J Vector Borne Dis 46, September 2009, pp. 171–183

Research Articles

Structural model of the *Plasmodium falciparum* Thioredoxin reductase: a novel target for antimalarial drugs

Amit Kumar Banerjee, Neelima Arora & U.S.N. Murty

Bioinformatics Group, Biology Division, Indian Institute of Chemical Technology, Hyderabad, India

Abstract

Background: Malaria, a scourge of mankind, imposes a huge socioeconomic burden in tropical countries. Emergence of multi-drug resistant malarial parasites impels us to explore novel drug targets. Thioredoxin reductase is a promising antimalarial drug target.

Methods: The Thioredoxin reductase enzyme of *Plasmodium falciparum* was characterized *in silico* and protein disorder was predicted using available online tools. Since the crystal structure of Thioredoxin reductase of *P. falciparum* is not yet available, its three-dimensional structure was constructed by homology modeling using the high-resolution Thioredoxin reductase type 2 of mouse as a template. Obtained model was further refined by Molecular Dynamics (MD).

Results: The model was stable during the simulation with the equilibrium root mean square deviation (RMSD) value of 1.2 Å. Stereochemical evaluation revealed that 99.1% residues of the constructed model lie in the most favoured and allowed regions, thus, indicating a good quality model.

Conclusion: Results of this study will provide an insight into the structure of the Thioredoxin reductase of malarial parasite and aid in rational drug designing.

Key words Homology modeling – malaria – oxidative stress – *Plasmodium* – redox system – Thioredoxin reductase

Introduction

Malaria is a life threatening disease caused by four species of *Plasmodium* belonging to apicomplexa. The perilous malaria accounts for approximately three million deaths and nearly five billion episodes of clinical illness annually[1]. Malaria was ranked the 8th highest contributor to the disability adjusted life years (DALY) leading to almost 3% of DALY globally[2]. This pandemic affects poorest of population residing in 107 countries[3]. The torment due to morbidity and debility and loss of productive man hours is colossal. Vicious cycle of malaria and poverty continues in its most severe form in the developing economies where poorest of poor do not have access to unaffordable costly treatment. Malaria is tightening its grip afflicting more number of people and is spreading to parts of globe where it was non-existent. Despite concerted efforts employing sanitation measures, insecticide sprays and treatment programmes, malaria eradication remains a far-fetched dream. The resurgence of drug-resistant *P. falciparum*, the most fatal human malarial parasite is one of the major problems in the battle against malaria. Several marvelous antimalarial drugs like chloroquine have turned futile owing to emergence of multi-drug resistant strains of *Plasmodium*. This augments an immediate need to explore and validate new drug targets in the parasite for chemotherapeutic interventions. Drug pipeline for malaria is nearly dry due to indifference of pharmaceutical

companies towards tropical diseases and their shifting focus to life-style diseases in order to gain more profits. This grave situation has raised an alarm in the international community and has resulted in increased collaboration among pharmaceutical industry and government for exploring novel drug targets and new antimalarial drugs. The hunt for novel drug targets has been triggered and sustained by dawn of structure-based drug design approaches.

Plasmodium inhabits the human erythrocytes for a fraction of its intricate life-cycle. The pathophysiology of malaria is caused by the developmental stages of *Plasmodium* and these stages are targeted for effectual drug interventions for malaria. An oxidative stress is known to occur in *Plasmodium* infected erythrocytes which render them vulnerable to oxidative challenge[4–6]. In order to thrive, the parasite requires efficient antioxidants to shield itself against damages like nucleic acid modifications, lipid peroxidation and oxidation of thiol containing proteins, caused by reactive oxygen species (ROS). Thus, approach to disrupt parasite's antioxidative systems by focusing on redox systems in *Plasmodium* holds promise for inhibiting the development and survival of the parasite in the host cell.

P. falciparum Thioredoxin reductase (PfTrxR) , belonging to the family of dimeric flavoenzymes, which includes lipoamide dehydrogenase, glutathione reductase, and mercuric ion reductase, is a high molecular weight enzyme that reduces Thioredoxin (Trx)[7]. Inhibition of this enzyme is likely to affect the parasite at several vulnerable points resulting in enhanced oxidative stress, ineffective DNA synthesis and cell division and disturbed redox regulatory processes. The significant structural and mechanistic difference of PfTrxR from the human counterpart makes it a promising drug target[8].

It has been inferred that there is not much information available on characterization of TrxR protein sequence of *P. falciparum* 3D7 involved in redox mechanism. Unavailability of 3-dimensional structure of enzymes is one of the major hindrances in elucidating the interactions of enzymes with possible inhibitors. Comparative homology modeling is promulgated as the most unswerving computer-based technique for deciphering the 3D structures in the absence of crystal structure of the protein. This article describes the modeling of TrxR of *P. falciparum* which will provide insight into its structure and aid in drug designing.

Material & Methods

The primary sequence of the TrxR (Accession No. XP_001352109.1) of *P. falciparum* 3D7 was obtained from the public domain protein sequence database of NCBI (*http://www.ncbi.nlm.nih.gov*). The obtained sequence of PfTrxR was characterized *in silico* using Expasy-ProtParam tool[9] (*http://www.expasy.ch/tools/protparam.html*). The protein has 541 amino acid residues and the estimated molecular weight is 59686.2 with an isoelectric point (pI) of 7.81. The instability index (II) is computed to be 35.60 indicating the stability of protein. The Grand average of hydropathy (GRAVY) index value of –0.287 shows its hydrophilic nature. Conservation of residues was calculated using ConSeq server[10], a web server for the identication of biologically important residues in protein sequences. Since the three-dimensional structure of the protein was not available in Protein Data Bank (PDB) (*http://www.rcsb.org/pdb*), the present task of developing the 3D model of Thioredoxin reductase was undertaken.

Secondary structure prediction: Secondary structure of this protein was predicted using different methods, viz. Double prediction method (DPM)[11], Discrimination of protein secondary structure class (DSC)[12], GOR4[13], Hierarchical neural network (HNN)[14], PHD[15], Predator[16], SIMPA96[17], Self-optimized prediction method with alignment (SOPMA)[18] and Sec.Cons[19].

Prediction of intrinsic disorder: In order to identify regions of higher flexibility, the tools DisEMBL[20], Globplot[21], Regional order neural network (RONN)[22] and Protein disorder prediction system

(PRDOS)[23] were applied to select complementary regions for the identification of consistent problematic regions in the protein.

Homology modeling of PfTrxR: Our aim was to gain an insight into the structural characteristics of the Thioredoxin reductase protein of *P. falciparum* employing comparative modeling techniques. Molecular modeling, being the method of choice in the absence of experimentally determined crystal structure, can provide rationally good and accurate structural model for wide array of applications.

The methodology used to derive the PfTrxR model is partitioned into four phases: template selection, sequence alignment, model generation followed by refinement and model evaluation.

