Analyzing a Potential Drug Target N-Myristoyltransferase of *Plasmodium falciparum* Through *In Silico* Approaches

Amit Kumar Banerjee, Neelima Arora, USN Murty

Bioinformatics Group, Biology Division, Indian Institute of Chemical Technology, Tarnaka, Uppal Road, Hyderabad, Andhra Pradesh, India

ABSTRACT

Background: Despite concerted global efforts to combat malaria, malaria elimination is still a remote dream. Fast evolution rate of malarial parasite along with its ability to respond quickly to any drug resulting in partial or complete resistance has been a cause of concern among researcher communities. **Materials and Methods:** Molecular modeling approach was adopted to gain insight about the structure and various analyses were performed. Modeller 9v3, Protparam, Protscale, MEME, NAMD and other tools were employed for this study. PROCHECK and other tools were used for stereo-chemical quality evaluation. **Results and Conclusion:** It was observed during the course of study that this protein contains 32.2% of aliphatic amino acids among which Leucine (9.5%) is predominant. Theoretical pI of 8.39 identified the protein as basic in nature and most of the amino acids present in N-Myristoyltransferase are hydrophobic (46.1%). Secondary structure analysis shows predominance of alpha helices and random coils. Motif analyses revealed that this target protein contains 2 signature motifs, i.e., EVNFLCVHK and KFGEGDG. Apart from motif search, three-dimensional model was generated and validated and the stereo-chemical quality check confirmed that 97.7% amino acid residues fall in the core region of Ramachandran plot. Molecular dynamics simulation resulted in maximum 1.3 Å Root Mean Square Deviation (RMSD) between the initial structure and the trajectories obtained later on. The template and the target molecule has shown 1.5 Å RMSD for the C alpha trace. A docking study was also conducted with various ligand molecules among which specific benzofuran compounds turned out to be effective. This derived information will help in designing new inhibitor molecules for this target protein as well in better understanding the parasite protein.

Key words: Bioinformatics, Comparative modeling, Malaria, NMT, N-myristoyltransferase, *Plasmodium falciparum*

INTRODUCTION

Every year, millions of people pay the cost of residing in malaria-endemic regions of the world with their life. It is speculated that 40% of population face the risk of malaria in 109 countries across the globe.[1] Most vulnerable population figures among the most downtrodden and is further weakened by poverty, malnutrition, unhygienic conditions, and above all with limited or no access to healthcare and life-saving drugs. Malaria now claims more life than ever before and is tightening its grip in newer regions. In 2006, 247 million cases were reported and 3.3 million live under the shadow of malaria.[1] A glance at malaria figures shows clearly that it is the biggest killer of children below 5 years of age and the number is bound to increase in next few years according to a recent estimate.[1]

Malaria is caused by four species of protozoan parasite *Plasmodium vivax, Plasmodium falciparum, Plasmodium ovale,* and *Plasmodium malariae* transmitted by vector *Anopheles* mosquitoes. *Plasmodium falciparum* has emerged as one of the most successful and dreaded parasite of our times. Hopes raised by initial success of insecticide Dichloro Diphenyl Trichloroethane (DDT) have been marred by increasing resistance of vectors to insecticides and drug resistance in parasite. Emergence of multi–drug resistant parasite has been the cause of failure of our efforts to control malaria. Efficacy of many conventional anti-malarial drugs has been compromised.[2-3] Malaria control still remains a far fetched dream and big claims of malaria elimination have been a major letdown despite the enormous funding for research,[4] efforts by national as well as international agencies and advances in basic and applied sciences. Widespread resistance of *Plasmodium* to chloroquine and sulfadoxine/

Address for correspondence:
Dr. USN Murty, E-mail: murty_usn@yahoo.com

pyrimethamine poses a major threat to malaria control programs.[5] Failure of conventional drugs for treatment of malaria has warranted the need for exploring new drug targets. Malaria drug pipeline is nearly empty due to lack of interest from big pharmaceutical giants owing to low profit margins. Few anti-malarial drugs that were registered during past several decades can be counted on fingers.[6] The woefully small number of available anti-malarial drugs in arsenal for battling malaria is a cause of concern. In order to replenish the drug pipeline, efforts for identification of novel chemotherapeutic targets must be intensified.

Myristoyl-CoA protein *N*-myristoyltransferase (NMT: EC 2.1.3.97) is a ubiquitous cytosolic[7] enzyme, which follows a bi–bi catalytic reaction mechanism to bring about co-translational transfer of the rare cellular fatty acid myristate (C14:0) from myristoyl CoA to the N-terminal glycine residue in a variety of eukaryotic cellular protein substrates.[8-10] N-terminal N-myristoylation results in significant change in important properties like lipophilicity of the protein and thus facilitates interactions of protein with hydrophobic domains and membranes.[11-15] NMT of *P. falciparum* is known to express in its asexual blood stage and has been cloned.[16] NMT of *P. falciparum* is an attractive drug target as established by comparative biochemical studies involving human and parasite enzyme.[17] Difference in properties can be exploited for development of specific *P. falciparum* NMT inhibitors. So, the logical step for characterization of protein is making use of available bioinformatics tools and techniques for determining important aspects of the enzyme. Crystal structure of *P. falciparum* NMT has not been solved yet. Comparative modeling provides a way for obtaining structural information in absence of experimentally derived structure. We have undertaken this study to characterize *P. falciparum* NMT *in silico* as knowledge of key properties and structural features of NMT will aid in development of new chemotherapeutic agents for treatment of malaria.

MATERIALS AND METHODS

Amino acid sequence of *P. falciparum* NMT (Accession Number: AAF18461.1) was obtained from the protein database of NCBI. *In silico* characterization of physiochemical properties of the protein was performed using Protparam[18] and Protscale available at EXPASY(http://www.expasy.ch/). Secondary structure predictions were made using multi-prediction server at Network Protein Sequence Analysis at PBIL. CONSEQ was used for determining important residues conserved during the course of evolution.[19] Multiple Em for Motif Elicitation (MEME) available at (http://meme.nbcr.net/

meme4_1_1/cgi-bin/meme.cgi) was used to predict motifs in the *Pf* NMT keeping default options (minimum width: 6, Maximum width: 50, Motifs to find: 3 and minimum sites >=2). For identifying potential domains, PROSCAN was used. Quasi motifinder was employed to predict signatures and motif like patterns keeping default cut-off *P* value for motif similarity of 0.05 and number of possible pseudomotifs predicted was kept 3.

CYSREC (http://linux1.softberry.com/) was used to predict the cysteine pairing pattern. As some proteins are known to contain several unstructured or disordered regions, metaProtein disorder prediction system (metaPrDOS)[20] was used for prediction of disorder regions applying PrDOS, DISOPRED2, DisEMBL, DISPROT (VSL2P), DISpro and IUPred predictors. Transmembrane regions were predicted by DAS,[21] TMPRED.[22] SOSUI,[23] HMMTOP,[24] TMHMM,[25] and SPLIT.[26] A search was made using target protein sequence as query against PDB database by Position Specific Iterative (PSI) BLAST[27] keeping default parameters so as to find homologous structure, which can serve as template for theoretical model construction. Sequence alignment between target and template sequence was created using CLUSTALX,[28] a multiple sequence alignment program.

Homology model was constructed using a computer program MODELLER 9v3,[29] which relies on satisfaction of spatial restraints. MODELLER works by accepting target–template alignment as an input and employs a series of model building steps for construction of the model. Obtained model was further refined and subjected to molecular dynamics simulations using strategy followed in our earlier work.[30,31]

Structural diagram of the model was prepared and viewed using VMD.[32] The model was subjected to various tests for assessing the quality. For checking the consistency as well as validity of the model, stereochemical evaluation was performed using PROCHECK,[33] VERIFY3D,[34] and ERRAT[35] server. As low folding energy indicates stability of the model, energetic properties were also determined using PROSA.[36] POLYVIEW[37] was used to predict secondary structure profile and solvent accessible area.

