STRUCTURAL COMPARISON OF PROKARYOTIC AND EUKARYOTIC FOLATE TRANSPORTERS BY COMPUTATIONAL APPROACH

DEEPA AGARWAL, ALOK JHA*

Department of Biotechnology and Life Science, Institute of Biomedical Education and Research, Mangalayatan University, Beswan, Aligarh, Uttar Pradesh, India. Email: alok.jha@mangalayatan.edu.in

ABSTRACT

Objective: *In silico* approach has particularly drawn attention in providing a realistic representation needed to understand the fundamental molecular structure of a transporter. The importance of folate metabolism and role in the internalization of antifolates in eukaryotes have been studied extensively, but the structural study of folate transporters in *Homo sapiens* (HFT), *Plasmodium falciparum* (PFT), and *Streptococcus* sp. (SFT) is still lacking. This study was conducted to study and compare the structures of prokaryotic and eukaryotic folate transporters.

Methods: HFT, PFT, and SFT were queried using blast and sequences were retrieved using National Center for Biotechnology and Information (NCBI) databases. This was superseded by structural and functional prediction of transporters. The structure has been generated using Swiss model which was visualized using PyMol and validated by Procheck and ERRAT analysis along with the values of different secondary structures mapping to diverse sections of the Ramachandran plot. The structural and functional comparison was performed by PROSO, ProFunc, TM Score, Porewalker, TMHMM, and Protscale.

Result: All the parameters for structural comparison suggest that *H. sapiens* folate transporter is 16.67% and 17.72% identical to *Plasmodium* and *Streptococcus* whereas *Plasmodium* is 21.59% identical to *Streptococcus*. The evaluation of transmembrane helices and hydrophobicity resulted in the presence of 1, 4, and 12 membrane-spanning segments with predicted US, UDUD, and UDS as pore shape in *Plasmodium*, *Streptococcus*, and humans.

Conclusion: Such folate receptors are the main targets for the specific conveyance of antifolates. The differences found between these species may offer possibilities for the development of new drugs in future.

Keywords: Human folate transporters, *Plasmodium* folate transporter, *Streptococcus* folate transporter, Structural study, Software.

INTRODUCTION

Folic acid (or folate) is an essential B group vitamin nutritionally important in synthesizing cofactors required for cell proliferation and tissue repair [1]. One of the major folates is tetrahydrofolate which is utilized in donating one-carbon units for thymidylate and purine nucleotide synthesis, conversion of homocysteine into methionine and interconversion of serine into glycine [2]. In addition to the role as a constituent of protein, it plays a major role in the synthesis of S-adenosylmethionine necessary for methylation of RNA, DNA, some neurotransmitter, phospholipid, and proteins like histones. Bacteria, protists, fungi, and plants synthesize folates *de novo,* but higher organisms such as animals and human beings lack key enzymes of synthesis pathway thus require them in the diet which implies the pivotal role of different types of folate transporters [3].

Folate transporter belongs to the major facilitator superfamily (MFS), largest known secondary transporters for the passage of a variety of ligands such as peptides, drugs, nucleotides, monosaccharides oligosaccharides, cations, and anions [4]. Folates are transported through three diverse types of transporters, including: (i) Reduced folate carrier (RFC) (SLC19A1): Which transports reduced folates specifically 5 methyltetrahydrofolate to the systematic cells at neutral pH. Structurally, it is an integral protein with 12 transmembrane domains (TMDs) with amino and carboxyl-terminal at cytoplasmic side [5,6]. These receptors are located in placenta and significantly in lung, kidney, leukocytes, bone marrow, and some portions of the central nervous system. (ii) Proton-coupled folate transporter (PCFT): That pumps the protons down the electrochemical gradient for transportation of folates into cells at low pH, unlike RFC. Human PCFTs are located on

chromosome 17q11.2 having 12 TMDs with cytosolic oriented C and N termini encoding 459 amino acids [7]. Functionally, they are active in apical brush border membrane of jejunum and duodenum besides the basolateral membrane of choroid plexus, kidney, and placenta [7,8]. (iii) Folate receptors: Which bind to diverse form of folate conjugates. There are three folate receptors - FRα, FRβ, and FRγ which are usually expressed in different cell membranes [9]. FRs uptake of cellular folate through invagination process and forms the endosome vesicle which on acidification dissociates into separate entities and exits the endosome by diffusion [10]. Both FRα and FRβ are functional in malignant tissues, thus are moieties for tumor target of toxins and cytotoxic agents.