Template selection: Similarity search was performed using Position specific iteration-basic local alignment search tool (PSI-BLAST)[24] against PDB database keeping default parameters like E-value threshold 10, word size 3 and Blosum 62 Matrix. Total three iterations of PSI-BLAST were considered as the BLAST search results converged after three iterations. So, high-resolution X-ray crystallography structure of the TrxR type 2 of mouse (PDB ID: 1ZDL: A chain) at resolution of 3 Å obtained from PDB[25] was selected as a template protein showing 43.1% identity with the target protein.

Target-template alignment: A sequence alignment of template and target was performed using the ClustalX program[26] and as illustrated in Fig.1. The Blosum scoring matrix was selected with a gap penalty of 10

```
target          MCKDKNEKKNYEHVNANEKNGYLASEKNELTKNKVEEHTYDYDYVVIGGGPGGMASAKEA
template        ----------------------------------------QSFDLLVIGGGSGGLACAKEA
                                                        .:* :*****.**:*.****
target          AAHGARVLLFDYVKPSSQGTKWGIGGTCVNVGCVPKKLMHYAGHMGSIFKLDSKAYGWKF
template        AQLGKKVAVADYVEPSPRGTKWGLGGTCVNVGCIPKKLMHQAALLGGMIR-DAHHYGWEV
                *  * :* : ***:**.:*****:*********:****** *. :*.::: *:: ***:.
target          DN-LKHDWKKLVTTVQSHIRSLNFSYMTGLRSSKVKYINGLAKLKDKNTVSYYLKGDLSK
template        AQPVQHNWKTMAEAVQNHVKSLNWGHRVQLQDRKVKYFNIKASFVDEHTVRGVDKG--GK
                 : ::*:**.:. :**.*::***:.: . *:. ****:*  *.: *::**    **  .*
target          EETVTGKYILIATGCRPHIPDDVEGAKELSITSDDIFSLKKDPGKTLVVGASYVALECSG
template        ATLLSAEHIVIATGGRPRYPTQVKGALEYGITSDDIFWLKESPGKTLVVGASYVALECAG
                   ::.::*:**** **: * :*:** * .******* **:.****************:*
target          FLNSLGYDVTVAVRSIVLRGFDQQCAVKVKLYMEEQGVMFKNGILPKKLTKMDD-KILVE
template        FLTGIGLDTTVMMRSIPLRGFDQQMSSLVTEHMESHGTQFLKGCVPSHIKKLPTNQLQVT
                **..:* *.** :*** ******* :  *. :**.:*. * :* :*.::.*:   :: *
target          FSDKTS-----ELYDTVLYAIGRKGDIDGLNLESLNMNVNKSNNKIIADHLSCTNIPSIF
template        WEDHASGKEDTGTFDTVLWAIGRVPETRTLNLEKAGISTNPKNQKIIVDAQEATSVPHIY
                :.*::*       :****:****  :   ****. .:..* .*:***.*  ..*.:* *:
target          AVGDVAENVPELAPVAIKAGEILARRLFKDSDEIMDYSYIPTSIYTPIEYGACGYSEEKA
template        AIGDVAEGRPELTPTAIKAGKLLAQRLFGKSSTLMDYSNVPTTVFTPLEYGCVGLSEEEA
                *:*****. ***:*.*****::**:*** .*. :**** :**:::**:***. * ***:*
target          YELYGKSNVEVFLQEFNNLEISAVHRQKHIRAQKDEYDLDVSSTCLAKLVCLKNEDNRVI
template        VALHGQEHVEVYHAYYKPLEFTVADRD--------------ASQCYIKMVCMREPPQLVL
                  *:*:.:***:   :: **::...*:              :* *  *:**:::  : *:
target          GFHYVGPNAGEVTQGMALALRLKVKKKDFDNCIGIHPTDAESFMNLFVTISSGLSYAAKG
template        GLHFLGPNAGEVTQGFALGIKCGASYAQVMQTVGIHPTCSEEVVKLHISKRSGLEPTVTG
                *:*::**********:**.::  ..  :. : :***** :*..::*.::  ***. :..*
target          GCGGGKCG
template        --------
```

Fig. 1: Sequence alignment between *Plasmodium falciparum* Thioredoxin reductase and template 1ZDL_A

and a gap extension penalty of 0.05. The alignment was used for consequent model generation.

Model generation: The model was generated using a comparative modeling program MODELLER9v3[27] which generates a refined three dimensional homology model of a protein sequence based on a given sequence alignment and selected template. MODELLER employs probability density functions (PDFs) derived analytically using statistical mechanics and empirically using a database of known protein structures as the spatial restraints rather than energy[27]. MODELLER infers distance and angle constraints from a template structure and coalesce them with energy terms for sufficient stereochemistry in an objective function which is afterward optimized in Cartesian space with conjugate gradients and molecular dynamic methods.

Molecular dynamics simulation: A total of 20 models were generated by MODELLER. The structure selected on the basis of least objective function was subjected to molecular dynamics and equilibration method. The calculations were done using NAMD2.5 (Nanoscale Molecular Dynamics) software[28] employing CHARMM27 (Chemistry at Harvard Macromolecular Mechanics)[29] force field for lipids and proteins[30] along with the TIP3P model for water molecules[31]. The energy of the structure was minimized with 100,000 steps in 250,000 runs and 500 ps of molecular dynamics assuming a cut-off of 12 Å (switching function starting at 10 Å) for Van der Waals interactions. Spherical periodic boundary conditions were also taken into consideration for this study. An integration time step of 2 fs was applied, allowing a multiple time stepping algorithm[32,33] involving covalent bond interactions computed for every time step. For every two time steps, short range non-bonded interactions were computed and in every four time steps, long-range electrostatic forces were computed. The pair list of the non-bonded interaction was recalculated for every four time process. The pair list of the non-bonded interaction was recalculated for every ten time steps with a pair list distance of 13.5 Å. The short range non-bonded interactions were defined as Van der Waals and electrostatic interactions between particles within 12 Å. A smoothing function was employed for the Van der Waals interaction at a distance of 10 Å. The equilibrated system was simulated for 1ps with a 500 kcal/mol/Å^2. Restraint on the protein backbone was adjusted under 1 atm constant pressure and 310 K constant temperature. The Langevin damping coefficient was set to 5 ps unless otherwise stated.

Structure validation: The stereochemical quality of the obtained structure was evaluated using PROCHECK program[34], WHATCHECK[35], WHATIF[36], PROSA[37], VERIFY 3D[38] and ERRAT program[39].This model was further subjected to identification of active site.

Active site identification: The binding sites of the protein were identified and assessed using CASTp (Computed Atlas of Surface Topology of Proteins)[40]. CASTp is a program for identifying and characterizing protein active sites, binding sites, functional residues located on protein surface and voids buried under the interior of proteins by measuring concave surface regions of three dimensional structures of proteins. It also measures the area and volume of the pocket or void by solvent accessible surface model (Richards surface and molecular surface model (Connolly surface). It also finds application in studies of surface features and functional regions of the protein.

