to be altered with respect to environmental salinity stress (Paital *et al.*, 2010) which may be related to its growth and pathophysiology. The mechanism of altered catalase activity in this crab with respect to their altered physical environment may be attributed to the spatial arrangement and altered amino acid compositions of both the active site and other functionally and structurally important sites.

Biological function of a protein is often a reflection of its tertiary structure. Knowledge of the structural organization of the protein is a prerequisite for understanding its functional aspects. Owing to technical difficulties, there exists a vast gap in pace of protein sequence generation and experimental determination of the 3D structure. In absence of crystal structure, homology modeling provides a way to obtain structural insight into proteins. Although catalase gene and its translated products are well studied in higher group of animals, still its structural and functional aspects are scantly studied in lower organisms in general and particularly in crabs. In this report, we have generated structure of catalase enzyme of the crab *S. serrata* by comparative modeling approach. The models were subjected to molecular dynamics (MD) simulations to study the time evolution as well as time averaged values of structural properties. The possible interactions between the crab catalase and its substrate H_2O_2 have been predicted. The predicted structure can be served as a model for various experimental purposes in relation to the crabs pathophysiological conditions that may helps in its culture.

2 Materials and methods

2.1 Functional characterization of catalase of crab

Domains and motifs were predicted using InterProscan (Zdobnov *et al.*, 2001) and Multiple Em for Motif Elicitation (MEME) (Bailey *et al.*, 1994) respectively using default parameters to gain insight about its function. Disulphide bonds are important in determining the functional linkages.

2.2 Prediction of protein disorder

Protein disorder was predicted using DisEMBL (Linding *et al.*, 2003a), GLOBPLOT (Linding *et al.*, 2003b), RONN (Yang *et al.*, 2005) and Protein DisOrder prediction System (PRDOS) (Ishida and Kinoshita, 2007) server.

2.3 Template search and sequence alignment

The amino acid sequence of catalase of *Scylla serrata* (target) was retrieved from the sequence database of NCBI (http://www.ncbi.nlm.nih.gov) (ACX46120). The three-dimensional structure of the protein was not yet available in Protein Data Bank, hence the present exercise of developing the 3D model of the crab catalase of *Scylla serrata* was undertaken. BLASTP (Altshul *et al.*, 1990) search was performed against Brookhaven Protein Data Bank (PDB) (Bermen *et al.*, 2000) with the default parameters to find suitable templates for homology modeling. Based on the maximum identity with high score and lower e-value crystal structure of NADPH binding site on Beef Liver Catalase (PDB code: 7CAT) (Fita and Rossmann, 1985) was selected as template. The sequence identity and similarity between the target and template are 71% and 82%, respectively. The sequence alignment of catalase of *Scylla serrata* and 7CAT was carried out using the CLUSTALW (Thompson *et al.*, 1994) (http://www.ebi.ac.uk/clustalw) program.

2.4 3D structure generation

The academic version of MODELLER9v6 (http://www.salilab.org/modeler) (Sali and Blundell, 1993), was used for 3D structure generation based on the information obtained from sequence alignment. The MODELLER software employs probability density functions (PDFs) as the spatial restraints rather than energy (Sali *et al.*, 1993; Sali and Overington, 1994; Sali *et al.*, 1995). The 3D model of a protein is obtained by optimization of the molecular pdf such that the model violates the input restraints as little as possible. The molecular pdf was derived as a combination of pdfs restraining individual spatial features of the whole molecule. Out of 20 models generated by MODELLER, the one with the best G-score of PROCHECK (Laskoswki *et al.*, 1993), and with the best VERIFY3D (Eisenberg *et al.*, 1997) profile

was subjected to energy minimization. Using the parameters as a distance-dependent dielectric constant 1.0 and non binding cutoff of 14 Å, CHARMm (Brooks *et al.*, 1993), force field and CHARMm-all-atom charges, initially a steepest descent algorithm was used to remove close van der Waals contacts, followed by conjugate gradient minimization until the energy showed stability in sequential repetition. All hydrogen atoms were included during the calculation. The energy minimization was started with main chain of the core and then all core side chains were subjected to the same. All calculations were performed by using ACCELRYS DS Modeling 2.0 (Accelrys Inc. San Diego, CA 92121, USA) software suite. During these steps, the quality of the initial model was improved. VERIFY3D (a structure evaluation server) were used to check the residue profiles of the obtained three-dimensional models. STRIDE (Frishman and Argos, 1995), was used in prediction of secondary structure of the modeled crab protein. In order to assess the stereo-chemical qualities of the three dimensional models, PROCHECK analysis was performed. Quality evaluation of the model for the environment profile was also predicted using ERRAT (structure evaluation server) (Colovos and Yeates, 1993). The final refined model was evaluated for its atomic contacts using the Whatif program (Vriend, 1990) to identify bad packing of side chain atoms or unusual residue contacts. This model was further subjected for identification of active site and docking study.

2.5 Molecular dynamics simulation

The MD simulations of modeled catalase protein were performed with the GROMACS 4.0.6 software package (Hess *et al.*, 2008) using GROMOS 96 force field (Chris *et al.*, 2004) and the flexible SPC water model. The initial structure was immersed in a periodic water box of cubic shape (0.2 nm thick). Electrostatic energy was calculated using the particle mesh Ewald method (Essmann *et al.*, 1995). Cutoff distance for the calculation of the coulomb and van der Waals interaction was 1.0. After energy minimization using a steepest descent for 1000 steps, the system was subjected to equilibration at 300k and normal pressure for 100 ps under the conditions of position restraints for heavy atoms. LINCS (Hess *et al.*, 1997) constraints were performed for all bonds, keeping the whole protein molecule fixed and allowing only the water molecule to move to equilibrate with respect to the protein structure. The system was coupled to the external bath by the Berendsen pressure and temperature coupling (Berendsen *et al.*, 1984). The final MD calculations were performed for 400 ps under the same conditions except that the position restraints were removed. The results were analyzed using the standard software provided by the GROMACS package. An average structure was refined further using a steepest descent energy minimization.