The presence of folate synthesizing and salvage genes in *Plasmodium falciparum* and *Streptococcus anginosus* unlike human explain the difference and reflect at folate synthesizing proteins as the target of antifolates effectively used for the treatment of malaria and other infectious disease like abscesses.

This intriguing difference in the role of transporters in prokaryotes and eukaryotes reflects the structural and functional study between them.

Folate intake is an ATP dependent symport mechanism mediating folate transporters. Wang *et al*. reported such uptake as a saturable process inhibited by probenecid and furosemide [11]. These studies suggested the sensitivity of folate transporters in the presence of inhibitors thus highlighting the accessibility of folate transporters as a drug target explaining the importance of drug susceptibility and parasite survival. However, the structural and functional comparison of folate transporters between prokaryotes and eukaryotes has not been conducted yet making it a novel attempt.

METHODS

Retrieval of sequences

The protein sequences of *Homo sapiens* folate transporter (HFT), *P. falciparum* (PFT), and *Streptococcus anginosus* (PFT) were retrieved from the NCBI databases. All the sequences were downloaded in FASTA format. Blastp [12] search homology with RCSB PDB database identifies the template. On the basis of query coverage, identities and E value, the template of folate transporter homolog of *H. sapiens*, *P. falciparum,* and *S. anginosus* were obtained from Blastp database.

Model generation and quality validation

Protein 3D homology models were obtained through Swiss model [13], a web server which computes models on the basis of ProMod3, by placing target sequence as input. The model with lowest Q means, Z score, and highest sequence identity were considered as the best model. Selected models were visualized by pymol and validated by several servers to check the stereochemical quality of constructed model such as Procheck [14] calculating ERRAT analysis, Z score [15], and the program verifying the protein structure based on statistics of non-bonded atom-atom interactions. The quality of models was further assessed by Ramachandran plot by RAMPAGE which shows the conformations of φ and ψ torsion angles of an amino acid along with the distribution of data points for structure validation [16].

Structural and functional study

Structural and functional study of transporter proteins was executed by studying various parameters such as amino acid composition, molecular weight, pI, atomic composition, extinction coefficient, instability index, aliphatic index, and grades average of hydropathicity using a web server tool named ProtParam server [17]. For prediction of transmembrane helices, TMHMM server 2.0 is considered, the output result of which is based on Hidden Markov Model predicts 97–98% correct helices along with the discrimination between soluble and membrane proteins [18]. Hydrophobicity analysis was accomplished using ProtScale (Kyte and Dolittle). Prediction of secondary structure was performed by a convenient tool which functions on structural and sequence-based method termed as Profunc [19]. For improved structural and functional analysis, ProFunc tool also reported binding site residues, cleft and nest analysis. Number of binding sites which may participate in ligand binding was figured out by DEPTH, a server dealing with the parameters such as ligand binding cavities and pKa of ionizable residues in proteins [20]. The characterized protein channel, pore shape, pore size, and pore axis were computed using fully automated tool named PoreWalker [20]. Leucine-rich repeats and motif scanning were carried out by leucine-rich repeats (LRR) Finder [21]. HOLLOW facilitates the cavity and channel prediction [22].

Refinement of models by energy minimization

Small errors in the constructed models may hinder its benefit in drug design and protein-protein interaction. To bring them closer to their native state, model refinement was done using 3D Refine by optimizing hydrogen bonding network and atomic level energy through MESHI program. Global test score (GDT_TA), high accuracy score (GDT_HA) at 1Å was examined to understand the quality of refined model with respect to its native structure [23].