Results & Discussion

A set of TrxR sequences was collected by using the TrxR protein sequence as a query in the CONSEQ server (BLAST *E*-value threshold: 10^{-2}, maximum number of homologs: 500, iteration: 1)[10]. Out of 199 PSI-BLAST hits, 170 of them were unique. The calculation was performed on the 50 sequences with the lowest E-value. A multiple sequence alignment was made using MUSCLE (v 3.51), with default parameters and used to construct an unrooted phylogeny using the tree building facility of CLUSTAL-W (Fig. 2a). The conservation scores versus residue number were determined and plotted as shown in Fig. 2b and 2c.

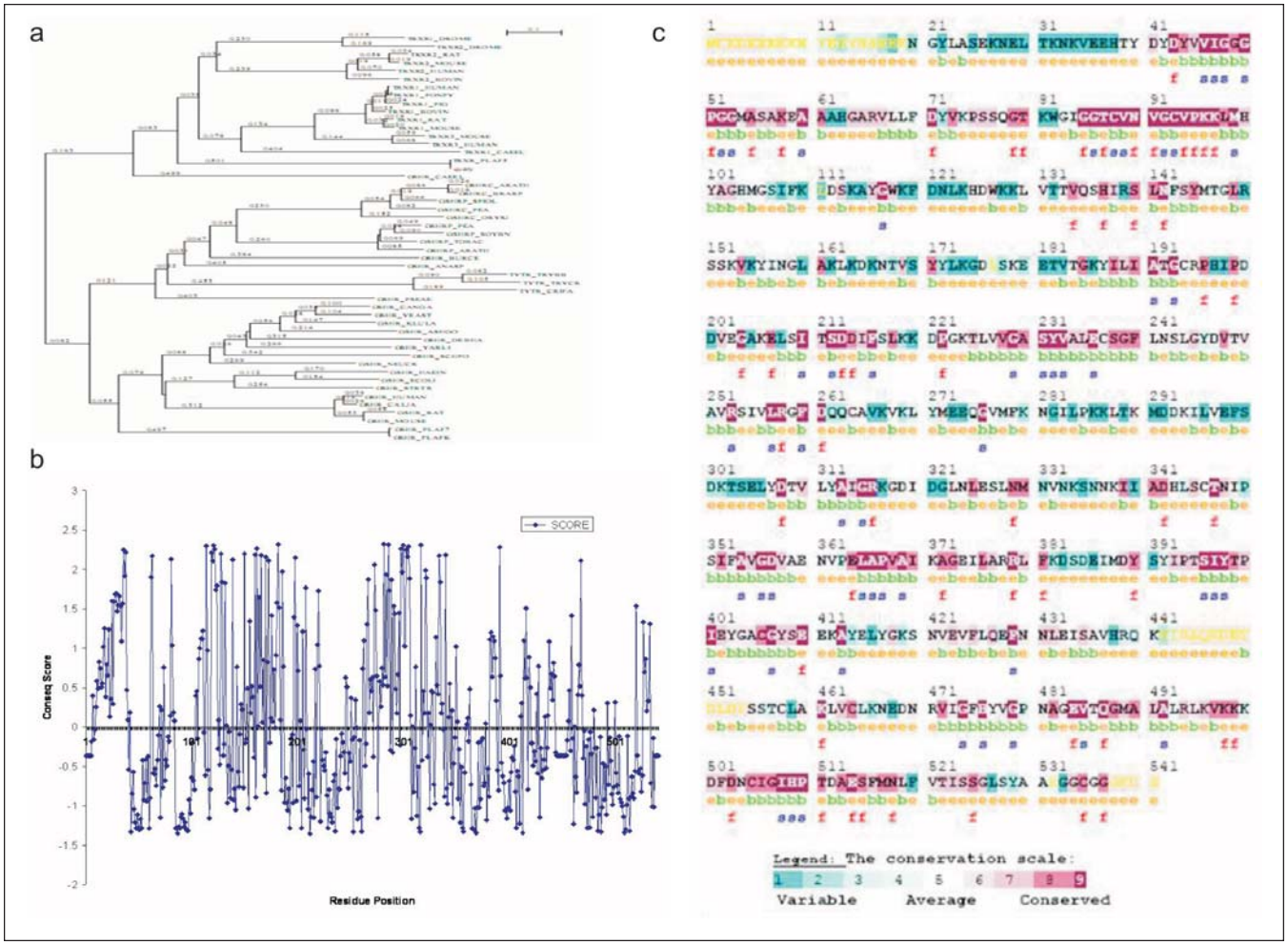

Fig. 2: (a) Phylogenetic tree obtained using CONSEQ; (b) Conservation score with residue number; and (c) Conservation scores of amino acids on a scale varying from 0–9 indicating variable to conserved amino acids where, e — An exposed residue according to the neural-network algorithm, b—A buried residue according to the neural-network algorithm, f — A predicted functional residue (highly conserved and exposed), s — A predicted structural residue (highly conserved and buried), X — Insufficient data–the calculation for this site was performed on <10% of the sequences

Table 1. Secondary structure elements as calculated by different tools

Secondary structure	DPM (%)	DSC (%)	GOR4 (%)	HNNC (%)	PHD (%)	Predator (%)	SIMPA96 (%)	SOPMA (%)	Sec.Cons. (%)
Alpha helix	21.81	25.51	31.24	32.35	26.43	20.33	36.23	29.02	27.91
3_{10} helix	0	0	0	0	0	0	0	0	0
Pi helix	0	0	0	0	0	0	0	0	0
Beta bridge	0	0	0	0	0	0	0	0	0
Extended strand	18.67	18.67	22	19.41	22.37	17.93	12.75	22.7	19.22
Beta turn	12.2	0	0	0	0	0	0	7.21	0
Bend region	0	0	0	0	0	0	0	0	0
Random coil	47.32	55.82	46.77	48.24	51.20	61.74	51.02	41.04	47.5
Ambigous states	0	0	0	0	0	0	0	0	5.36
Other states	0	0	0	0	0	0	0	0	0

Table 2. Intrinsic disorder of the protein as calculated by different tools

Server	Disordered	Disordered by REM465	Disordered by loops/coil definition	Disordered by HOTLOOPs definition
PRDOS	1-20,26-30,532-541			
RONN	1-34,438-443,532-541			
GLOBPROT		2-13,20-34,52-63, 198-215,362-370, 435-452, 532-541		
DISEMBL		1-10,54-61, 532-541	1-53,70-97,105-124,161-83, 191-227,275-291,316-330, 342-366,382-413,443-455, 463-485,499-514,527-541	1-28,76-88,160-182, 75-290,445-455,528-541

Secondary structure prediction: The secondary structure indicates whether a given amino acid lies in a helix, strand or coil. The results from the different secondary structure prediction servers used in the analysis revealed that random coils dominated among secondary structure elements followed by alpha helix, extended strand and beta turns (Table 1).