Molecular dynamics simulation was performed on the molecule showing lowest MODELLER objective function. These studies employed NAMD 2.5 (Nano-scale Molecular Dynamics)[38] by applying CHARMM27 force field[39] for lipids and proteins[40,41] along with the TIP3P model for water. Energy minimization of the structure was done using 10,000 steps. Multiple time-stepping algorithm was used[42,43]

with an integration time step of 2 fs. Various interactions were computed in 1, 2, and 4 time steps for covalent bonds, short-range non- bonded interactions and long range electrostatic forces, respectively. For every ten time steps, non-bonded interactions were with a pair list distance of 13.5 Å. Van-der-Waals and electrostatics interactions were defined as interactions between short-range non-bonded interactions between particles within 12 Å. For Van-der-Waals interactions at a distance of 10 Å, a smoothing function was employed. Simulations were performed on the equilibrated system for 1 ps while maintaining a restraint of 500 kcal/mol/A° 2 on the protein backbone under constant pressure and temperature of 1 atm and 310 K, respectively, with Langevin damping coefficient set to 5 ps. Structure showing least energy with converged root mean square deviation was used for subsequent exercise. Final structure was visualized using VMD.[32]

Binding pockets were determined and explored using CASTp (Computed Atlas of Surface topology of Proteins)[44] and residues lining the functional site were also identified. The results were compared with highest scoring amino acids (score=9) predicted by Conseq results so as to identify the amino acids conserved in other species. Based on such observation, atom 3739 of Phenylalanine 226 was selected as target atom for docking. A total of 41 ligand molecules were selected based on literature[45-47] for docking studies to determine binding affinities of the ligand molecules toward the modeled NMT.

AMPAC software (http://www.semichem.com/ampac/default.php) was used for molecular mechanical calculation where AM1 calculations with SCF were performed applying restricted Hartree-Fork method. GOLD 2.0 package, (Cambridge Crystallographic Data Centre, Cambridge, UK)[48] which performs exhaustive and exclusive search for different conformations efficiently maintaining the flexibility of the ligand molecule provided by variation in dihedral angles was employed for docking studies. Default annealing parameters were considered for Van der Waals force and hydrogen bond calculation. Parameters considered for genetic algorithm (GA) are as follows: Population size 100, Selection pressure 1.1, No. of operations: 100000, No. of Islands: 5, Niche Size: 2, Migrate: 10, Mutate: 95, and Crossover: 95. Active site radius was set to 15 Å for docking calculations. Molecular interactions between the ligands and the protein were analyzed using SILVER.

RESULTS AND DISCUSSION

Sequence and structural analysis of the *P. falciparum* NMT was carried out using bioinformatics tools. Amino acid

composition of a protein reveals a lot about its nature. Amino acid composition was calculated using Protparam[18] from Expasy (http://us.expasy.org/tools/protpar/) and the obtained results are summarized in Table 1.

It was found that majority of amino acids present in NMT are hydrophobic (46.1%). Polar amino acids and charged amino acids constitute 23.2% and 27.5%, respectively, while glycine constitutes only 3.2% of all amino acids. Figure 1a shows the detailed representation of amino acid composition of NMT.

Other properties calculated using Protparam are summarized in Table 2.

Instability index of NMT indicates about the instability of the protein as a value above 40 is hallmark of unstable proteins. The aliphatic index (AI), which denotes the

Table 1: Amino acid composition of *P. falciparum* N-myristoyltransferase

Amino acid	Number	Percentage
Ala (A)	18	4.40
Arg (R)	16	3.90
Asn (N)	37	9.00
Asp (D)	30	7.30
Cys (C)	8	2.00
Gln (Q)	8	2.00
Glu (E)	21	5.10
Gly (G)	13	3.20
His (H)	7	1.70
Ile (I)	37	9.00
Leu (L)	39	9.50
Lys (K)	39	9.50
Met (M)	6	1.50
Phe (F)	21	5.10
Pro (P)	12	2.90
Ser (S)	25	6.10
Thr (T)	17	4.10
Trp (W)	7	1.70
Tyr (Y)	24	5.90
Val (V)	25	6.10

Table 2: Properties determined using Protparam

Property	Value
Number of amino acids	410
Molecular weight	47970
Theoretical pI	8.39
Total number of negatively charged residues	51
Total number of positively charged residues	55
Ext. coefficient (assuming ALL Cys residues appear as half cystines)	74760
Ext. coefficient (assuming NO Cys residues appear as half cystines)	74260
Estimated half-life (mammalian reticulocytes, *in vitro*)	30 hrs
Instability index	41.22
Aliphatic index	94.37
Grand average of hydropathicity (GRAVY)	0.327

relative volume of a protein occupied by aliphatic side chains was found to be 94.37. GRAVY value of 0.327 tells about its hydrated state. Theoretical pI of 8.39 classifies the protein as basic. Other important physico-chemical properties were calculated using Protscale.[18] Protscale assigns value to each amino acid using a scale and the results are presented below [Table 3].

Numerous NMT sequences were retrieved by the CONSEQ server[19] using *P. falciparum* NMT protein sequence as the query keeping the default options of BLAST *E*-value threshold: 0.001, maximum number of homologs: 50, iteration: 1. 32 out of 33 PSI-BLAST hits were found to be unique and the calculation was performed on the unique hits. The conservation scores versus residue number were determined and are shown in Figure 2a. An unrooted phylogenetic tree was constructed using the tree building facility of CLUSTAL-W employing the multiple sequence alignment obtained from MUSCLE [Figure 2b].

It was found that among the secondary structure elements, alpha helices were found to be predominant followed by random coils, extended strands and beta turns in descending order as predicted by DPM, DSC, GOR3, HNNC, SOPMA, Sec. Consensus. Results of three servers viz. MLRC, PHD, Predator indicated that random coils outnumbered alpha helices and extended strands in the protein [Table 4].

MEME was used for the elucidation of motifs in *P. falciparum* NMT with the parameters set to their default values. Three motifs predicted using MEME server along with their positions are shown in Table 5.