2.6 Active site identification

Pocket Finder (Laurie and Jackson, 2005), a program for identifying and characterizing protein active sites, binding sites and functional residues located on protein surfaces was used to identify binding pockets of crab catalase. Pocket-Finder works by scanning probe radius 1.6 angstroms along all gridlines of grid resolution 0.9 angstroms surrounding the protein. Cubic diagonals were also scanned by using this probe. Grid points are defined to be a part of a site when the probe is within range of protein atoms followed by free space followed by protein atoms.

2.7 Protein ligand docking studies

The chemical structure of H_2O_2 molecule was extracted from pubchem database (http://pubchem.ncbi.nlm.nih.gov). Structure of the ligand was retrieved into two-dimensional MDL/SDF format and three dimensional coordinates were generated using the CORINA program (Tetko *et al.*, 2005). The molecules were then read into Discovery Studio 2.0 for further treatment such as energy minimization for 100 steps with CHARMm force field. Genetic optimization for ligand docking (GOLD) version 4.1.1 (Cambridge Crystallographic Data Centre, Cambridge, UK) was used for docking H_2O_2 for 50 times in the standard default settings (The standard default settings, consisting of population size-100, number os islands-5, selection pressure-1.1, niche size-2, migrate-10, cross over-95, number of operations-1,00,000, number of docking 10). For protein - ligand binding, 10 docking conformations (poses) were tested and the best GOLD score was selected for studies. GOLD uses a genetic algorithm to explore the full range of ligand conformational flexibility with partial flexibility of the protein (Jones *et al.*, 1997). Docking procedure consisted of three interrelated components: (a) identification of binding site, (b) a search algorithm to effectively sample the search space (the set of possible ligand positions and conformations on the protein surface), and (c) a scoring function. To estimate the protein-ligand complexes, the scoring function, GOLD score was employed which is based on four components: (1) protein-ligand hydrogen bond energy (external H-bond); (2) protein-ligand van der Waals energy (external vdw); (3) ligand internal van der Waals energy (internal vdw); (4) ligand interamolecular hydrogen bond energy (internal H-bond). The binding affinity between the protein and ligand was estimated by using the consensus scoring function X-Score V2.1 (Wang *et al.*, 2003). The ligand showing maximum interactions with the protein were plotted using the program LIGPLOT (Wallace *et al.*, 1995). Hydrogen bond interactions were double-checked with the software GETNEARES, available with the program DOCK (Ewing *et al.*, 2001).

3 Results

3.1 Model building

The structure of catalase protein in the crab *Scylla serrata* has determined by using homology modeling protocol. BLASTP search was performed against PDB with default parameters to find suitable templates for homology modeling. Based on the maximum identity with high score and lower e-value 7CAT was used as the template for homology modeling. Sequence alignment between catalase of crab and 7CAT was done using ClustalW program. The sequence - structure alignment used for model building shown in Fig. 1. The alignment was characterized by some insertions and deletions in the loop regions. Modeling was carried out from 3 to 515 residues followed by a rigorous refinement of the model by means of energy minimization using CHARMm force field. The final stable structure of crab catalase protein is shown in Fig. 2.

```
CLUSTAL  2.0.12  multiple sequence alignment

crab        ---RDRAAEQLNEFK--KNQTKEDVLTTGFGCPLSDKLNSLTVGPRGPILLQDIQLLDEM 55
7CAT        ADNRDPASDQMKHWKEQRAAQKPDVLTTGGGNPVGDKLNSLTVGPRGPLLVQDVVFTDEM 60
               ** *::*::.:*  :   * ****** * *:.*************:*:**: : ***

crab        AHFDRERIPERVVHAKGAGAFGYFEVTHDISQYTKAKIFSEIGKRTPLAVRFSTVGGESG 115
7CAT        AHFDRERIPERVVHAKGAGAFGYFEVTHDITRYSKAKVFEHIGKRTPIAVRFSTVAGESG 120
            ********************************::*:***:*..******:*******.****

crab        SADTARDPRGFAVKFYTEEGNWDLVGNNTPIFFIRDPVLFPSFIHTQKRNPATHLKDADM 175
7CAT        SADTVRDPRGFAVKFYTEDGNWDLVGNNTPIFFIRDALLFPSFIHSQKRNPQTHLKDPDM 180
            ****.*************:*****************.:*******:***** *****.**

crab        FWDFITLRPETTHQVSFLFSDRGTPDGYRHMNGYGSHTFKLVNKEGKPVYCKFHYKTDQG 235
7CAT        VWDFWSLRPESLHQVSFLFSDRGIPDGHRHMDGYGSHTFKLVNADGEAVYCKFHYKTDQG 240
            .*** :****: *********** ***:***:*********** :*:.************

crab        IKCLSAERADFLAGSDPDYAIRDLYNAISEGNYPSYTMYIQVMTFEQAEKWEFNPFDLTK 295
7CAT        IKNLSVEDAARLAHEDPDYGLRDLFNAIATGNYPSWTLYIQVMTFSEAEIFPFNPFDLTK 300
            ** **.* *  ** .****.:***:***: *****:*:*******.:** : ********

crab        VWPHADFPLIPVGRITLDRNPQNYFAEVEQLAFTPSNLVPGIEPSPDKMLQGRLFAYTDT 355
7CAT        VWPHGDYPLIPVGKLVLNRNPVNYFAEVEQLAFDPSNMPPGIEPSPDKMLQGRLFAYPDT 360
            ****.*:******::.*:*** *********** ***: ******************.**

crab        HRHRLGANYHQIPVNCPYRARSKNYQRDGPMTVNDNQTCAPNYFPNSFSGPMDCKQFEVP 415
7CAT        HRHRLGPNYLQIPVNCPYRARVANYQRDGPMCMMDNQGGAPNYYPNSFSAPEHQPSALEH 420
            ******.** ***********  ******** : ***  ****:*****.* .  .

crab        KEKLSGDVMRYSSADEDNFTQVCTFYKNVLNEEERQRLVNNIAGHIVNAQEFLQERAIKN 475
7CAT        RTHFSGDVQRFNSANDDNVTQVRTFYLKVLNEEQRKRLCENIAGHLKDAQLFIQKKAVKN 480
            : ::**** *:.**::**.*** *** :*****:*:** :*****: :** *:*::*:**

crab        FSQACPEYGAGIRSALNRIKAAQSSN 501
7CAT        FSDVHPEYGSRIQALLDKYNEEKPKN 506
            **:. ****: *:: *:: :  :..*
```

Fig. 1 Sequence alignment between crab catalase and template (7CAT).