Structural comparison of models between prokaryotes and eukaryotes

An efficient tool named iPBA was used for the structural comparison which aligns the protein structure based on PB substitution matrices and anchor-based alignment methodology [24]. Structural and functional comparison of all three transporter species was then tabulated involving the results of all the parameters including amino acid composition, molecular weight, root-mean-square deviation (RMSD) value difference, TMH, hydrophobicity, pore shape and size, secondary structures, binding sites and its residues, number of cavities and pockets, and leucine-rich repeats. Structural comparison was also done to predict the residues at putative protein interface and also explore the solvent accessible areas using SPPIDER, protein interface recognition surface [25].

RESULTS AND DISCUSSION

Retrieved sequence composition

The protein sequences of *H. sapiens* retrieved were 257 amino acids in length with accession number (AN_NP_057937.1), 505 amino acids in *P. falciparum* with accession number (PKC47564.1), 186 amino acids in *S. auringosus* with accession number (AIK78459.1).

Model features of HFT, PFT, and SFT: The selected template using Blastp for *H. sapiens* was crystal structure of FR alpha transporter (PDB ID: 5iq.6.A) with 100% sequence identity, *Escherichia coli* transport protein (PDB ID: 2aw0.1.B) for *P. falciparum* with 100% sequence identity, and ECF transporter (PDB ID: 5d3m.2.C) for *S. auringosus* with 30% sequence identity. These template sequences were the input for a model generation (Fig. 1). The models were validated using ERRAT analysis scoring 98.8 in *H. sapiens*, 90.22 in *P. falciparum,* and 79.66 in *S. auringosus*. In addition, Z score, an energy-based qualifier to determine structural errors ranging −6.98 in *H. sapiens*, −9.69 in *P. falciparum,* and −2.39 in *S. auringosus*. These values are credible as it indicates negative scoring energy characteristic of TM regions. *H. sapiens* gave better results in perspective of Ramachandran plot stipulating zero number of residues in outlier region, 96% in favored region, and 3.9% in allowed region as compared to *S. auringosus* showing 0.5% residues in outlier region, 94.0% in favored region, and 5.4% in allowed region whereas *P. falciparum* showed 2.7% residues in outlier region, 90.5% in favored region, and 6.8% in allowed region. These verification studies validate the model to be stable.

Structural and functional analysis of folate transporters

Folate transporter belongs to highly conserved MFS which functions using electrochemical gradient. The physiochemical parameters of folate transporter protein of *H. sapiens* (HFT), *P. falciparum* (PFT), and *S. auringosus* (SFT) are tabulated in Table 1.

Analysis of amino acid composition reveals that all three of the proteins are leucine-rich in common whereas isoleucine is found to be highest in *P. falciparum* as compared to alanine and tryptophan in *H. sapiens* and valine in *S. auringosus* whereas atomic acid composition shows that all these proteins are rich in hydrogen and carbon. The charge on amino acid residues displays net positive charge on all the three proteins relating to highest extinction coefficient in *H. sapiens* as compared to *P. falciparum* and *S. auringosus* validating the tryptophan-rich human folate transporter. Highest instability coefficient in human folate transporter implies the most stable structure followed by *S. auringosus* and *P. falciparum.* Estimated lifespan was found similar for all the three proteins.

The 3D structure of each protein demonstrates the presence of N and C terminal domains having two-fold symmetry axis perpendicular to
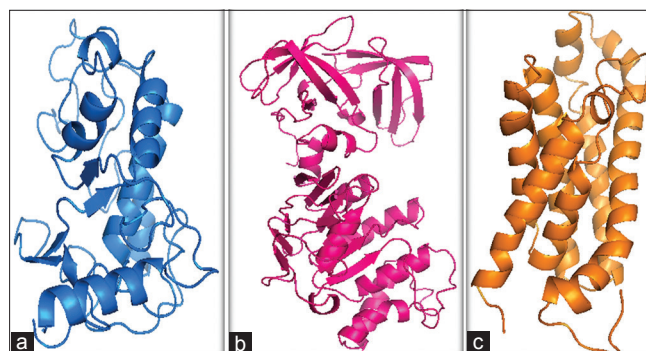