Identification of intrinsic disorder in protein: The results from all the servers revealed the regions CKDKNEKKN (2–9) and KGGCGGGKCG (532–541) as common (Table 2). The intrinsic disorder pro-

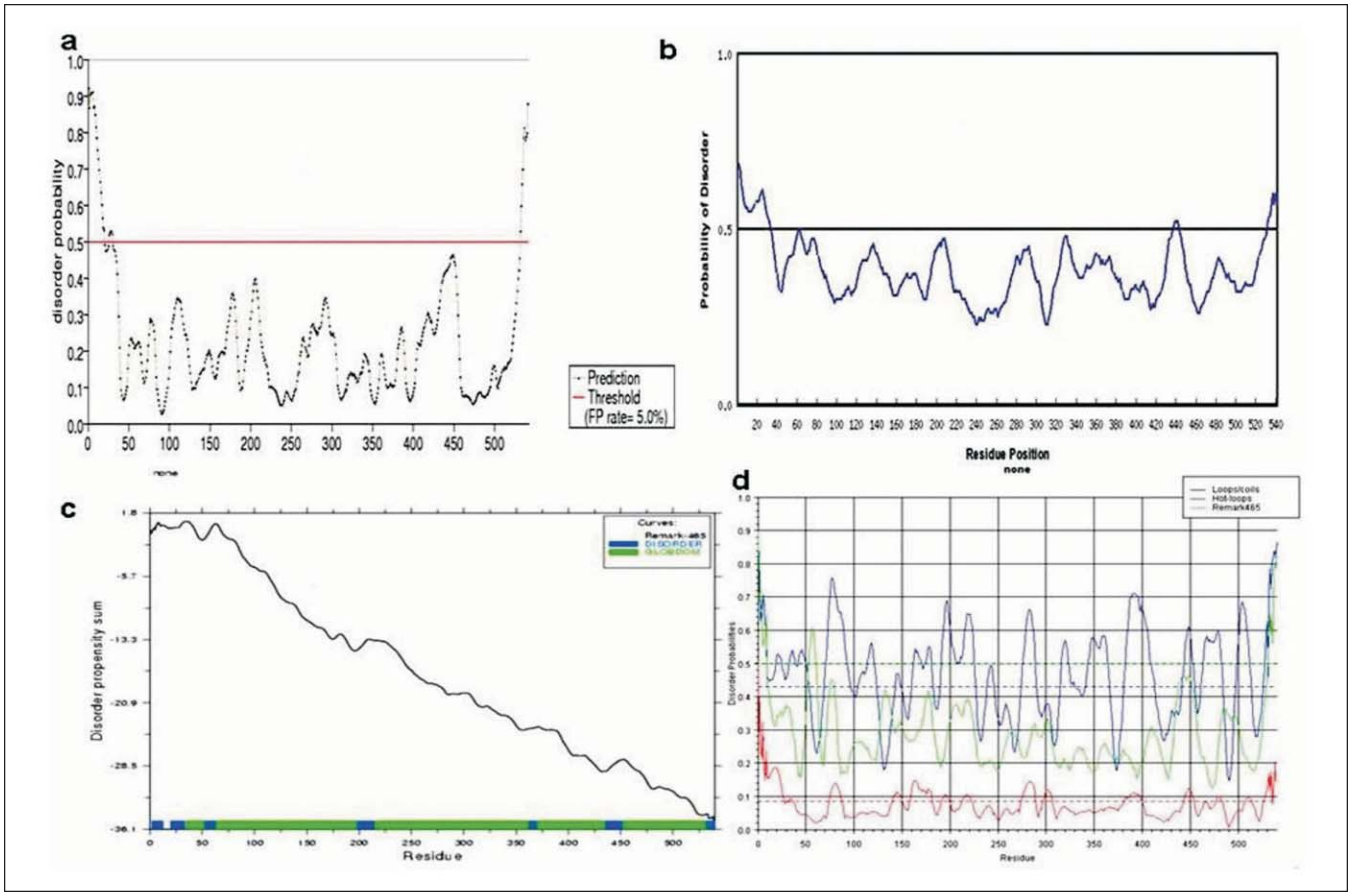

Fig. 3: Protein intrinsic disorder profile from different servers—(a) PRDOS; (b) RONN; (c) GLOBPROT; and (d) DISEMBL

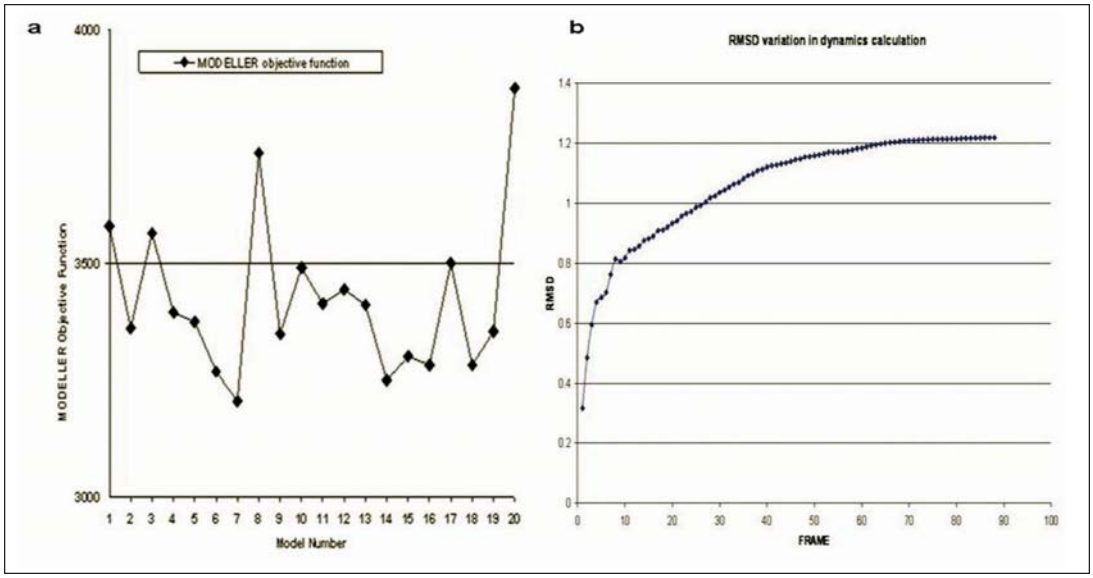

Fig. 4: (a) Generated models and their respective objective functionsl; and (b) RMSD variation in dynamics calculation

file of TrxR obtained using different servers is illustrated in Fig. 3.

Homology modeling: Homology modeling studies have been conducted for deriving the structure of other enzymes of *Plasmodium* species, for instance, apical membrane antigen-1[41], heat shock protein 90[42] Dihydrofolate reductase[43] and yielded satisfactory results. The modeling approach can be summarized as follows: from an initial alignment, it consists of four steps—(i) the derivation of sets of models by MODELLER9v3, (ii) selection of the best model on the basis of relative objective function values from the set of models generated, (iii) molecular dynamics calculation, and (iv) a final validation step.