3.2 Protein structure validation

The stereochemical quality of the homology model of crab catalase was analyzed using PROCHECK. It was found that the phi/psi angles of 91.2% residues fell in the most favored regions. The overall PROCHECK G-factor for the homology modeled structure was -0.07 (Table 1). PROCHECK commonly used for structure validation (Singh and Dubey, 2009; Singh *et al.*, 2008a and 2008b). The model was also validated using ERRAT graph (Fig. 3). The quality factor of 74.11 suggests that the model has good quality as a score higher than 50 is acceptable for a reasonable model. The energetic architecture as predicted by PROSA score was negative (−8.88) for the modeled protein. The value is quite nearer to that of template (−8.98), which indicates the reliability of the model. High quality of model is also confirmed from VERIFY 3D server as 72.69% of residues of modeled protein showed a score higher than 0.2. Results of WHATCHECK also indicate about the correctness of the modeled structure (Table 2). Based on these results, it was ascertained that the obtained structure has reasonably good quality. The final model was submitted to the Protein Model Data Base

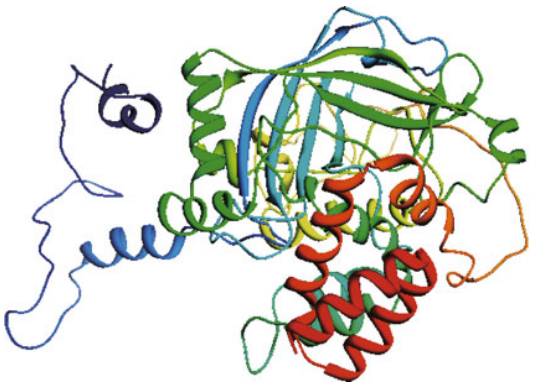

Fig. 2 Ribbon representation of modeled crab catalase protein. The α-helices and β-sheets are shown as helices and ribbons, respectively. The rest are shown as loops. The figure was prepared by PyMol

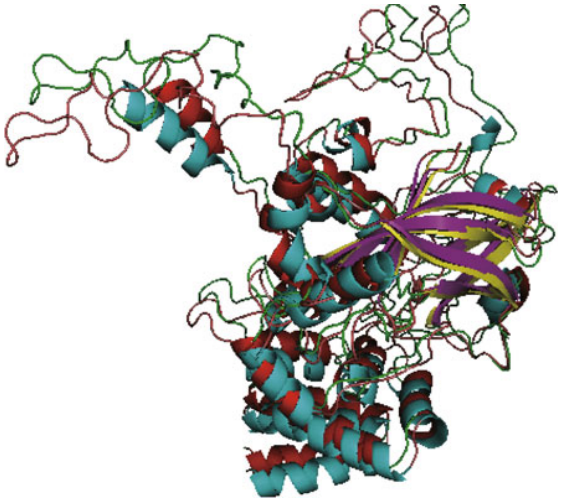

Fig. 3 Superimposition of Cα trace of crab catalase protein and 7CAT (template).

Table 1 Quality of main chain and side chain parameters of modeled crab catalase proteins giving comparative account of values observed for modeled structure (parameter value) and values obtained for well refined structures at same resolution (typical values)

Chain	*Stereochemical parameter*	*No. of data pts.*	*Parameter value*	*Typical value*	*Band width*	*No. of band widths from mean*
Main Chain Parameter	% distribution residues in A, B, L	430	91.2	83.8	10.0	0.7
	Omega angles st. deviation	495	3.8	6.0	3.0	−0.7
	Bad contacts/100 residues	4	0.8	4.2	10.0	−0.3
	Zeta angle st. deviation	465	1.3	3.1	1.6	−1.2
	H-bond energy st. deviation	268	0.8	0.8	0.2	−0.1
	Overall G-factor	497	−0.1	−0.4	0.3	1.1
Side Chain Parameter	Chi-1 gauche minus st. Dev.	66	5.4	18.1	6.5	−1.9
	Chi-1 trans st. dev.	143	8.7	19.0	5.3	−1.9
	Chi-1 gauche plus st. Dev.	190	7.0	17.5	4.9	−2.1
	Chi-1 pooled st. dev.	399	7.5	18.2	4.8	−2.2
	Chi-2 trans st. dev.	110	9.4	20.4	5.0	−2.2

Table 2 Quality indicators as calculated from WHATCHECK

Structure Z-score		*RMS-Z score*	
2nd generation packing quality	−2.512	Bond lengths	0.924
Ramachandran plot appearance	−0.289	Bond angles	1.272
Chi-1/chi-2 Rotamer normality	0.378	Omega angle restraints	0.699
Backbone conformation	−5.432	Side chain planarity	0.291
		Improper dihedral distribution	0.959
		Inside/Outside distribution	1.161

(http://mi.caspur.it/PMDB). The PMDB id for the submitted model was PM0076843.

The structural superimposition of Cα trace of the target model after MD simulation over template structure 7CAT (Fig. 4) resulted in a root mean square deviation (RMSD) of 1.3 Å (Z-score 8.3) using CE program (http://cl.sdsc.edu/ce.html). It indicates a valid structure of the model.