Fig. 1: *In silico* **models of folate transporter generated by Swiss model server, (a):** *Homo sapiens***, (b):** *Plasmodium falciparum***, (c):** *Streptococcus auringosus*

Table 1: Physiochemical features of HFT, PFT, SFT

Parameters	HFT	PFT	SFT
Number of amino acids	257	505	186
Molecular weight	29819.3	58060.42	21000.44
Theoretical pI	8.30	8.05	9.85
Total number of negative charged residues	22	35	8
Total number positive charged residues	20	38	17
Formula of atomic composition	$C_{1334}H_{1991}N_{371}O_{366}S_{23}$	$C_{270}4H_{4192}N_{620}O_{736}S_{28}$	$C_{1005}H_{1573}N_{233}O_{239}S_{08}$
Total number of atoms	4085	8280	3058
Extinction coefficient	100920	52105	32430
Instability coefficient	41.69	34.80	35.54
Aliphatic composition	64.20	116.79	126.77
Estimated half-life	30	30	30
Grand average of hydropathy	0.418	0.490	0.801

H. sapiens: Homo sapiens, P. falciparum: Plasmodium falciparum, S. anginosus: Streptococcus anginosus, HFT*: Folate transporter of H. sapiens,* FPT*: P. falciparum,*
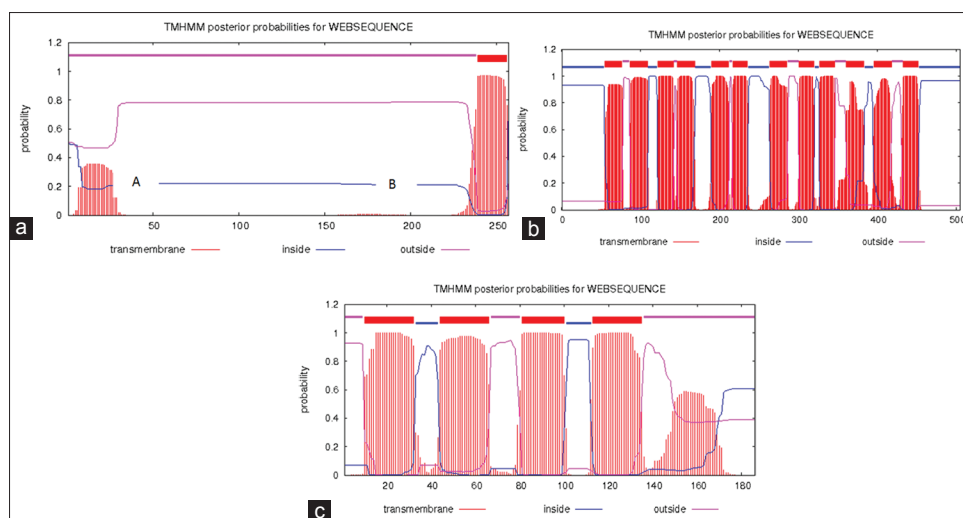SFT*: S. anginosus*

Fig. 2: Transmembrane helices prediction by TMHMM2.0 server (a): *Homo sapiens*, (b): *Plasmodium falciparum*, (c): *Streptococcus auringosus*
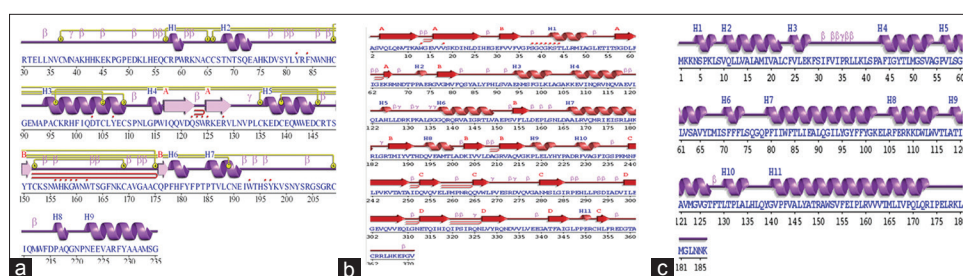