5 motifs were predicted using PROSCAN and 2 signature motifs viz Myristoyl-CoA: Protein N-myristoyltransferase signature 1(EVNFLCVHK) and Myristoyl-CoA: Protein N-myristoyltransferase signature 2(KFGEGDG) were found. Pattern, probability and description of the motifs
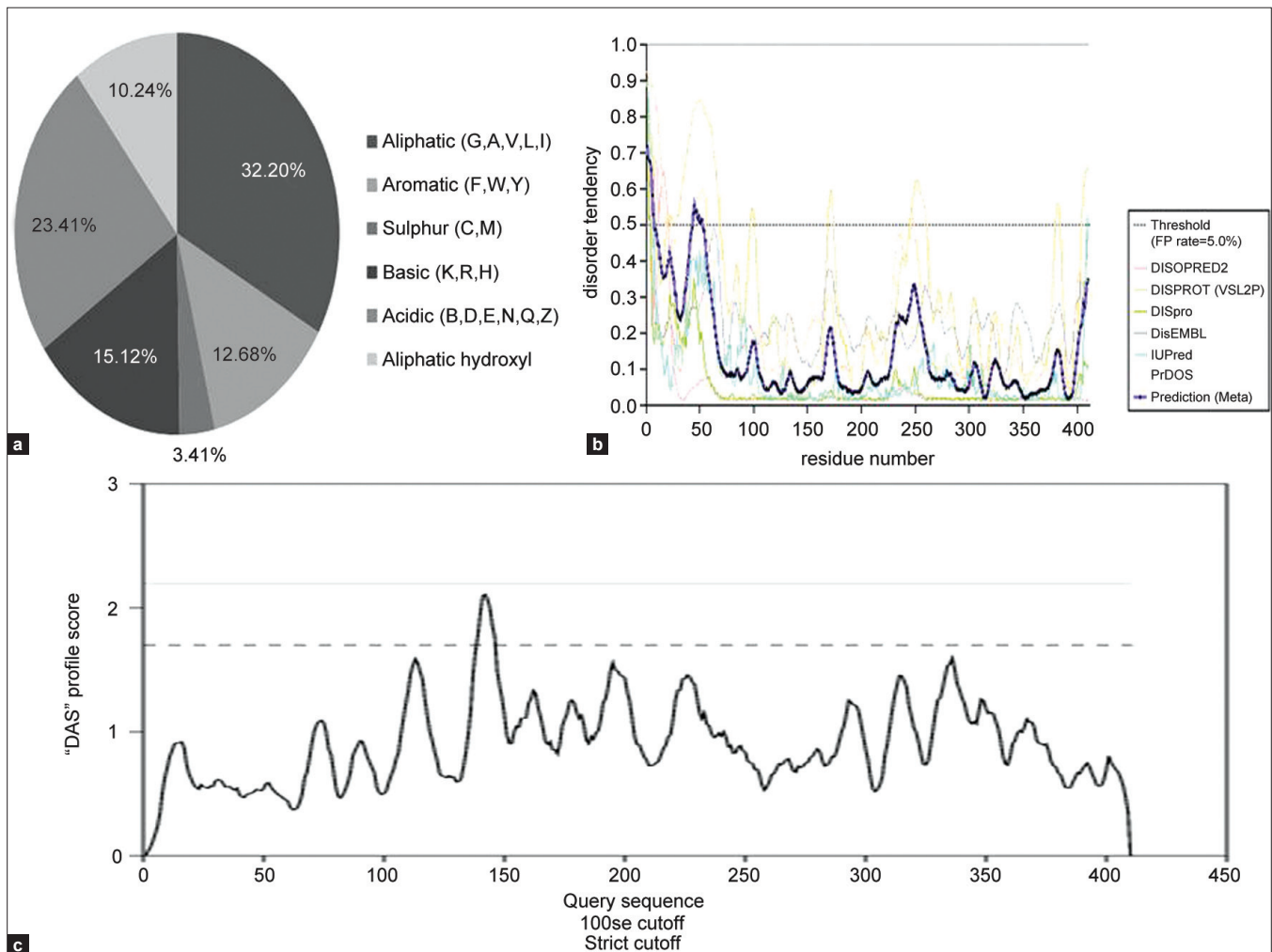

Figure 1: (a) Pie chart diagram representation of composition of *P. falciparum* N-Myristoyltransferase; (b) Protein disorder predicted using metaProtein disorder prediction system. 2 peaks clearly visible above threshold value of 0.5 denote the disordered regions; (c) Transmembrane region predicted using DAS server. The peak above 1.7 indicates the transmembrane region

are presented [Table 6]. 3 pseudomotifs were also predicted using Quasimotifinder apart from those predicted using PROSCAN [Table 7].

Disuphide bridges known as "switches for protein function",[49] result from covalent bonding of suphur from cysteine residues. Disulphide bridges play an important part in folding of protein and are also responsible for

Table 3: Important physicochemical properties calculated using protscale

Property	Minimum	Maximum	Average
Bulkiness	11.118	18.376	14.747
Polarity (Zimmerman)	0.458	39.127	19.7925
Recognition factors	82.778	95.222	89
Hydrophobicity (Kyte and Doolittle)	-2.6	2.167	-0.2165
% accessible residues	4.067	7.478	5.7725
% Buried residue	2.789	8.956	5.8725
Ratio hetero end/side	0.182	1.13	0.656
Average flexibility	0.381	0.489	0.435
Relative mutability	49.667	103.667	76.667
Refractivity	11.676	27.503	19.5895
Transmembrane tendency	-2.174	0.834	-0.67
Average area buried	104.811	157.8	131.3055

Table 4: Secondary structure prediction using NPS server

Server	Alpha helix (%)	Extended strand (%)	Beta turn (%)	Random coil (%)
DPM	36.59	25.61	8.29	29.51
DSC	48.05	7.32	0.00	44.63
GOR3	50.98	19.27	0.00	29.76
HNNC	41.22	18.54	0.00	40.24
MLRC	34.88	17.56	0.00	47.56
PHD	30.98	20.00	0.00	49.02
Predator	28.78	14.15	0.00	57.07
SOPM	36.83	24.88	7.32	30.98
Sec.cons	42.93	15.61	0.00	37.56

NPS: Network protein sequence, DPM: Double prediction method, DSC: Discrimination of protein Secondary structure Class, GOR: GARNIER OSGUTHORPE and ROBSON (Third improvement), HNNC: Hierarchical neural network, MLRC: Multivariate linear regression combination, PHD: PHD-an automatic mail server for protein secondary structure prediction, SOPM: Self-optimized prediction method
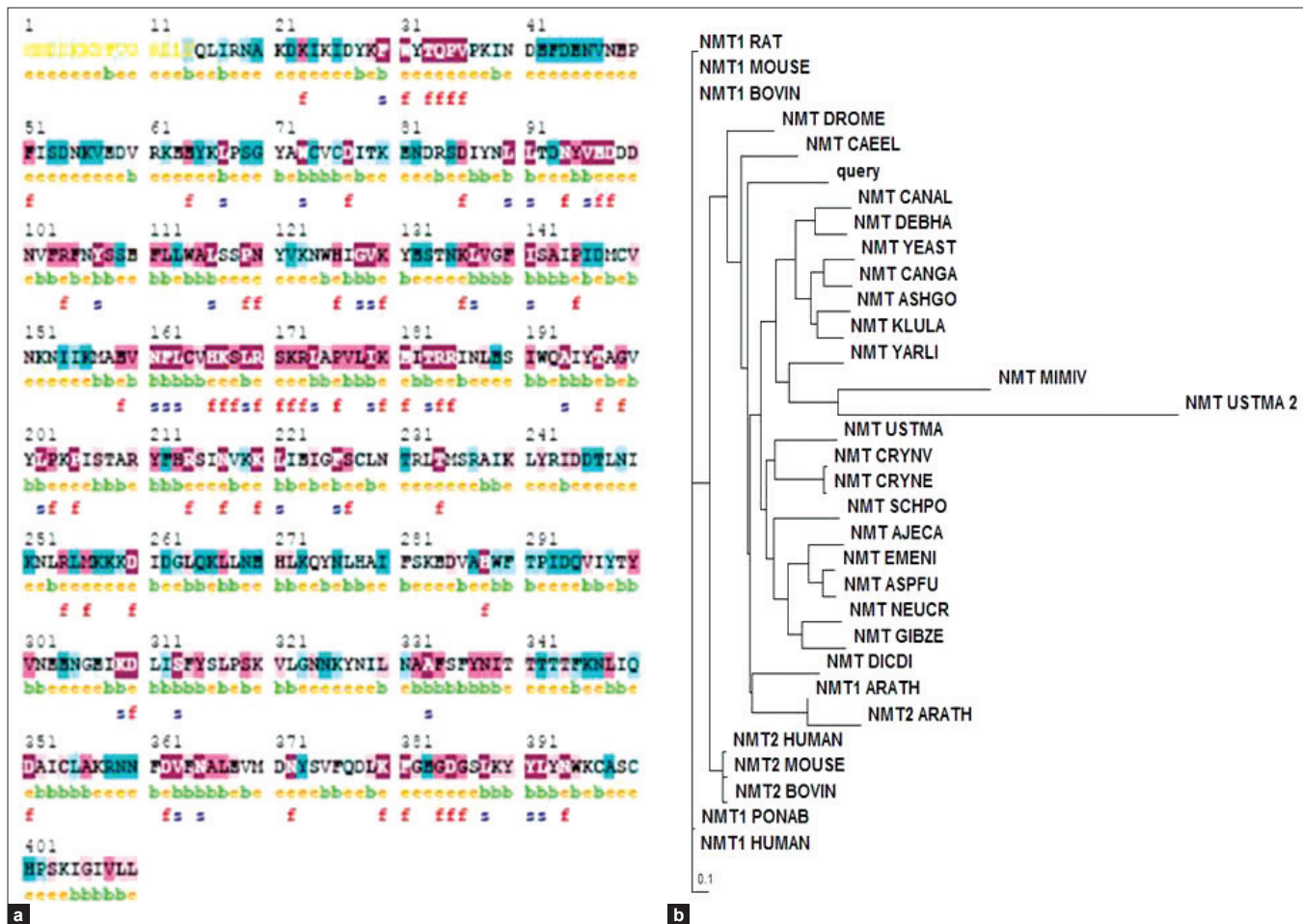