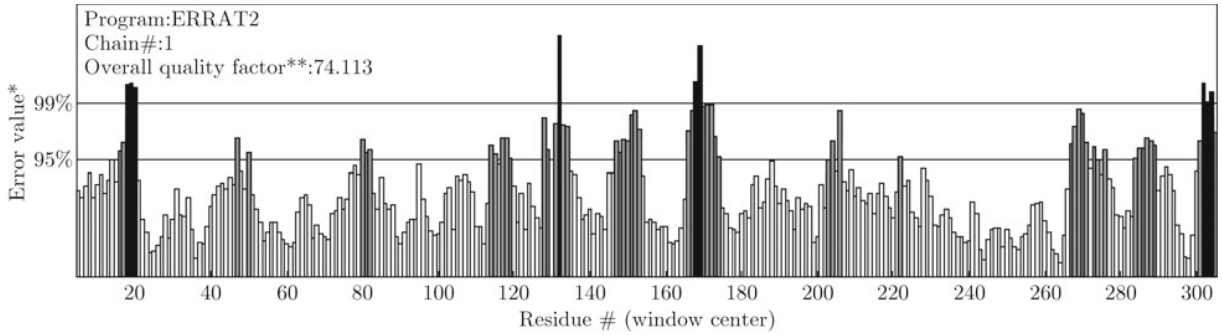

Fig. 4 The 3D profiles of crab catalase protein verified using ERRAT server. Overall quality score indicates residues are reasonably folded.

3.3 Molecular dynamics simulation

The MD simulation of the modeled catalase protein was performed and the resulted trajectory was analyzed to study the motional properties of the protein. The time evolution of root mean square deviation (RMSD) was computed for the modeled structure of the protein by taking the whole protein as initial structure (Fig. 5). It is evident from the Fig. 6 that RMSD increased slowly up to 320 ps and then decreases up to 350 ps then again slightly increases up to 400 ps and attained the equilibrium. Based on intrinsic dynamics, structural stability and improved relaxation of the modeled structure, the energy (Fig. 6) of the energy minimized structure was also calculated. The energy and RMSD calculations demonstrated that the protein is highly conserved in nature i.e. the protein is not much flexible. RMSF indicates the flexibility of the protein. RMSF of Cα is presented as a function of residue numbers (Fig. 7).

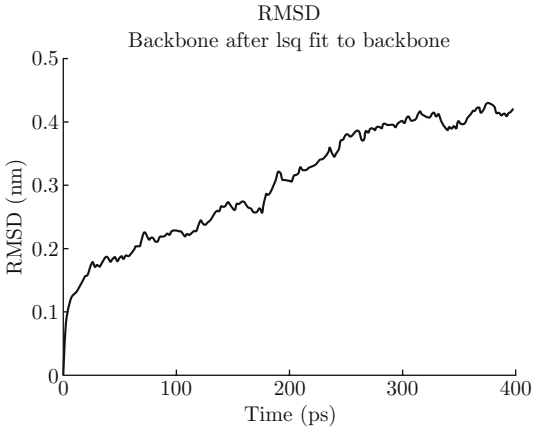

Fig. 5 Trajectories of the overall Cα (RMSD) of the crab catalase protein structure with respect to the starting structure over 400 ps MD simulation. The x axis represents the simulation time in picoseconds. The y axis represents RMSD in nm unit.

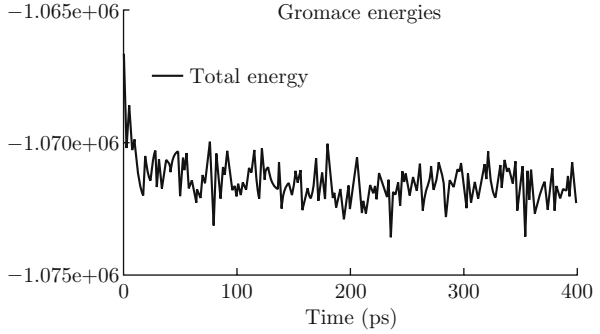

Fig. 6 Calculated energy vs time plot using GROMACS software. The x axis represents the simulation time in picoseconds. The y axis represents energy in nm unit.

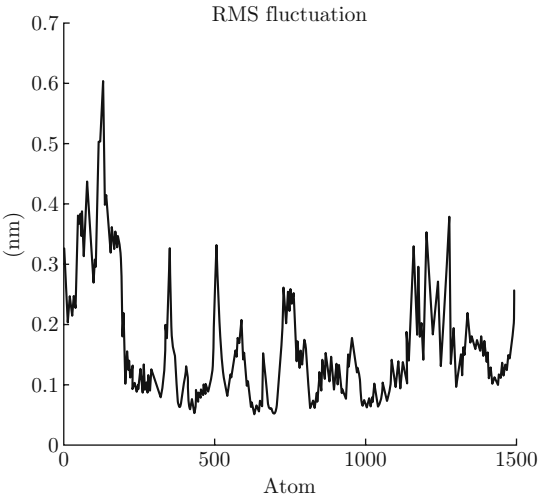

Fig. 7 Residue-wise RMSF profiles of Cα atoms of the catalase protein structure computed after stabilization of the RMSD trajectories. The x axis represents the residue number. The y axis represents RMSF in angstrom unit.