Fig. 3: Secondary structure prediction of folate transporter of (a): *Homo sapiens*, (b): *Plasmodium falciparum*, (c): *Streptococcus auringosus*

the membrane plane. Each domain is further divided into TMH. The HFT comprise only 1TM of 27 amino acids while PFT has 12 TM of 257 amino acids and SFT has 4TM of 101 amino acids. Since the number of amino acids exceeds 18, we suggest that all the folate transporters are transmembrane proteins. The probabilities for the N-terminus to be on cytoplasmic side for folate transporter of human, *Plasmodium,* and *Streptococcus* are 0.49, 0.933, and 0.07, respectively (Fig. 2). In human folate transporter, a higher number of amino acid lies on the negative side of hydropathy scale implying the hydrophilic nature of protein whereas a higher number of amino acids in *Plasmodium Streptococcus* and lies on the positive side of scale suggesting the hydrophobic character.

Secondary structure prediction results in HFT, PFT, and SFT indicated the number of helices (Fig. 3). The HFT secondary structure reveals 9

alpha chains, 28 beta turns, 2 gamma turns, 7 helix-helix interactions, and 8 disulfides. The PFT structure has 11 alpha chains, 38 beta turns, 10 gamma turns, 9 helix-helix interactions, and 7 beta hairpins whereas SFT structure indicates the presence of 11 alpha chains, 6 beta turns, 1 gamma turns, and 23 helix-helix interactions.

We analyzed tertiary structure for cleft and binding sites using ProFunc server and found putative binding sites in the cytoplasmic region inside the transmembrane region in Plasmodium in contrast with human and *Streptococcus* having binding sites outside the transmembrane as shown in Fig. 4. ProFunc also predicts "nest" near putative phosphorylation site which is characteristic functional motifs. We also studied features of a pore to understand the geometry of the channel in these proteins. The pore is found to be cylinder shape in human whereas conical

shape in *Plasmodium* and diamond shape in *Streptococcus* resulting in unique types of pore from each other by which ligand may pass. All are reported to have pore size as 3 Å (Fig. 5).
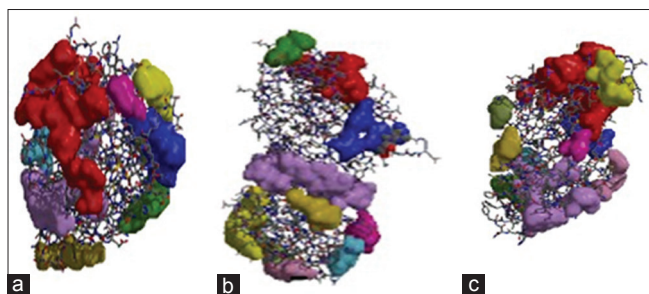

Fig. 4: Binding site prediction by profunc (a): *Homo sapiens*, **(b):** *Plasmodium falciparum*, **(c):** *Streptococcus auringosus*
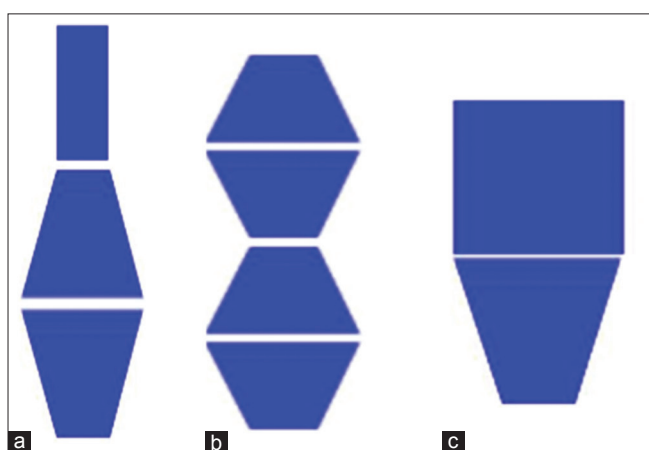