Figure 2: (a) Conservation scores of amino acids on a scale varying from 0–9 indicating variable to conserved amino acids where e-An exposed residue according to the neural-network algorithm, b-A buried residue according to the neural-network algorithm, f-A predicted functional residue (highly conserved 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 (b) Phylogenetic tree obtained using CONSEQ

stabilization of protein structure. Keeping this in mind, disulphide bridges were calculated using CYS-REC. The search resulted in eight cysteine residues and the most probable pattern of pairs are 74−400 and 76−397 [Table 8 in supplementary].

Some regions in protein occur as dynamic and unstructured ensembles and are called disordered regions. Identification of protein disorder is important for understanding protein function.[50] The disorder in protein facilitates its molecular interaction with multiple partners and is implicated in provision for various modification sites.[51] Disorder prediction by metaProtein disorder prediction system showed the presence of two such regions viz. MNDDKKDFVGRD (1-12) and DEFDENVNEPFISDN (41-55) with a value higher than the threshold of 0.5 [Figure 1b].

DAS[21] predicted a transmembrane region spanning from 139−146 at a cutoff value of 1.7 [Figure 1 c], while TMPRED predicted three helices (inside to outside

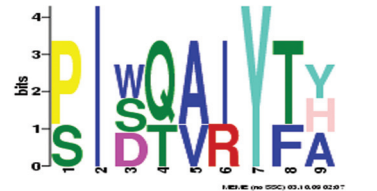## Table 5: Detailed information of motifs predicted using Meme

Motif	Relative entropy	Width	Sites	llr	E-value	Sequence logo	Start	*P* value
Motif 1	29.3	7	2	41	3.0e+000		69	1.13e-10
							30	4.47e-09
Motif 2	31.8	9	3	66	2.5e+001		291	8.72e-11
							189	1.46e-10
							204	4.38e-09
Motif 3	23.8	6	2	33	1.2e+002		394	2.02e-09
							113	6.48e-07

Table 6: Output of PROSCAN along with the probability and patterns of motifs predicted

Description	Prosite access number	Pattern	Randomized probability	Site
N-glycosylation site	PS00001	N-{P}-[ST]-{P}	5.138e-03	Site: 106 to 109 NYSS Site: 338 to 341 NITT Site: 372 to 375 NYSV
Protein kinase C phosphorylation site	PS00005	[ST]-x-[RK].	1.423e-02	Site: 134 to 136 TNK Site: 168 to 170 SLR Site: 171 to 173 SKR Site: 183 to 185 TRR Site: 208 to 210 TAR Site: 344 to 346 TFK Site: 387 to 389 SLK
Casein kinase II phosphorylation site	PS00006	[ST]-x(2)-[DE]	1.482e-02	Site: 282 to 285 SKED Site: 291 to 294 TPID
Myristoyl-CoA:protein N-myristoyltransferase signature 1	PS00975	[DEK]-[IV]-N-[FS]-L-C-x-H-K	2.191e-10	Site: 159 to 167 EVNFLCVHK
Myristoyl-CoA:protein N-myristoyltransferase signature 2	PS00976	K-F-G-x-G-D-G	4.316e-08	Site: 380 to 386 KFGEGDG

Table 7: Pseudomotifs predicted using QuasiMotif

Description	Prosite code	Pattern	Conservation score	Physiochemical score	Position	Pseudomotif
Kringle domain signature	PS00021	[FY]-C-[RH]-[NS]-x(7,1)-[WY]-C-.	-1.08675	1.12492	95-107	YVEDDDNVFRFNY
Kringle domain signature	PS00772	V-[DN]-Y-[EQD]-F-V-[DN]-C-.	-0.681647	-1.06868	160-167	VNFLCVHK
Histone H3 signature 2	PS00959	P-F-x-[RA]-L-[VA]-[KRQ]-[DEG]-[IV]-.	-0.872559	-1.0624	161-169	NFLCVHKSL

102(102)-122(120); 186(188)-209 (206); 329(329)-349(349) and two outside to inside helices (191(193)-209(209; 329(329)-345(345)). Results from SOSUI[23] server classified the protein as a soluble protein. HMMTOP, TMHMM, and SPLIT server did not predict any region spanning the membrane in the protein.

Protein structure often reflects its function. This fact

Table 8: Details of patterns of Cystine–Cystine binding

Position	Status	Score
CYS 74	Probably not SS-bounded	-2
CYS 76	Probably not SS-bounded	-0.6
CYS 149	Not SS-bounded	-43.8
CYS 164	Not SS-bounded	-44
CYS 228	Not SS-bounded	-25.4
CYS 354	Not SS-bounded	-25.7
CYS 397	Probably not SS-bounded	-14.2
CYS 400	Probably not SS-bounded	-5.9

makes the protein structure prediction a lucrative exercise in wake of unavailability of experimentally derived protein structures. Comparative modeling involves assigning the structure to a protein for which the structure has not been determined based on its sequence similarity to already known protein structures. This is based on the assumption that similarity at primary structure level is often indicative of structural similarity. This method assures reasonably good protein structure prediction which can provide deep insight in the mechanism of protein function in absence of its crystal structure.