3.4 Active site identification of crab protein

After getting the final model, the possible binding sites of modeled catalase were searched using the Pocket-Finder server. Ten possible binding sites were obtained and these pockets were compared with active site of the template and were found that pocket1 is highly conserved in comparison to the template. Since the catalase protein and 7CAT are well conserved in both sequence and structure; it is predicted that

their biological function may be identical. From the structure-structure comparison of template and final refined models of catalase using SPDBV program (Guex and Peitsch, 1997), it was found that the residues in the pocket1 such as Arg68, Val69, Val70, His71, Arg108, Ser110, Val112, Gly113, Gly114, Arg123, Asp124, Pro125, Arg126, Gly127, Phe128, Ala129, Val142, Gly143, Asn144, Thr146, Ile148, Phe149, Phe150, Phe157, Pro158, Phe160, Ile161, His162, Thr163, Gln164, Lys165, Lys173, Asp174, Trp179, Phe181, Ile182, Gln191, Val192, Leu195, Phe196, Gly212, Ser213, His214, Leu295, Leu328, Ala329, Phe330, Met346, Gly349, Arg350, Ala353, Tyr354, Thr357, His358, Arg361, His462 are highly conserved in comparison to the active site of template. Therefore, pocket1 was chosen as the binding site to dock with the substrate hydrogen peroxide.

3.5 Protein-ligand interaction studies

Docking with hydrogen peroxide (H_2O_2) was carried out on the binding site of crab catalase protein using GOLD software applying the parameters of standard default setting with fifty Genetic Algorithm runs. To substantiate the estimations done by the GOLD program, we used consensus scoring program X-Score. The scoring schema used in the software X-Score computes a binding score for a given protein-ligand complex structure, and this binding score correlates to experimental binding constants well. The H_2O_2-catalase complex has best GOLD fitness value of 13.14. The predicted binding energy for the docked complex is −5.17 kcal/mol and predicted average −log Kd is 3.79 using X-Score program. Ligplot shows hydrogen bonding (R68, V70 and R108) and hydrophobic interacting residue (H71) with the H_2O_2 ligand in the binding site of catalase protein (Fig. 8).

4 Discussion

The crystal structure of catalase enzyme in invertebrates like the crab *Scylla serrata* has not yet been determined. Therefore, we built a model following comparative modeling approach to understand the structure of catalase. The fact that the 3D structure of catalase of crab is highly conserved justifies the utility of catalase comparative models in structural studies. The initial models used in this study were stereochemically good, with 91.2% residues falling in the most favored regions of the Ramachandran map and with a PROCHECK G-score of −0.07. The VERIFY3D plots for these models also showed satisfactory 3D-1D scores for all the residues in the sequence. The ERRAT graph result shows that the quality factor of the protein is 74.11 which is higher than 50, the recommended value for a reasonable model. The energetic architecture predicted by PROSA score was negative (−8.88) for the modeled protein which almost same for template (−8.98). The WHATCHECK results also indicate about the accuracy of the model. The structural superimposition of C^{α} trace of the model over its template 7CAT also validates the model. The structure was also simulated for 400 ps and this ensured sufficient relaxation of the structures. The above findings indicate that the predicted structure can be accepted as a good model for the obtained sequence. Domains and motifs indicate about the functional profile of a protein. Table 3 shows the location of predicted domains. Details

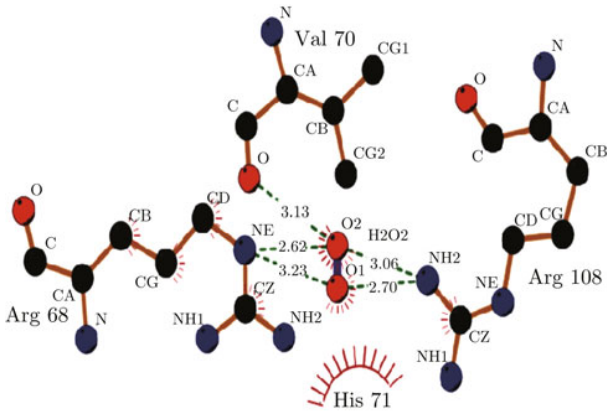

Fig. 8 Ligplot showing interacting residues (hydrogen bonding and hydrophobic) with the H_2O_2 in the binding site of crab catalase protein generated by Ligplot program.

Table 3 Location of domains predicted using Interproscan

PRINTS	*PR00067*	*CATALASE*	*3e-76 [37-60]T 3e-76 [100-118]T 3e-76 [121-138]T 3e-76 [140-158]T 3e-76 [305-332]T 3e-76 [337-363]T*
PFAM	Pf06628	Catalase-rel	1.3e-17[427-494]T
PROSITE	PS00437 PS00438	CATALASE_1 CATALASE_2	NA [350-358]T NA [60-76]T
GENE3D	G3DSA:2.40.180.10		3e-210[11-494]T
PROFILE	PS51402	CATALASE_3	79.833[20-499]T
PANTHER	PTHR11465	CATALASE	6.2E-266[99-496]T
SUPERFAMILY	SSF56634	Heme-dependent catalases	6.6e-227[16-494]T

about location and occurrence of motifs calculated using MEME server are summarized in Table 4. Disordered regions in a protein facilitate interactions of the protein with numerous proteins and allow more modification sites in the protein. Disordered regions were predicted using different servers and are summarized in Table 5. The intrinsic disorder profiles of crab protein obtained is illustrated in Figs. 9 (a), 9(b), 9(c) and 9(d).

Table 4 Motifs predicted using MEME

S. No.	*Motif*	*Width*	*Sites*	*Llr*	*E-value*	*Start position*	*p-value*	*Domain site*
1	Motif 1	6	3	47	1.0e+001	435	1.81e-8	VMRYSSADED NFTQVC FYKNVLNEE
						477	3.14e-08	QEFLQERAIK NFSQAC PEYGAGIRSA
						391	1.13e-06	NYQRDGPMTV NDNQTC NYFPNSFS
2	Motif 2	10	3	71	2.3e+001	220	3.21e-11	GYGSHTFKLV NKEGKPVYCK FHYKTDQGIK
						403	4.60e-11	NQTCAPNYFP NSFSGPMDCK QFEVPKEKLS
						365	1.60e-10	TDTHRHRLGA NYHQIPVNCP YRARSKNYQR
3	Motif 3	9	3	62	1.3e+003	171	5.46e-12	IHTQKRNPAT HLKDADMFW DFITLRPETT
						71	7.97e-09	DRERIPERVV HAKGAGAFG YFEVTHDISQ
						462	2.84e-08	RQRLVNNIAG HIVNAQEFL QERAIKNFSQ