Total 20 models were generated using MODELLER (Fig. 4a). Model with least objective function of 3204.0662 was selected for further analysis. The model obtained was refined by molecular dynamics method and the graph was drawn between the time (ps) and RMSD of C alpha which showed that model is stable above 50 ps of molecular dynamics simulations as shown in Fig. 4b. The refined model (Fig. 5a) was used for further analysis. The root-mean-square deviation (RMSD) between the backbone atoms of the template and the homology model was 1.29 Å, again indicating a close homology (Fig. 5b). Therefore, we felt that we had a reasonable and reliable conformation for further analysis.

Structure validation: The geometry of model was evaluated with Ramachandran's plot calculations using PROCHECK. Stereochemical evaluation of backbone Psi and Phi dihedral angles revealed that 78.3, 17.6, 3.2 and 0.9% of residues were falling within the most favoured regions, additionally allowed regions, generously allowed regions and disallowed regions of Ramachandran's plot respectively (Fig. 5c). Nevertheless, Ramachandran plot analysis showed that several residues Tyr232, Asn324, Asn361 were placed out of energetically favoured regions in the plot. Remaining residues are in the core regions of Ramachandran's plot, predicting the final structure to be highly reliable for further studies. Totally, 99.1% of the residues are in the most favoured and allowed regions. The high quality of the structure is further evident by the G-factor of –0.4 computed in PROCHECK and overall quality factor of 92.441 indicates acceptable protein environment. Similarly, WHATCHECK revealed that RMS Z-score for bond lengths, RMS Z-score for bond angles, omega angle restraints, side chain planarity, improper dihedral distribution, outside distribution are 0.656, 1.155, 1.507,

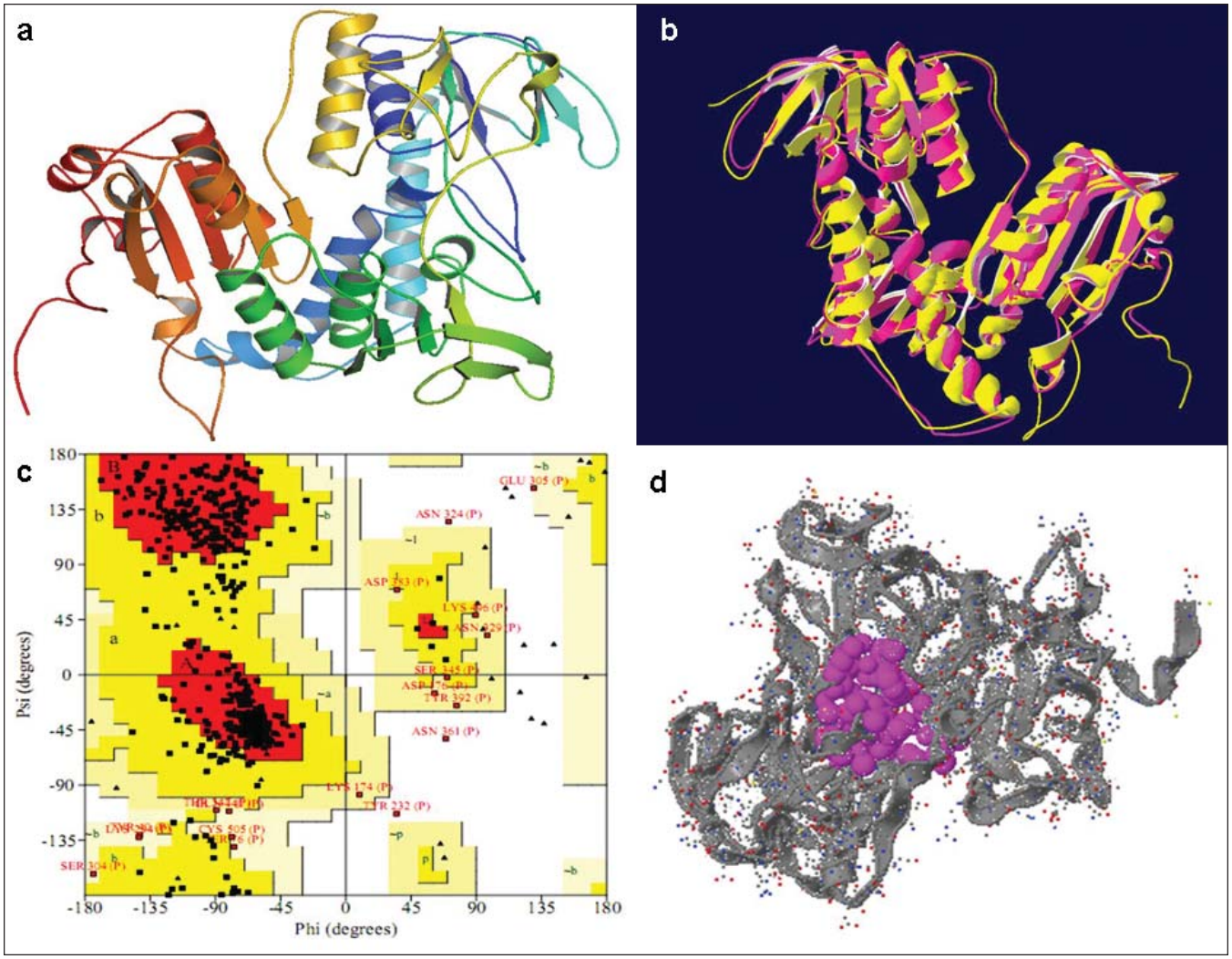

Fig. 5: (a) Model of *Plasmodium falciparum* thioredoxin reductase; (b) Superimposition of target (Yellow) and template (Rose) structure; (c) Ramachandran plot analysis of the modeled structure. The stereo chemical spatial arrangement of amino acid residues are shown in Ramachandran plot. The Plot statistics are: residues in most favoured regions [A, B, L] - 131 (84.5%); residues in additional allowed regions [a,b,l.p] - 21 (13.9%) (area represented in red); residues in generously allowed regions (area represented in yellow) [-a,-b,-l,-p] - 3 (1.9%); residues in disallowed regions–0 (area represented in white); number of non-glycine and non-proline residues-155(100%); number of end residues (excl. Gly and Pro)-2; number of glycine residues (shown as triangles)-13; number of proline residues-5; total number of residues- 175; and (d) Projection of the predicted active site

1.507, 1.340, 0.874 and 1.078, respectively for the modeled structure which are all positive values (Positive is better than average). The model has a normal distribution of residue types over the inside and the outside of the protein.

WHATIF program predicted the RMSD Z-Score of backbone-backbone contacts (–1.79), backbone-sidechain contacts (–1.14), sidechain-backbone contacts (–1.68), sidechain-sidechain contacts (–1.06) as shown in Fig. 6b. Moreover, the structural integrity of final model was reflected by a Z-Score of –1.84. These values suggest that structural average for second generation quality control values are within normal range of 2.0 and hence, confirm the reliability of modeled structure. Arg 425 and Tyr 430 were found to deviate from planarity by more than 4-times the expected value. Sixteen pairs of atoms showed abnormally short inter-atomic distances. There are also 28 residues with abnormal bond angles.

