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/
pyrimethamine poses a major threat to malaria control programs. 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. 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 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 of the protein and thus facilitates interactions of protein with hydrophobic domains and membranes. NMT of \textit{P. falciparum} is known to express in its asexual blood stage and has been cloned. NMT of \textit{P. falciparum} is an attractive drug target as established by comparative biochemical studies involving human and parasite enzyme. Difference in properties can be exploited for development of specific \textit{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 \textit{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 \textit{P. falciparum} NMT \textit{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 \textit{P. falciparum} NMT (Accession Number: AAF18461.1) was obtained from the protein database of NCBI. \textit{In silico} characterization of physiochemical properties of the protein was performed using ProtParam and Protoscale 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. 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 \textit{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 motiffinder was employed to predict signatures and motif like patterns keeping default cut-off \textit{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) was used for prediction of disorder regions applying PrDOS, DISOPRED2, DisEMBL, DISPROT (VSL2P), DISSpro and IUPred predictors. Transmembrane regions were predicted by DAS, TMPRED, SOSUI, HMMTOP, TMHMM, and SPLIT. A search was made using target protein sequence as query against PDB database by Position Specific Iterative (PSI) BLAST 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, a multiple sequence alignment program.

Homology model was constructed using a computer program MODELLER 9v3, 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.

Structural diagram of the model was prepared and viewed using VMD. 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,VERIFY3D, and ERRAT server. As low folding energy indicates stability of the model, energetic properties were also determined using PROSA. POLYVIEW 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) by applying CHARMM27 force field for lipids and proteins along with the TIP3P model for water. Energy minimization of the structure was done using 10,000 steps. Multiple time-stepping algorithm was used.
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 electrostatic 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/Å² 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 [19] from Expsy (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

| 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       |
| Glu (G)   | 8      | 2.00       |
| Gln (Q)   | 11     | 2.70       |
| 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       |

| 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) | 74/260 |
| Ext. coefficient (assuming NO Cys residues appear as half cystines) | 74/260 |
| 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. 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 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, S0PMA, 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].

Disulfide bridges known as “switches for protein function”,[49] result from covalent bonding of sulfur 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     | 0.702     | 0.554     |
| 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      | 0.566     |
| 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          | -2.124    | 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           | 25.62               | 8.29          | 29.51           |
| DSC        | 48.05           | 7.32                | 0.00          | 44.63           |
| GOR3       | 50.98           | 19.27               | 0.00          | 29.76           |
| HNNC       | 43.12           | 18.54               | 0.00          | 40.14           |
| 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   | 42.33           | 15.62               | 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](image-url) (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.\(^{[94]}\) The disorder in protein facilitates its molecular interaction with multiple partners and is implicated in provision for various modification sites.\(^{[83]}\) Disorder prediction by metaProtein disorder prediction system showed the presence of two such regions viz. MNDDKDVFGRD (1-12) and DEFCDNVEPFISDN (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 | lnr | E-value | Sequence logo | Start | P value |
|-------|-----------------|-------|-------|-----|---------|----------------|-------|---------|
| Motif 1 | 29.3 | 7 | 2 | 41 | 3.0e+000 | ![](image1.png) | 69 | 1.13e-10 |
| | | | | | | | 30 | 4.47e-09 |
| Motif 2 | 31.8 | 9 | 3 | 66 | 2.5e+001 | ![](image2.png) | 291 | 8.72e-11 |
| | | | | | | | 189 | 1.46e-10 |
| | | | | | | | 204 | 4.38e-09 |
| Motif 3 | 23.8 | 6 | 2 | 33 | 1.2e+002 | ![](image3.png) | 394 | 2.02e-09 |
| | | | | | | | 113 | 6.48e-07 |

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

| Description | Prosite code | Pattern | Randomized probability | Site |
|-------------|-------------|---------|------------------------|------|
| N-glycosylation site | PS00001 | N-[P]-[ST]-[P] | 5.38e-03 | Site: 106 to 109 NYSS Site: 338 to 344 NITT Site: 372 to 375 NYSV |
| Protein kinase C phosphorylation site | PS00005 | [ST] x [RK] | 1.43e-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: 378 to 389 SLK |
| Casein kinase II phosphorylation site | PS00006 | [ST] x (2) - [DE] | 1.48e-02 | Site: 282 to 285 SKL Site: 291 to 294 TPID |
| Myristoyl-CoA:protein N-myristoyltransferase signature 1 | PS00975 | [DEK]-[IV]-N-[FS]-L-C-x-H-K | 2.31e-10 | Site: 159 to 167 EVNFLCVHK |
| Myristoyl-CoA:protein N-myristoyltransferase signature 2 | PS00976 | K-F-G-x-G-D-G | 4.31e-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(3,1)-[WY]-C-V-[DN]-V-[EGD]-F-V-[DN]-C-V | -1.08675 | 1.12492 | 95-107 | YVEDDDNVFRFNY |
| Kringle domain signature | PS00772 | V-[DN]-V-[EGD]-F-V-[DN]-C-V | -0.68167 | -1.06868 | 160-167 | VNFCVHK |
| Histone H1 signature 2 | PS00959 | P-F-x-[RA]-L-[IVA]-[KQR]-[DEG]-[IV] | -0.872559 | -1.0624 | 161-169 | NLFKVHKS |

Banerjee, et al.: In silico analysis of Plasmodium falciparum N-myristoyltransferase
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 Ångstrom 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].