Table 5 Predicted disordered region

Server	*Disordered*	*Disordered by REM465*	*Disordered by loops, coil definition*	*Disordered by HOTLOOPs definition*
PRDOS	1-21, 114-119, 390-392, 408-411, 497-517			
RONN	1-16, 108-126, 339-345, 376-398, 417-429, 498-514, 516-517			
GLOBPLOT		3-17, 54-70, 112-123, 242-248, 413-429, 493-515		
DISEMBL		1-9, 112-122, 497-517	15-48, 59-86, 97-105, 135-176, 182-190, 196-239, 244-256, 264-273, 285-321, 327-348, 358-436	11-22, 111-125, 364-375, 501-517

Each monomer of catalase contains a haem group which is present inside the structure of its active site. It is accessible from the surface through a hydrophobic channel. We aligned the template (7cat) and the target (crab catalase) structure along with haem using all atom superposition in a structural comparison program (Subbarao and Haneef, 1991) to construct a complex of haem along with the template. This complex was subjected to EM to study the binding residues involved in haem binding in this protein in order to understand similarities with other known proteins. There were hydrophobic interaction between haem and protein involved in binding of haem moiety (Fig. 10).

The docking studies of H_2O_2 binding to catalase is based on the stable receptor structure derived from MD simulation of homology models. The docking results indicate that conserved amino-acid residues in catalase enzyme play an important role in maintaining a functional conformation and are directly involved in donor substrate binding. From the docking results, the binding of the ligand to the active site of catalase seems to predominantly depend on the hydrogen bonding interactions that the ligand makes with the protein, indicating that binding specificity arises largely as a consequence of the volume and geometry of the binding site. The functional aspect of the protein is mainly determined by the composition and structural arrangement of the amino acid residues of its active site. The possible role of Tyr^{357}, His^{74} and Asn^{147} for enhancing H_2O_2 catalysis is described in bovine catalase (Boon *et al.*, 2010). From the docking experiment, we found four amino acids residues are coordinated and present in vicinity to have possible maximum interaction with H_2O_2 (Fig. 10). Unlike Tyr and Asn in bovine catalase, two Arg residues at 68 and 108 positions and a Val at 70 position along with His at 71 position seem to inter-

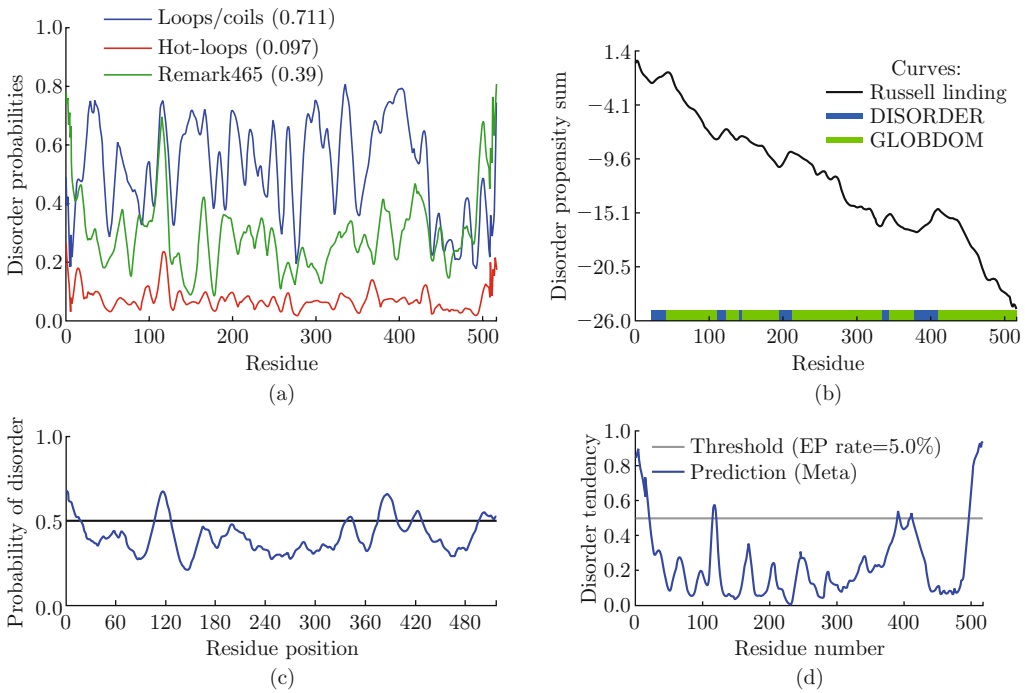

Fig. 9 Disordered regions in the crab catalase protein. (a) Disorder by DIEMBL; (b) Disorder by Globplot; (c) Disorder by RONN; (d) Disorder by PRDOS