Chain A of myristoyl-CoA: Protein N-myristoyltransferase (PDB ID: 1RXT_A) determined by X-ray diffraction at a resolution of 3.0 Angstrom was selected as a template as the protein shared 44.3% identity with the target protein sequence. Sequence alignment between target and template was generated using CLUSTALX and used as input for model construction [Figure 3a].
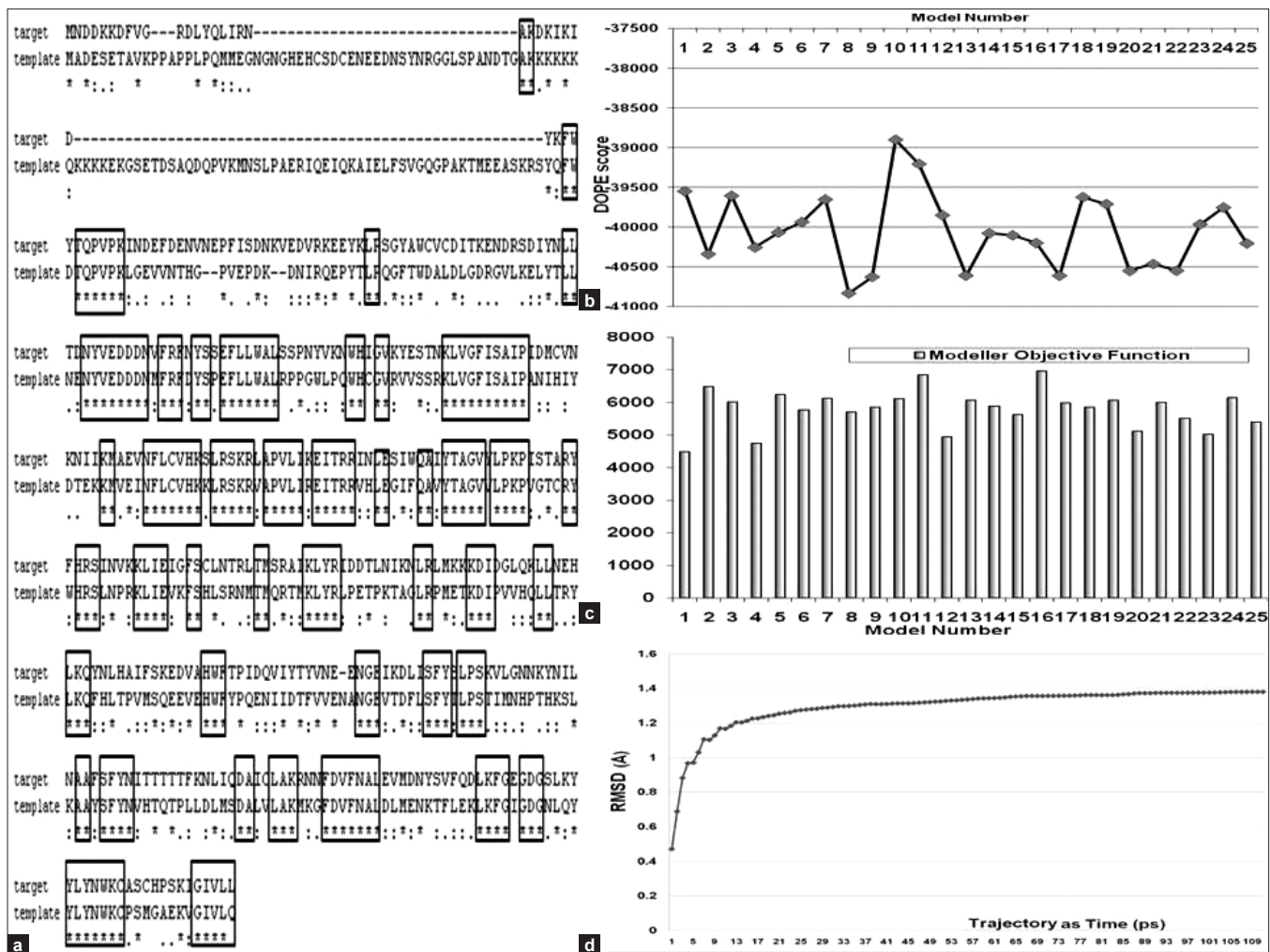

Figure 3: (a) Sequence alignment of template and target protein. Blocked regions represent the conservation between the target and template; (b) DOPE score of models generated using Modeller; (c) Modeller objective function of models generated using Modeller; (d) RMSD variations in dynamics calculation

A three-dimensional model was constructed using comparative modeling employing MODELLER9v3 based on probability density functions (PDFs). Total 25 models were generated using Modeller and evaluated based on DOPE score and Modeller objective function [Figure 3b-c]. Model having least objective function, i.e., 4486.498 was selected for further refinement. Model was refined by NAMD simulations using strategy discussed elsewhere.[30,31]

Obtained model was further refined using molecular dynamics. A graph was plotted between the trajectories generated as function of time (ps) and RMSD of C alpha trace of the protein molecule. During the MD simulation run, a total of 109 frames were generated and it was observed clearly that the obtained RMSD data showed a variation from 0.47 to 1.30 and thereafter, attained a plateau state with minor variation far away from the decimal points. Model was found to be stable above 1.3 ps of molecular

dynamics simulation [Figure 3d]. A schematic presentation of the modeled protein is shown in Figure 4a.

Superimposition of target and template indicates the structural similarities and differences and a close homology between template and target is expected as revealed by RMSD of 1.5 [Figure 4b]. This further reinforced the reliability of our model and this model was used for subsequent analysis. The protein belongs to alpha class of protein.

Stereochemical properties of the model were evaluated using PROCHECK. Ramachandran plot analysis of Psi and Phi dihedral angles showed that 97.7% fall in the allowed regions of the plot while 2.3% of the residues fall in disallowed regions [Table 9 and Figure 4c]. These results are comparable to the template structure. Several residues Asn 19, Gln 94, Lys25, Asp22, Tyr65, Tyr28, Ser33, Phe345, Phe51, Leu16, Lys23, Thr134, Tyr393, Phe226, and Ala198 fall outside
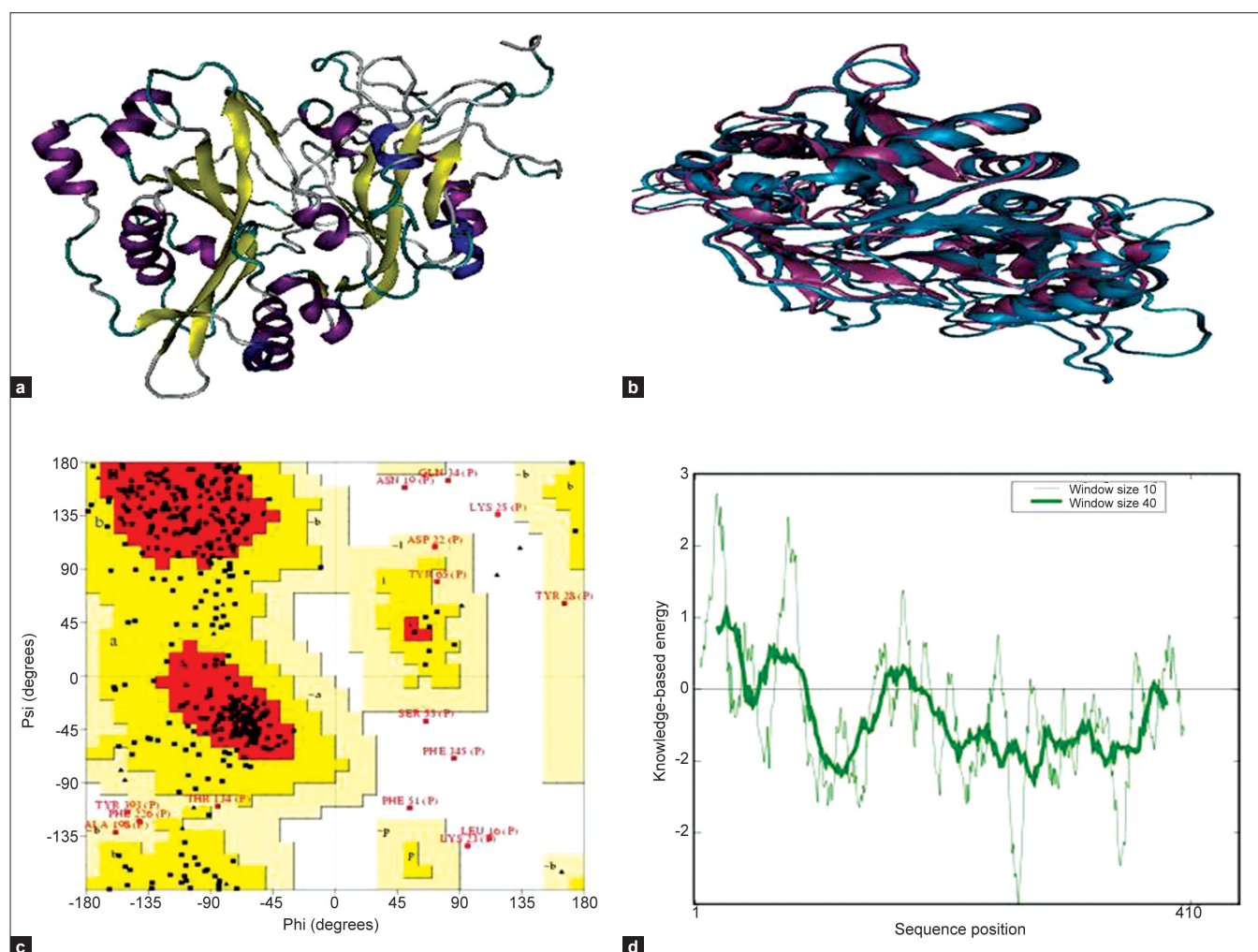