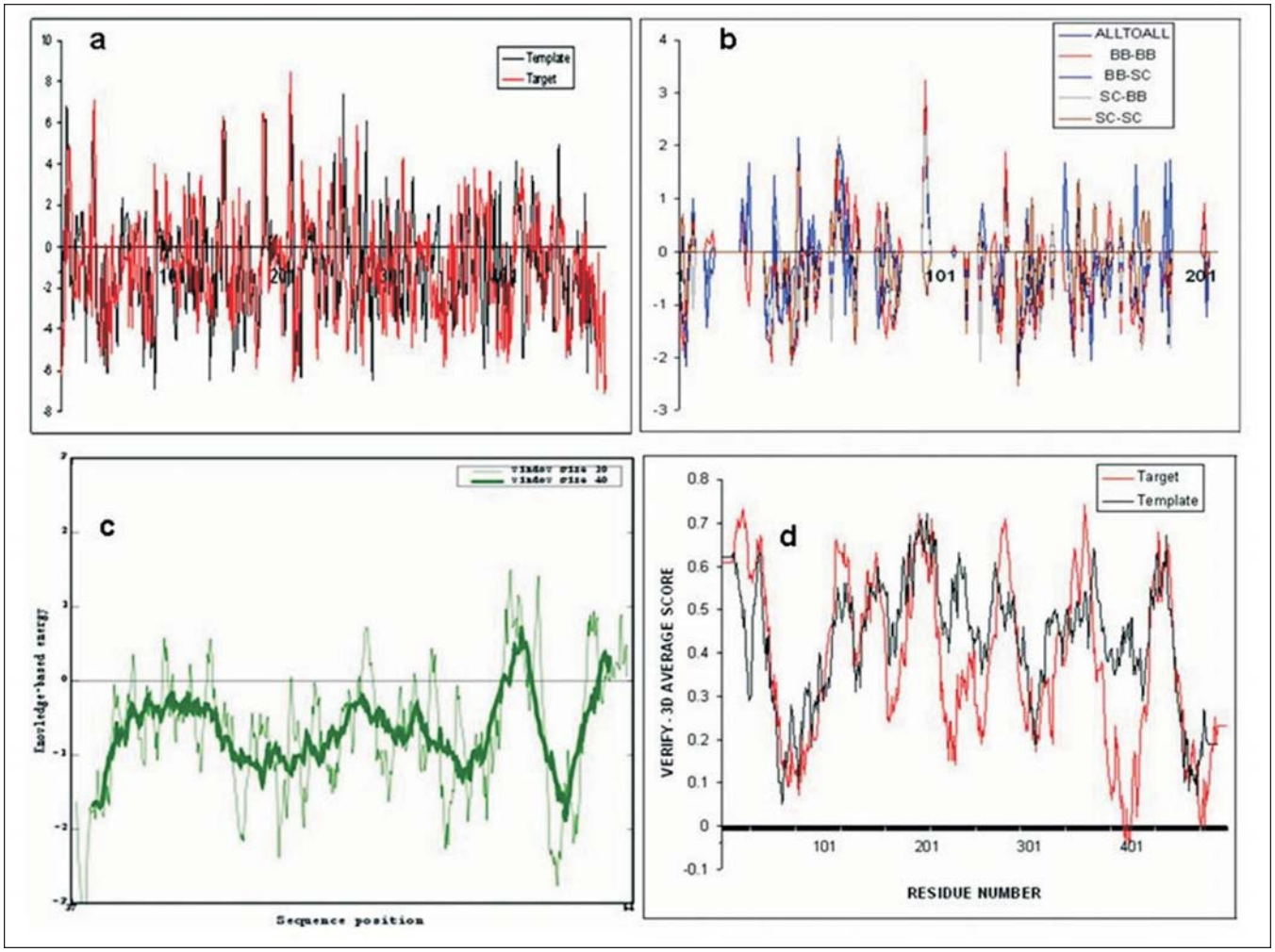

Fig. 6: (a) Packing quality assessment values calculated for the modeled protein; (b) Fine packing quality of the final model; (c) PROSA energy plot of the Thioredoxin reductase model. The graphs are smoothed over a window size of 50 residues; and (d) VERIFY3D validation of the target modeled protein

Another check is the packing quality of each residue as assessed by the WHATIF program. Fig. 6a represents the profile obtained with respect to the residues. As the evaluation criterion corresponds in this case to a threshold of –5, one can see that all residues show satisfactory packing values except for 27 residues for which the score is below the threshold. Therefore, the WHATIF evaluation also indicates that the homology modeled structure is very reasonable.

PROSA is widely used to check 3D models of protein structures for potential errors. The Z-score indicates overall model quality and measures the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. The PROSA[44] score was negative for the modeled protein, which indicates its correctness. Fig. 6c displays the PROSA profiles calculated for the protein model which were found similar to the energy of the template structures. Overall model quality of target as reflected with Z-score was –9.25 which was quite similar to that of template (–10.94). The PROSA curves represent the residue interaction energies; negative values correspond to stable parts of the molecules. VERIFY 3D analyzes the compatibility of an atomic model (3D) with its own amino acid sequence. For our protein homology model, 80.36 of the residues have a score of greater than 0.2 which indicates a good quality model (Fig. 6d).

ERRAT works by analyzing the statistics of non-bonded interactions between different atom types and a score of greater than 50 is normally acceptable. For our model, the ERRAT score is 92.441which is well within normal range for a high quality model (Fig. 7); the ERRAT score for the template is 88.723.

All these analyses suggested the reliability of the structure and this model was compared with the models obtained using online modeling servers, viz. CPH model[45], Geno3D[46], Magos[47] and Swiss Model[48] (Table 3).

Active site identification: Once the final model was built, the possible binding sites of TrxR were searched using the CASTp server[41]. Total 91 possible binding sites were obtained. The active site was predicted to have an area of 666.2 and volume of 791.2 (Fig. 5d). The residues forming the pocket were Pro51, Thr87, Val91, Gly92, Cys94, Lys96, Cys194, Ser212, Phe 216, Glu236, Val233, Tyr232, Ser231, Cys237, Ala313, Ile314, Gly315, Arg316, Gly356, Asp357, Pro363, Glu364, Leu365, Ala366, Pro367, Ala369, Pro394, Ser396, Ile397, Tyr398, Gly483 and Gln487.

Conclusion

Malaria is the leading cause of mortality attributable to a communicable disease, with an unprecedented death toll of above two million annually. More than 107 countries and territories are reported to be in the grip of malaria in 2004. Antimalarial drug resistance is recognized to be one of the greatest coercion to our ability to battle against malaria. The situation continues to be more frightening, with the geographical spread of resistance widening to previously unaffected areas and a ruthless augmentation both in the incidence and degree of drug resistance.