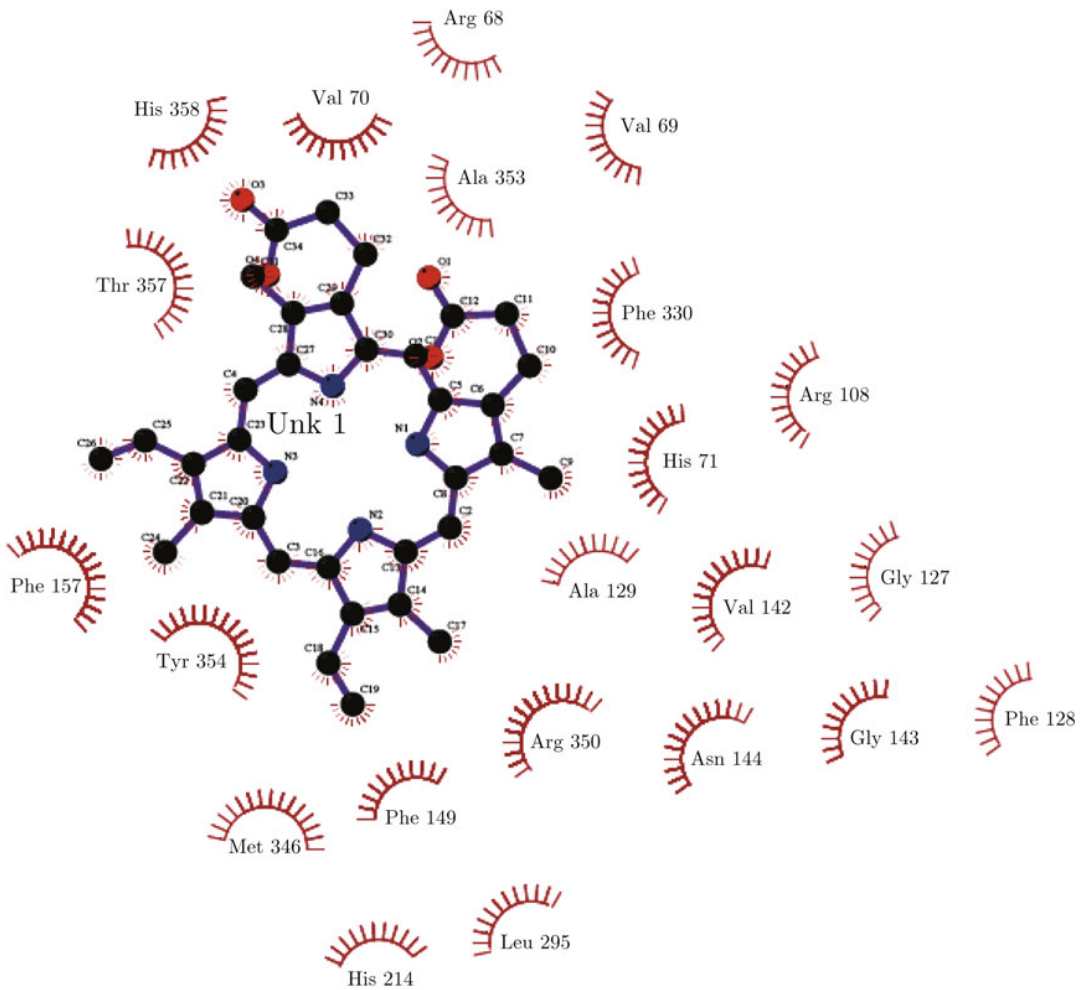

Fig. 10 Ligplot for interaction of haem with catalase protein.

act to make the 1st phase of peroxide bond breakage in case of crabs catalase. Further, in higher groups of animals Asn in the active site of catalase is also believed to have important role in the 1st step of catalytic reaction. Apparently, we did not find any involvement of Asn in case of crab catalase.

The interaction between the enzyme and the substrates proposed in this study are useful for understanding the potential mechanism of enzyme and the substrate binding. It is well known that hydrogen bonds play an important role for the structure and function of biological molecules particularly for the enzyme catalysis. The GOLD result (Section 3.5), indicates that Arg68, Val70 and Arg108 are important for the strong hydrogen bonding interaction with the substrate. The interaction between the enzyme and the substrates proposed in this study may be useful for understanding the potential mechanism of catalase enzyme and its substrate binding in invertebrates. The predicted model of crab catalase can be useful for different experimental purposes in relation to the pathophysiological conditions of crabs which in turn may helps in their culture.

5 Conclusion

Our analysis provides insight into the structural properties of crab catalase and defines the active binding sites with substrate. These data are importance for further studies on catalases of invertebrates in general and that of crabs in particular.

Acknowledgements SK is thankful to the Director, Institute of Life Sciences. BRP is extremely grateful to the University Grants Commission, New Delhi for providing a fellowship under RFSMS scheme (F.4-3/2006 (XI plan/BSR/7-173/2007(BSR). GBNC is thankful to University Grants Commission for sanction of a scheme (No. F.No. 31-284/2005 (SR)) under Major Research Project head. The authors gratefully acknowledge the three anonymous referees and the editor for their critical and helpful comments.

References

[1] Altshul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. Basic local alignment search tool. J Mol Biol 215, 403–410.

[2] Bailey, T.L., Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: Proceedings of the 2nd International Conference on Intelligent Systems for Molecular Biology, Menlo Park, California, 28–36.

[3] Berendsen, H.J.C., Postma, J.P.M., Gunsteren, W.F., DiNola, A., Haak, J.R. 1984. Molecular dynamics with coupling to an external bath. J Chem Phys 81, 368–3690.

[4] Bermen, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. 2000. The protein data bank. Nucleic Acids Res 28, 235–242.

[5] Bohacova, V., Zamocky, M., Godocikova, J., Buckova, M., Polek, B. 2006. The Expression and diversity of catalases in isolates of genus comamonas in response to the oxidative stress of a polluted environment. Curr Microbiol 53, 430–434.

[6] Boon, E.M., Downs, A., Marcey, D. 2010. Catalase: H_2O_2: H_2O_2 Oxidoreductase. http://www.callutheran.edu/BioDev/omm/catalase/cat1.htm.

[7] Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. 1993. CHARMm: A program for macromolecular energy minimization and dynamics calculations. J Comp Chem 4, 187–217.

[8] Chris, O., Alessandra, V., Alan, E., Wilfred, F.V.G. 2004. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25, 1656–1676.

[9] Colovos, C., Yeates, T.O. 1993. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci 2, 1511–1519.

[10] Eisenberg, D., Luthy, R., Bowie, J.U. 1997. VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol 277, 396–404.

[11] Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G. 1995. A smooth particle mesh ewald method. J Chem Phys 103, 8577–8593.

[12] Ewing, T.J., Makino, S., Skillman, A.G., Kuntz, I.D. 2001. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule database. J Comput Aided Mol Des 15, 411–428.