Figure 4: (a) Three-dimensional structure of *P. falciparum* NMT where alpha helix are shown in purple, 3_10_helix in blue, Pi_helix in red, extended beta sheets in yellow, bridge beta in tan, turn in cyan and coil in white color (b) Superimposition of target and template structure (Target in cyan color, template in mauve color) (c) Ramachandran plot analysis (d) PROSA curves representing the residue interaction energies

Table 9: Stereochemical evaluation of obtained protein model in Ramachandran Plot analysis using Procheck

Region of plots	Target		Template	
	Number	Percentage	Number	Percentage
Residues in most favoured regions [A, B, L]	263	68.4	184	60.3
Residues in additional allowed regions [a, b, l, p]	106	27.7	109	35.7
Residues in generously allowed regions [~a, ~b, ~l, ~p]	6	1.6	12	3.9
Residues in disallowed regions [XX]	9	2.3	0	0
Number of non glycine & non proline residues	383	100	305	100
Number of end-residues (excluding Gly and Pro)	2		1	
Glycine residues	13		17	
Proline residues	12		19	
Total number of residues	410		342	

energetically favorable regions. The ERRAT score of 76.294 (which is greater than 60.542 score of template) confirms about the reliability of the structure as a score of greater than 50 indicates the high quality of the model (data not shown). About 77.41% of residues in the model have a score >0.2 as revealed by VERIFY3D. This also underlines the reliability of structure (data not shown). PROSA score is a measure of model quality and calculates the deviation of the total energy of the structure with respect to an energy distribution derived from random conformations. PROSA score of model protein is -6.55, which is negative and is comparable to -9.36 [Figure 4d]. As negative value of residue interaction energy is the criterion for the correctness of the model, this further indicates good quality of model.

Secondary structure of a protein is considered as the local spatial arrangement of main chain atoms. Obtained model
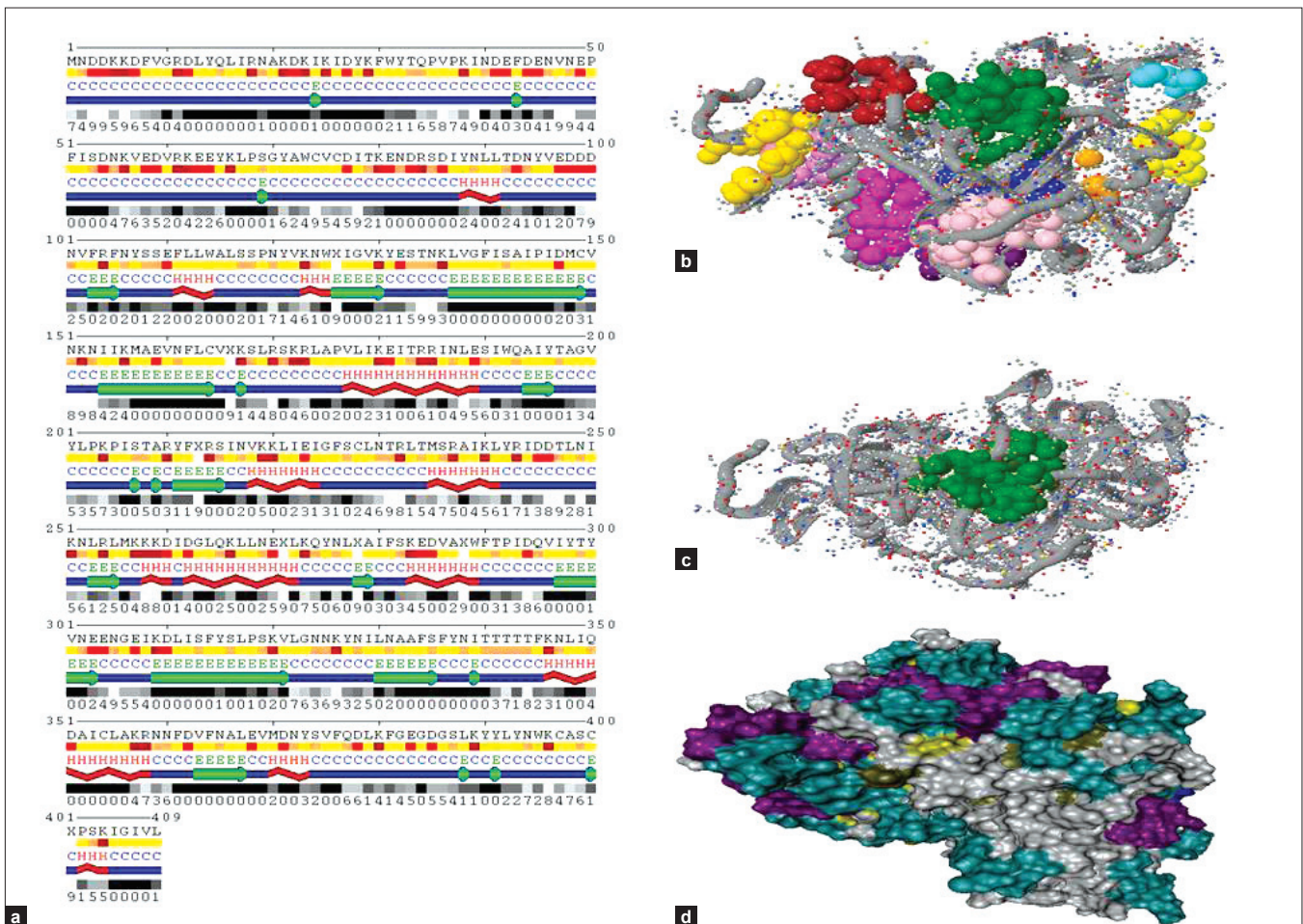

Figure 5: (a)Secondary structure, physiochemical profile and solvent accessible surface area as predicted by POLYVIEW ⌇ H-α and other helices (view 1), ▬ H-α and other helices (view 2), ▬▶ E-β-strand or bridge, ▬ C–coil, Relative solvent accessibility (RSA) ▪▪▪▫▫ 0-completely buried (0-9% RSA), 9-fully exposed (90-100% RSA), HAPNC where H-hydrophobic: A,C,F,G,I,L,M,P,V; A-amphipathic: H,W,Y; P-polar: N,Q,S,T and N/C-charged: D,E–negative, R,K–positive; (b)Top 10 possible binding sites as predicted by CASTP where Pocket 1=green, Pocket 2=blue, Pocket 3=cyan, Pocket 4=yellow, Pocket 5=magenta, Pocket 6=pink, Pocket 7=orange, Pocket 8=purple, Pocket 9=brown, Pocket 10=gold. (c) Active site of modeled protein. (d) Surface occupied by Pf NMT in spacefill model

was analyzed for determining the secondary structure and physiochemical profiles, solvent accessible surface area using POLYVIEW.[37] This provided the information on presence of secondary structure elements at specific locations [Figure 5a]. It was found that the protein contains 11 α helices, 46 turns, and 20 β strands altogether. A saddle shaped mixed β sheet was found in the core, which was further surrounded by several other α helices.