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  |

Banerjee, et al.: In silico analysis of Plasmodium falciparum N-myristoyltransferase

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
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

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

| Region of plots                              | Target | Percentage | Template | 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 [X]          | 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          | 1        | 1          |
| Glycine residues                             | 13     | 17         |          |            |
| Proline residues                             | 12     | 19         |          |            |
| Total number of residues                     | 410    | 342        |          |            |

Figure 5: (a) Secondary structure, physiochemical profile and solvent accessible surface area as predicted by POLYVIEW. H-α and other helices (view 1), E-β and other helices (view 2), C-coil, Relative solvent accessibility (RSA). 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 physicochemical profiles, solvent accessible surface area using POLYVIEW. 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, I17, 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, 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=2.192) 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](image)

![Figure 7: Second (a) and third (b) best benzofuran molecule found in the docking calculations](image)
We have performed a thorough validation servers and was found to be reasonably good. Model obtained was assessed using various structure constructed based on a sequence homology of 44.3%. by us resulted in a good quality model as the model was target and template sequences. The methodology adopted Plasmodium falciparum. determining important features and properties of the NMT structures. We have applied bioinformatics tools for details about protein structure in absence of crystal modeling is often envisioned to play a role in providing 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.

ACKNOWLEDGMENTS

Authors are thankful to the Director, Indian Institute of Chemical Technology, for his constant support and encouragement throughout the study. AKB thanks Council of Scientific and Industrial Research (C.S.I.R) for Senior Research Fellowship (SRF).

REFERENCES

1. WHO report projections of mortality and the burden of disease. 2006.
2. Bioland 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. Wellens TE. Plasmodium chloroquine resistance and the search for a replacement antimalarial drug. Science; 2002; 298:124-46.
4. Snow RW, Guerra CA, Meltzer H, Hay SI. International funding for malaria control in relation to populations at risk of stable Plasmodium falciparum transmission. PLoS Medicine 2008; 7:e142.
5. Olivarra PI, Taylor WR. Antimalarial compounds: From bench to bedside. J Exp Biol 2003; 206:3753-9.
6. Trouiller P, Olivarra P, Torrecle 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, Zupe ME, Lu HF, Nagarajan S, Krishne 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. Bhattacharjee RS, Schell 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;36:6780-8.
9. Rudnick DA, McWherter CA, Roquefu WJ, Lennon PJ, Getman DP, Gordon JL. 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 JL. Structures of Saccharomyces cerevisiae N-myristoyltransferase with bound myristoyl-CoA 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 JL. The biology and enzymology of protein N-myristoylation. J Biol Chem 2001;276:39014-4.
13. Gordon JL. Protein N-myristoylation: Simple questions, unexpected answers. Cell Res 1990;38:517-28.
14. Hanks KK, Martingale A. Post-translational chemical modification(s) of proteins. In J Biochem 1992;24:19-28.
15. Johnson DR, Bhattacharjee RS, Knoll L, Gordon JL. Genetic and biochemical studies of protein N-myristoylation. Annu Rev Biochem 1994;63:809-914.
16. Gunaratne RS, Saidj M, Ling IT, Tripathi R, Pachetja JA, Holder AA. Characterization of N-myristoyltransferase from Plasmodium falciparum. Biochem J 2000;348:459-63.
Banerjee, et al.: In silico analysis of Plasmodium falciparum N-myristoyltransferase

17. Price HP, Menon MR, Panethmitak 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. Jurcic D, Zoranic L, Zucic D. Basic charge clusters and predictions of membrane protein topology. J Chem Inf Comput Sci 2002;42:620-32.

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

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;25:2153-60.

29. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial alignment surface method. Prot Eng 1997;10:673-6.

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

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

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

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

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

35. Slenders 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.

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

37. Frearson JA, Brand S, McElroy SP, Cleghorn LA, Smid O, Stojanovski I, et al. N-myristoyltransferase inhibitors as new leads to treat sleeping sickness. Mol Microbiol 2011;81:612-24.

38. Jennings BC, Nadolski MJ, Ling Y, Baker MB, Harrison MJ, 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 Biol Chem 1998;273:233-42.

39. Jones G, Willett P, Glen RC, Leach AR, Taylor RJ. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1996;268:201-25.

40. MacKerell AD, Brooks CL, Dunbrack RL Jr., 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.

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

42. Grubmuller H, Heller H, Windemuth A, Schulten K. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Mol Simul 1991;1:61-42.