[13] Fita, I., Rossmann, M.G. 1985. The NADPH binding site on beef liver catalase. Proc Natl Acad Sci USA 82, 1604–1608.

[14] Frishman, D., Argos, P. 1995. Knowledge-based protein secondary structure assignment. Proteins 23, 566–579.

[15] Guex, N., Peitsch, M.C. 1997. SWISS-MODEl and the Swiss-PdbViewer: An environment for comparative protein modeling. Electrophoresis 18, 2714–2723.

[16] Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M. 1997. LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18, 1463–1472.

[17] Hess, B., Kutzner, C., Spoel, D., Lindahl, E. 2008. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4, 435–447.

[18] Ishida, T., Kinoshita, K. 2007. PrDOS: Prediction of disordered protein regions from amino acid sequence. Nucleic Acids Res 35, W460–W464.

[19] Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R. 1997. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 267, 727–748.

[20] Laskoswki, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M. 1993. PROCHECK: A program to check the sterochemical quality of protein structures. J Appl Cryst 26, 283–291.

[21] Laurie, A.T.R., Jackson, R.M. 2005. Q-SiteFinder: an energy-based method for the prediction of protein-ligand binding sites. Bioinformatics 21, 1908–1916.

[22] Linding, R., Jensen, L.J., Diella, F., Bork, P., Gibson, T.J., Russell, R.B. 2003a. Protein disorder prediction: implications for structural proteomics. Structure 11, 1453–1459.

[23] Linding, R., Russell, R.B., Neduva, V., Gibson, T.J. 2003b. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res 31, 3701–3708.

[24] Mates, J.M., Sanchez-Jimenez, F. 1999. Antioxidant enzymes and their implications in pathophysiologic process. Front Biosci 4, 339-345.

[25] McDowall, J. 2010. Catalase. http://www.ebi.ac.uk/interpro/potm/2004_9/Page1.htm).

[26] Mutsuda, M., Ishikawa, T., Takeda, T., Shigeoka, S. 1996. The catalase-peroxidase of Synechococcus PCC 7942 purification, nucleotide sequence analysis and expression in *Escherichia coli.* Biochem J 15, 251–257.

[27] Paital, B., Chainy, G.B.N. 2010. Antioxidant defence and oxidative stress parameters in tissues of mud crab (*Scylla serrata*) with reference to changing salinity. Comp Biochem Physiol C 151, 142–151.

[28] Sali, A., Blundell, T.L. 1993. Comparative protein modeling by satisfaction of spatial restraints. J Mol Biol 234, 779–815.

[29] Sali, A., Overington, J.P. 1994. Derivation of rules for comparative protein modeling from a database of protein structure alignments. Protein Sci 31, 1582–1596.

[30] Sali, A., Matsumoto, R., McNeil, H.P., Karplus, M., Stevens, R.L. 1993. Three-dimensional models of four mouse mast cell chymases, identification of proteoglycan-binding regions and protease-specific antigenic epitops. J Biol Chem 268, 9023–9034.

[31] Sali, A., Pottertone, L., Yuan, F., Vlijmen, V.H., Karplus, M. 1995. Evaluation of comparative protein modeling by MODELLER. Proteins 23, 318–326.

[32] Singh, B.K., Dubey, V.K. 2009. *In Silico* studies on tryparedoxin peroxidase of *Leishmania infantum*: Structural aspects. Curr Pharma Biotech 10, 626–630.

[33] Singh B.K., Nandini, S., Jagannadham, M.V., Dubey, V.K. 2008a. Modeled structure of trypanothione reductase of *Leishmania infantum*. BMB Reports 41, 444–447.

[34] Singh B.K., Nandini, S., Dubey, V.K. 2008b. Modeled structure of trypanothione synthetase of *Leishmania infantum* for development of novel therapeutics for leishmaniasis. Curr Trend Biotech & Pharma 2, 390–395.

[35] Subbarao, N., Haneef, I. 1991. Defining topological equivalences in macromolecules. Protein Eng 4, 877–884.

[36] Switala, J., Loewen, P.C. 2002. Diversity properties among catalase. Arch Biocehm Biophys 401, 145–154.

[37] Tetko, I.V., Gasteiger, J., Todeschini, R., Mauri, A., Livingstone, D., Ertl, P., Palyulin, V.A., Radchenko, E.V., Zefirov, N.S., Makarenko, A.S., Tanchuk, V.Y., Prokopenko, V.V. 2005. Virtual computational chemistry laboratory - design and description. J Comput Aided Mol Des 19, 453–463.

[38] Thompson, J.D., Higgins, D.G., Gibson, T.J. 1994. 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, 4673–4680.

[39] Vriend, G. 1990. WHAT IF: A molecular modeling and drug design program. J Mol Graph 8, 52–56.

[40] Wallace, A.C., Laskowski, R.A., Thornton, J.M. 1995. LIGPLOT: A program to generate schematic diagrams of protein-ligand interaction. Protein Eng 8, 127–134.

[41] Wang, R., Lu, Y., Wang, S. 2003. Comparative evaluation of 11 scoring functions for molecular docking. J Med Chem 46, 2287.

[42] Winston, G.W., Giulio, R.T.D. 1991. Prooxidant and antioxidant mechanisms in aquatic organisms. Aquat Txicol 19, 137–161.

[43] Wood, J.M., Decker, H., Hartmann, H., Chavan, B., Rokos, H., Spencer, J.D., Hasse, S., Thornton, M.J., Shalbaf, M., Paus, R., Schallreuter, K.U. 2009. Senile hair graying: H_2O_2-mediated oxidative stress affects human hair color by blunting methionine sulfoxide repair. Faseb J 23, 2065–2075.

[44] Yang, Z.R., Thomson, R., McNeil, P., Esnouf, R.M. 2005. RONN: The bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics 21, 3369–3376.

[45] Zdobnov, E.M., Apweiler, R. 2001. InterProScan-an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17, 847–848.