The local environment inside binding pockets and functional groups of amino acid lining the cavities on protein determines the function of protein and influence substrate binding. Active site of the modeled structure was determined using CASTP. Total 94 binding sites were predicted. Top 10 binding sites based on area are shown [Figure 5b]. Binding site having area of 754.9 and volume of 979 was further explored [Figure 5c]. Active site was found to consist of L16, I 17, N19, A20, K21, W73, D83, R84, Y95, V96, E97, D98, D99, N101, F103, F105, Y107, K167, T197, A198, G199, V200, Y211, U213, F226, L316, S318, L330, A332, F334,V363, N365, L367, D385, G386, S387, L388, Y390 and L410 [Figure 5c]. Volume and surface of the protein model are represented in Figure 5d.

Based on GOLDSCORE, which is dependent on the protein-ligand hydrogen bond energy (external H-bond), protein-ligand van der Waals (VDW) energy (external VDW), ligand internal VDW energy (internal vdw), and ligand torsional strain energy (internal torsion), the ligand protein-binding efficiency was computed. Molecule 7 in our study, which is a benzofuran compound with CH2CH2Ph and CH3 side chains[45], showed highest fitness score (65.93). It showed the presence of two hydrogen bonds, i.e., N-37 of ligand molecule with Serine 387(Bond length=2.422) and O12 of the ligand molecule with Tyrosine 211 (Bond length=2.154). Besides, it showed close contacts with C19 of ligand molecule with Glutamic acid 97 (Bond length=(2.689) and H52 of ligand molecule with Leucine 410 (Bond length=1.846)) [Figure 6].

Molecule 9 in the study,[45] belongs to the similar compound group with CH2SPh and CH3 side chains showed the second best fitness score of 65.74 in best rank file and displayed several close contacts which did not come up to the range of hydrogen bond. Following close contacts were observed: S53 of the ligand molecule with ALA198 of protein (Bond length=2.529), C43 of the ligand molecule with TYR95 of protein (Bond length=2.477), C41 of the ligand molecule with TYR95 of protein (Bond length=2.589), C5 of the ligand molecule with GLU97 of protein (Bond length=2.588), C34 of the ligand molecule with SER319 of protein (Bond length=2.586), N37 of

the ligand molecule with SER319O of protein (Bond length=2.451) [Figure 7a]

Molecule 14 in the study (A benzofuran compund with a Phenyl R group)[45] displayed score of 63.67 and showed the formation of 2 hydrogen bonds; one between O24 of the ligand molecule with Tyrosine 95 (Bond length=2.644) and another between O13 of the ligand molecule with TYR211 (Bond length=2.192) and 4 close contacts [Figure 7b]. Ranks of the rest of the molecules based on the GOLDSCORE fitness values are represented in Table 10.

CONCLUSION

Despite advent of high throughput methods, there exists a huge gap in number of available protein sequences and experimentally derived protein structures. Comparative
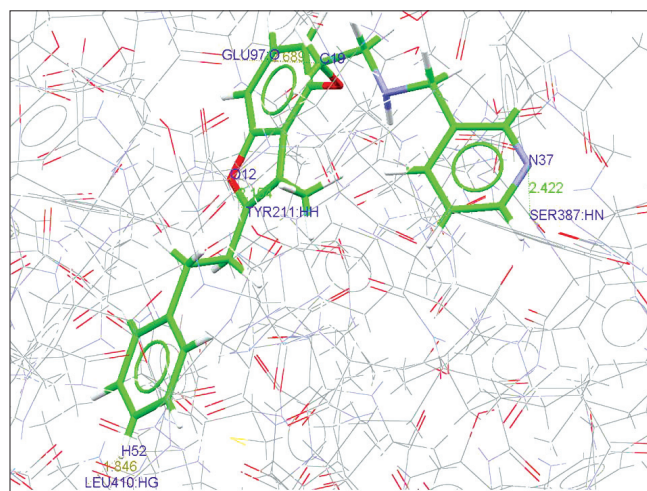

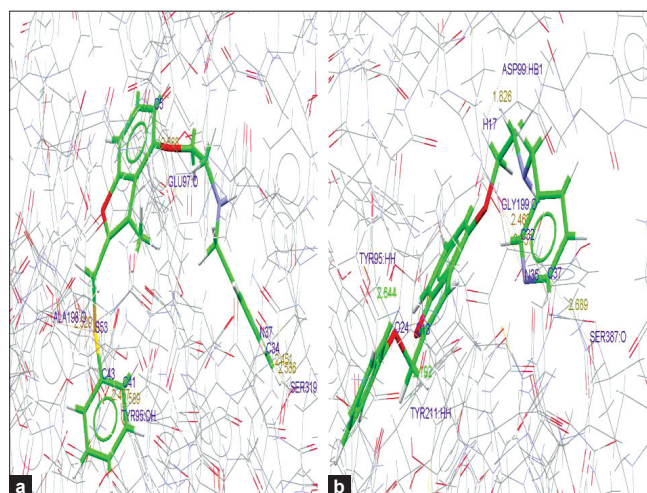**Figure 6:** Top scoring benzofuran compound obtained in the docking study

Figure 7: Second (a) and third (b) best benzofuran molecule found in the docking calculations

Table 10: Fitness scores of ligand molecules considered for the docking studies

Fitness	Ligand name	Reference
43.41	Mol1	[45]
60.98	Mol10	[45]
60.51	Mol11	[45]
59.1	Mol12	[45]
60.8	Mol13	[45]
63.67	Mol14	[45]
60.93	Mol15	[45]
62.18	Mol16	[45]
62.7	Mol17	[45]
59.45	Mol18	[45]
56.73	Mol19	[45]
48.31	Mol2	[45]
58.8	Mol20	[45]
61.68	Mol21	[45]
48.92	Mol22	[45]
60.98	Mol23	[45]
60.87	Mol24	[45]
52.71	Mol25	[45]
46.67	Mol26	[45]
62.53	Mol27	[45]
56.88	Mol28	[45]
58.2	Mol29	[45]
46.86	Mol3	[45]
55.56	Mol30	[45]
42.11	Mol31	[45]
42.77	Mol32	[45]
44.24	Mol33	[45]
45.39	Mol35	[45]
39.13	Mol36	[45]
43.3	Mol37	[46]
40.8	Mol38	[46]
43.38	Mol39	[46]
46.56	Mol4	[46]
53.57	Mol40	[46]
49.07	Mol41	[46]
44.48	Mol42	[46]
58.74	Mol5	[46]
53.51	Mol6	[47]
65.93	Mol7	[47]
57.98	Mol8	[47]
63.74	Mol9	[47]

modeling is often envisioned to play a role in providing details about protein structure in absence of crystal structures. We have applied bioinformatics tools for determining important features and properties of the NMT of *Plasmodium falciparum*.

The extent of reliability of structure prediction is dependent totally on the degree of sequence similarity between the target and template sequences. The methodology adopted by us resulted in a good quality model as the model was constructed based on a sequence homology of 44.3%. Model obtained was assessed using various structure validation servers and was found to be reasonably good. We have performed a thorough *in silico* characterization of

N-myristoyltransferase of malarial parasite. The outcome of this study will enhance our knowledge about this enzyme and will boost drug designing process as various aspects of structural characteristics of the enzyme explored in this study can provide a basis for effective inhibitor design. The model so obtained can be used as a potential target and results of this study can be exploited for the development of effective treatment plans for combating this dreaded disease.