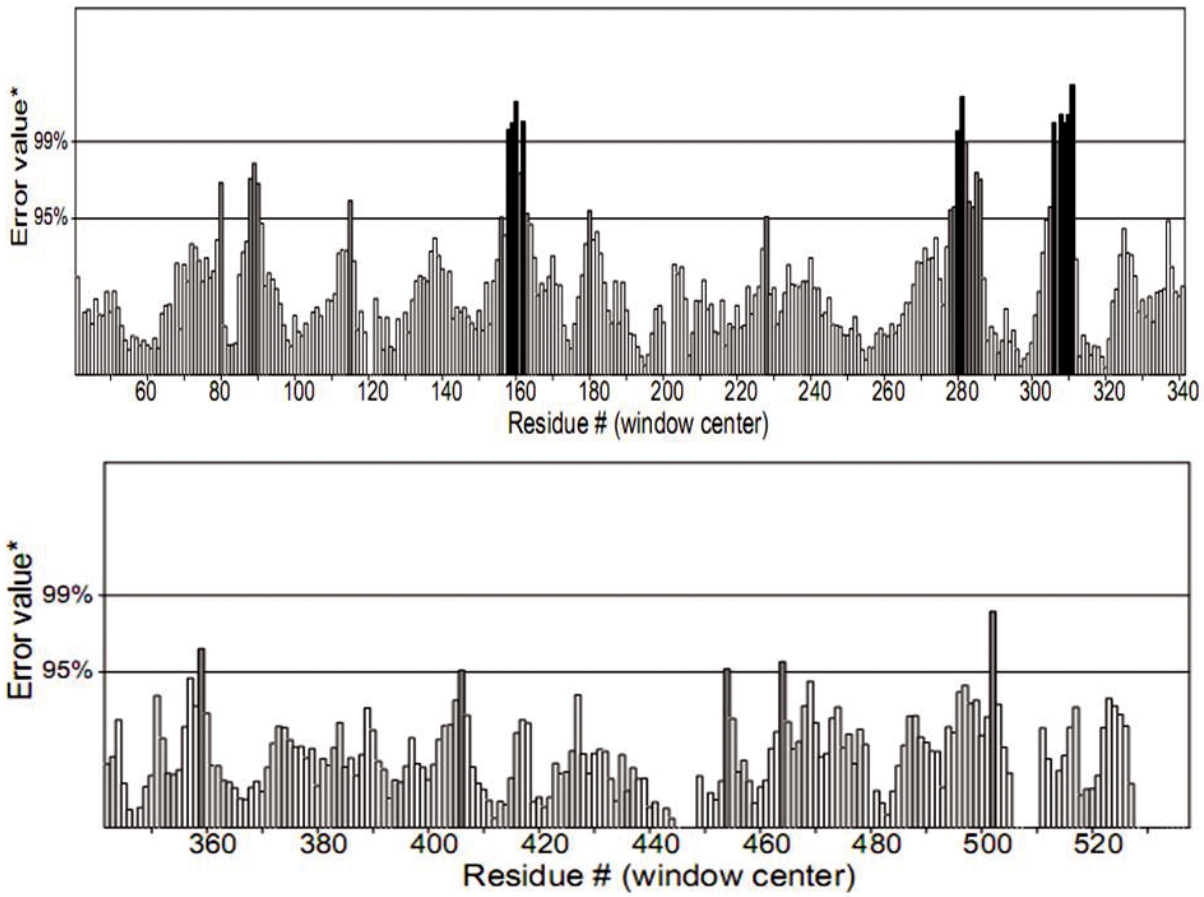

Fig. 7: 3D profiles of thioredoxin reductase using ERRAT server

Table 3. Comparative results of structure validation of different models obtained from MODELLER and other online servers

Server	Property	CPH model	Geno3D	Homer	Magos	SwissModel	Modeller
ERRAT	Overall quality factor	85.744	86.820	33.824	84.979	72.878	92.441
PROVE	Z-Score Mean	0.242	4.395	5.090	3.712	0.191	5.975
	Z-Score std deviation	1.387	60.514	71.972	55.644	1.627	72.802
	RMS Z-score	1.408	60.655	72.129	55.750	1.638	73.026
PROCHECK RC PLOT	Residues in most favoured regions	84.1%	67.4%	71.7%	67.3%	75.3%	78.3%
	Residues in additional allowed regions	13.1%	28%	23.1%	26.5%	22.6%	17.6%
	Residues in generously allowed regions	2.5%	2.6%	3.5%	3.9%	1.4%	3.2%
	Residues in disallowed regions	0.2%	2.1%	1.6%	2.3%	0.7%	0.9%
WHATCheck	RMS Z-score for bond lengths	0.482	0.249 (tight)	0.664 (tight)	0.251 (tight)	0.621 (tight)	0.656 (tight)
	RMS Z-score for bond angles	0.88	0.460 (tight)	0.800	0.493 (tight)	1.184	1.155
	Omega angle restraints	0.915	0.100 (tight)	0.202 (tight)	0.114 (tight)	1.039	1.507 (loose)
	Side chain planarity	2.373	0.389 (tight)	0.067 (tight)	0.449 (tight)	1.747	1.340
	Improper dihedral distribution	1.894 (loose)	0.378	0.852	0.424	1.554 (loose)	0.874
	Outside distribution	1.054	1.109	1.352 (unusual)	1.116	1.041	1.078
	2nd generation packing quality	–1.333	–3.136 (poor)	–3.928 (poor)	–3.036 (poor)	–2.011	–2.007
	Ramachandran plot appearance	–4.622	–4.583 (bad)	–5.698 (bad)	-4.258 (bad)	–4.655 (bad)	–3.556 (poor)
	Chi-1/Chi-2 rotamer normality	–0.773	–2.592	4.198	–2.508	–0.210	–2.801
	Backbone conformation	–8.608 (bad)	–8.194 (bad)	–10.294 (bad)	–8.860 (bad)	–7.943 (bad)	–6.269
VERIFY 3D	(3D-1D score > 0.2)	85.4	82.14%	34.02%	80.65%	87.6	80.57

The increasing drug resistance in *Plasmodium* persuades us to explore novel drugs to reduce the malaria burden. Comparative protein modeling is of great assistance during the rational design of drug molecules. In the dearth of experimental data, model-building on the basis of the known crystal structure of a homologous protein is the lone unswerving method to gain structural information. The enzymes of redox system of *Plasmodium* species are considered interesting potential drug targets as their inhibition affects several vulnerable points in redox mechanism required for dealing with oxidative challenge in host erythrocytes.

In this work, we have constructed a three-dimensional model of the TrxR of malarial parasite by homology modeling. After energy minimization and molecular dynamics simulations, the refined model structure was obtained. The final refined model was further assessed by various structure validation methods which con-

firmed the reliability of the model. In the absence of the experimental structure, this model will provide a foundation for elucidating structure function relationship and pave a way for rational drug design.