REFERENCES

1. WHO report projections of mortality and the burden of disease. 2006.
2. Bloland PB, Lackritz EM, Kazembe PN, Were JB, Steketee R, Campbell CC. Beyond chloroquine-implications of drug resistance for evaluating malaria therapy efficacy and treatment policy in Africa. J Infect Dis 1993;167:932-7.
3. Wellems TE. *Plasmodium chloroquine* resistance and the search for a replacement antimalarial drug. Science 2002; 298:124-46.
4. Snow RW, Guerra CA, Mutheu J, Hay SI. International funding for malaria control in relation to populations at risk of stable *Plasmodium falciparum* transmission. PLoS Medicine 2008; 7:e142.
5. Olliaro PL, Taylor WR. Antimalarial compounds: From bench to bedside. J Exp Biol 2003; 206:3753-9.
6. Trouiller P, Olliaro P, Torreele E, Orbinski J, Laing R, Ford N. Drug development for neglected diseases: A deficient market and a public-health policy failure. Lancet 2002; 359:2188-94.
7. Devadas B, Freeman SK, Zupec ME, Lu HF, Nagarajan SR, Kishore NS, *et al*. Design and synthesis of novel imidazole-substituted dipeptide amides as potent and selective inhibitors of Candida albicans myristoylCoA: Protein N-myristoyltransferase and identification of related tripeptide inhibitors with mechanism-based antifungal activity. J Med Chem 1997;40:2609-25.
8. Bhatnagar RS, Schall OF, Jackson-Machelski E, Sikorski JA, Devadas B, Gokel GW, *et al*. Titration calorimetric analysis of AcylCoA recognition by myristoylCoA: Protein N-myristoyltransferase. Biochemistry 1997;36:6700-8.
9. Rudnick DA, McWherter CA, Rocque WJ, Lennon PJ, Getman DP, Gordon JI. Kinetic and structural evidence for a sequential ordered Bi Bi mechanism of catalysis by Saccharomyces cerevisiae myristoyl-CoA: Protein N-myristoyltransferase. J Biol Chem 1991;266:9732-9.
10. Farazi TA, Waksman G, Gordon JI. Structures of Saccharomyces cerevisiae N-myristoyltransferase with bound myristoylCoA and peptide provide insights about substrate recognition and catalysis. Biochemistry 2001;40:6335-43.
11. Boutin JA. Myristoylation. Cell Signal 1997;9:15-35.
12. Farazi TA, Waksman G, Gordon JI. The biology and enzymology of protein N-myristoylation. J Biol Chem 2001;276:39501-4.
13. Gordon JI. Protein N-myristoylation: Simple questions, unexpected answers. Clin Res 1990;38:517-28.
14. Han KK, Martinage A. Post-translational chemical modification(s) of proteins. Int J Biochem 1992;24:19-28.
15. Johnson DR, Bhatnagar RS, Knoll LJ, Gordon JI. Genetic and biochemical studies of protein N-myristoylation. Annu Rev Biochem 1994;63:869-914.
16. Gunaratne RS, Sajid M, Ling IT, Tripathi R, Pachebat JA, Holder AA. Characterization of N-myristoyltransferase from *Plasmodium falciparum*. Biochem J 2000;348:459-63.

17. Price HP, Menon MR, Panethymitaki C, Goulding D, McKean PG, Smith DF. Myristoyl-CoA: Protein N-myristoyltransferase, an essential enzyme and potential drug target in kinetoplastid parasites. J Biol Chem 2003;278:7206-14.

18. 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. United States: Humana Press; 2005.p. 571-607.

19. Berezin C, Glaser F, Rosenberg J, Paz I, Pupko T, Fariselli P, *et al.* ConSeq: The identification of functionally and structurally important residues in protein sequences. Bioinformatics 2004;20:1322-4.

20. Ishida T, Kinoshita K. Prediction of disordered regions in proteins based on the meta approach. Bioinformatics 2008;24:1344-8.

21. Cserzo M, Wallin E, Simon I, Von Heijne G, Elofsson A. Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: The dense alignment surface method. Prot Eng 1997;10:673-6.

22. Hofmann K, Stoffel W. TMbase-A database of membrane spanning proteins segments Biol. Chem. Hoppe-Seyler 1993;374:166.

23. Hirokawa T, Boon-Chieng S, Mitaku S. SOSUI: Classification and secondary structure prediction system for membrane proteins. Bioinformatics 1998;14:378-9.

24. Tusnády GE, Simon I. The HMMTOP transmembrane topology prediction server. Bioinformatics 2001;17:849-50.

25. Moller S, Croning MD, Apweiler R. Evaluation of methods for the prediction of membrane spanning regions. Bioinformatics 2001;17:646-53.

26. Juretic D, Zoranic L, Zucic D. Basic charge clusters and predictions of membrane protein topology. J Chem Inf Comput Sci 2002;42:620-32.

27. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. A basic local alignment search tool. J Mol Biol 1990;215:403-10.

28. 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.

29. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779-815.

30. Banerjee AK, Arora N, Murty US. Structural model of the *Plasmodium falciparum* Thioredoxin reductase: A novel target for antimalarial drugs. J Vector Borne Dis 2009;46:171-83.

31. Duddela S, Sekhar NP, Padmavati GV, Banerjee AK, Murty US. Probing the structure of human glucose transporter 2 and analysis of protein ligand interactions. Med Chem Res 2009;1-18.

32. Humphrey W, Dalke A, Schulten K. VMD-Visual Molecular Dynamics. J Mol Graph 1996;14:33-8.

33. 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.

34. Eisenberg D, Lüthy R, Bowie JU. VERIFY3D: Assessment of protein models with three-dimensional profiles. Methods Enzymol 1997;277:396-404.

35. Colovos C, Yeates TO. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci 1993;2:1511-9.

36. 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.

37. Porollo A, Adamczak R, Meller J. POLYVIEW: A flexible visualization tool for structural and functional annotations of proteins. Bioinformatics 2004;20:2460-2.

38. 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.

39. Schlick T, Skeel R, Brunger A, Kale L, Board JA Jr., Hermans J, *et al.* Algorithmic challenges in computational molecular biophysics. J Comput Phys 1999;151:9-48.

40. MacKerell A Jr., Bashford D, Bellott M, Dunbrack RL Jr., Evanseck J, Field MJ, *et al.* All-hydrogen empirical potential for molecular modeling and dynamics studies of proteins using the CHARMM22 force field. J Phys Chem B 1998;102:3586-616.

41. MacKerell AD, Brooks, CL, Nilsson L, Roux B, Won Y, Karplus M. CHARMM: The energy function and its parameterization with an overview of the program. Edited: Schleyer. The encyclopedia of computational chemistry. Chichester, UK: Wiley; 1998b. p. 271-7.

42. Jorgensen WL, Chandresekhar J, Madura DR, Impey W, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926-35.

43. Grubmuller 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.

44. 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.

45. Deokar HS, Purushottamachar P, Kulkarni VM. QSAR analysis of N-myristoyltransferase inhibitors: Antifungal activity of benzofurans. Med Chem Res 2009; 18:206-20.

46. Frearson JA, Brand S, McElroy SP, Cleghorn LA, Smid O, Stojanovski L, *et al.* N-myristoyltransferase inhibitors as new leads to treat sleeping sickness. Nature 2010;464:728-32.

47. Jennings BC, Nadolski MJ, Ling Y, Baker MB, Harrison ML, Deschenes RJ, Linder ME. 2-Bromopalmitate and 2-(2-hydroxy-5-nitro-benzylidene)-benzo[b]thiophen-3-one inhibit DHHC-mediated palmitoylation *in vitro.* J Lipid Res 2009;50:233-42.

48. Jones G, Willett P, Glen RC, Leach AR, Taylor RJ. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727-48.

49. Hogg PJ. Disulfide bonds as switches for protein function. Trends Biochem Sci 2003;28:210-4.

50. Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: Intrinsic disorder as an ID for recognition, regulation and cell signaling. J Mol Recognit 2005; 18:343-84.

51. Wright P, Dyson H. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J Mol Biol 1999;293:321-31.