Acknowledgement

Authors are thankful to Dr J.S. Yadav, Director, Indian Institute of Chemical Technology, for his constant support and encouragement throughout the study. AKB thanks Council of Scientific and Industrial Research (CSIR) for Senior Research Fellowship (SRF). NA thanks Department of Science and Technology (DST) for Research Associate fellowship.

References

1. Breman JG, Alilio MS, Mills A. Conquering the intolerable burden of malaria: What's new, what's needed: a summary. *Am J Trop Med Hyg* 2004; *71*(2): 1–15.
2. *The world health report 2002: reducing risks, promoting healthy life*. Geneva : World Health Organization 2002.
3. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present and future. *Lancet Infect Dis* 2004; *4*(6): 327–36.
4. Clark IA, Hunt NH. Evidence for reactive oxygen intermediates causing hemolysis and parasite death in malaria. *Infect Immun* 1983; *39:*1–6.
5. Wozencraft AO. Damage to malaria infected erythrocytes following exposure to oxidant generating systems. *Parasitology* 1986; *92:* 559–67.
6. Vennerstrom JL, Eaton JW. Oxidants, oxidant drugs, and malaria. *J Med Chem* 1988; *31*(7): 1269–77.
7. Williams CH Jr, Zanetti G, Arscott LD, McAllister JK. Lipoamide dehydrogenase, glutathione reductase, thioredoxin reductase, and thioredoxin. *J Biol Chem* 1967; *242*(22): 5226–31.
8. Andricopulo AD, Akoachere MB, Krogh R, Nickel C, McLeish MJ, Kenyon GL, *et al.* Specific inhibitors of *Plasmodium falciparum* thioredoxin reductase as potential antimalarial agents. *Bioorg Med Chem Lett* 2006; *16:* 2283–92.
9. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel, RD, *et al.* Protein identification and analysis tools on the ExPASy server. In: John M. Walker, editor. *The Proteomics Protocols Handbook.* Humana Press 2005.
10. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, Casadio R, Ben-Tal N. ConSeq: the identification of functionally and structurally important residues in protein sequences**.** *Bioinformatics* 2004; *20:* 1322–4.
11. Deleage G, Roux B. An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng* 1987; *1*(4): 289–94.
12. King RD, Sternberg MJ. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci* 1996; *5*(11): 2298–310.
13. Garnier J, Gibrat JF, Robson B. GOR secondary structure prediction method version IV. *Meth Enzymol* 1996; *266:* 540–53.
14. Guermeur Y, Geourjon C, Gallinari P, Deleage G. Improved performance in protein secondary structure prediction by inhomogeneous score combination. *Bioinformatics* 1999; *15*(5): 413–21.
15. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol* 1993; *232*(2): 584–99.
16. Frishman D, Argos P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng* 1996; *9*(2): 133–42.
17. Levin JM, Robson B, Garnier J. An algorithm for secondary structure determination in proteins based on sequence similarity. *FEBS Lett* 1986; *205*(2): 303–8.
18. Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput Appl Biosci* 1995; *11*(6): 681–4.
19. Deleage G, Blanchet C, Geourjon C. Protein structure prediction: implications for the biologist. *Biochimie* 1997; *79*(11): 681–6.
20. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, Russell RB. Protein disorder prediction: implications for structural proteomics. *Structure* 2003; *11:* 1453–9.
21. Linding R, Russell RB, Neduva V, Gibson TJ. GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res* 2003; *31:* 3701–8.
22. Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* 2005; *21:* 3369–76.
23. Ishida T, Kinoshita K. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res* 2007; *35:* W460–4.
24. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.

A basic local alignment search tool. *J Mol Biol* 1990; *215:* 403–10.

25. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, *et al.* The Protein Data Bank. *Nucleic Acids Res* 2000; *28*(1): 235–42.
26. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997; *24:* 4876–82.
27. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 1993; *234*(3): 779–815.
28. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, *et al*. NAMD2: greater scalability for parallel molecular dynamics. *J Comput Phys* 1999; *151:* 283–312.
29. Schlick T. Computational molecular biophysics today: a confluence of methodological advances and complex biomolecular applications. *J Comput Phys* 1999; *151:* 1–8.
30. MacKerell AD, Bashford D, Bellott M, Dunbrack Jr, RL, Evanseck JD, Field MJ, *et al*. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 1998; *102:* 3586–616.
31. Schlenkrich M, Brickmann J, MacKerell Jr AD, Karplus M. Empirical potential energy function for phospholipids: criteria for parameter optimization and applications. In: Merz KM, Roux B, editors. *Biological membranes: a molecular perspective from computation and experiment.* Boston: MABirkhauser B. 1996; p. 31–81.
32. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983; *79:* 926–35.
33. Grubmüller H, Heller H, Windemuth A, Schulten K. Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol Simul* 1991; *6:* 121–42.
34. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993; *26:* 283–91.
35. Hooft RWW, Vriend G, Sander C. Abola EE. Errors in protein structures. *Nature* 1996; *381:* 272.
36. Vriend G. WHAT IF: a molecular modeling and drug design program. *J MolGraph* 1990; *8*(1): 52–6, 29.
37. Wiederstein M, Sippl M. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res* 2007; *35:* W407–10.
38. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol* 1997; *277:* 396–404.
39. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 1993; *2:*1511–9.
40. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J. CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* 2006; *34:* W116–8.
41. Mamatha DM, Nagalakshmamma K, Dev VA, Rajesh, Sheerin VS. Protein modeling of apical membrane antigen-1(AMA-1) of *Plasmodium cynomolgi*. *Afr J Biotechnol* 2007; *6* (22): 2628–32.
42. Kumar R, Pavithra SR, Tatu U. Three-dimensional structure of heat shock protein 90 from *Plasmodium falciparum*: molecular modelling approach to rational drug design against malaria. *J Biosci* 2007; *32*(3): 531–6.
43. Giulio Rastelli, Sara Pacchioni, Marco Daniele Parenti. Structure of *Plasmodium vivax* dihydrofolate reductase determined by homology modeling and molecular dynamics refinement. *Bioorganic Medicinal Chem Lett* 2003; *13:* 3257–60.
44. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins* 1993; *17*(4): 355–62.
45. Lund O, Nielsen M, Lundegaard C, Worning P. CPHmodels 2.0: X3M a Computer Program to Extract 3D Models Abstract at the CASP5 conferenceA102, 2002.
46. Combet C, Jambon M, Deléage G, Geourjon C Geno3D: Automatic comparative molecular modelling of protein. *Bioinformatics* 2002; *18:* 213–4.
47. Garnier N, Friedrich A, Bolze R, Bettler E, Moulinier L, Geourjon C, *et al.* MAGOS: multiple alignment and modelling server. *Bioinformatics* 2006; *22*(17): 2164–5.
48. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 2003; *31:* 3381–5.

Corresponding author: Dr U.S.N. Murty, Bioinformatics Group, Biology Division, Indian Institute of Chemical Technology, Hyderabad–500607, India.
E-mail: murty_usn@yahoo.com

Received: 8 December 2008 *Accepted in revised form:* 6 July 2009