Fig. 5: Pore shape by porewalker (a): *Homo sapiens*, **(b):** *Plasmodium falciparum*, **(c):** *Streptococcus auringosus*
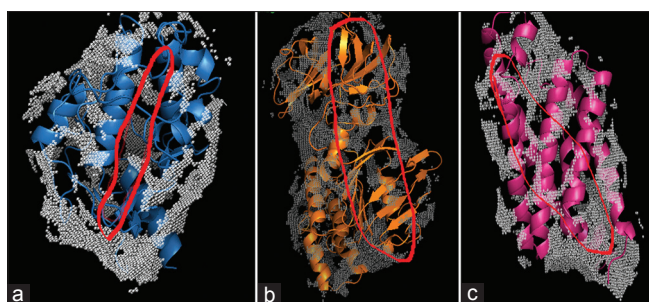

Fig. 6: Cavity prediction by HOLLOW (a): *Homo sapiens*, **(b):** *Streptococcus auringosus,* **(c):** *Plasmodium falciparum*

Structural refinement of transporters

Superimposition of energy minimized structures with that of modeled structure result in the fluctuation of RMSD value from 2.90 to 0.13 in *H. sapiens* while 0.14–2.65 and the highest reduction in RMSD value were found to be in *S. auringosus*, 3.95–0.14. Such reduction in RMSD value involves the elimination of 5 atoms from modeled structure in human and Streptococcus whereas 7 atoms in *Plasmodium*. Higher reduction in RMSD values in S. auringosus is subjected to the highest 3D refine score and Mol Prob score followed by *H. sapiens* and *P. falciparum* (Table 2).

Structural comparison between human, *Plasmodium,* **and** *Streptococcus* **folate transporter**

All the parameters studied under the structural analysis have been compared and tabulated between HFT, PFT, and SFT. Structural comparison suggests that *H. sapiens* folate transporter is 16.67% and 17.72% identical to *Plasmodium* and *Streptococcus* whereas *Plasmodium* is 21.59% identical to *Streptococcus*. Although all the membrane proteins are rich in leucine, higher positive residues in human make it more compatible for antimicrobial activity and hemolytic activity. In fact presence of high amount of valine in *Streptococcus* dramatically reduces toxicity and therapeutic index. Presence of 3 LRRs in *Plasmodium* and 1 each in human and *Streptococcus* indicate the importance of protein phosphorylation playing a fundamental role in drug transport across the cell membrane. Previous studies reported the effect of altering phosphorylation sites in mammalian transport homolog, PfMDR1 on drug action. All the three proteins have conserved residues that make up the predicted nest. No. of nest in humans, *Plasmodium,* and *Streptococcus* are 8, 5, and 1 which are functional motifs. Being eukaryotic, human, and *Plasmodium* folate transporter share some signature sequences in two cytosolic loops in MFS family.

In respect of TMH, *Plasmodium* comprises 12 TM with N and C terminus on cytosolic side validating the same in MFS proteins. Hydrophobicity of human folate transporter as compared to *Plasmodium* and *Streptococcus* predicts the rate of protein crossing the membrane. Greater number of binding site residues and solvent accessible areas in *Streptococcus* higher will be the affinity of molecules to bind. The different shapes of the pore of transporter protein will help in designing unique ligands.

Prediction of the cavity is important in ligand binding, molecular transport, and enzyme catalysis. 103 number of binding residues in Streptococcus

Table 2: Scoring by 3d refine of HFT, PFT, SFT

Model	3D refine score	GDT-HA score	Mol Prob	RMSD value
Human	12895.2	0.131	1.849	0.13
Plasmodium	1.0000	0.999	2.082	0.14
Streptococcus	13455.6	0.142	3.17	0.14

RMSD: Root-mean-square deviation, *H. sapiens: Homo sapiens,*
P. falciparum: Plasmodium falciparum, S. anginosus: Streptococcus anginosus,
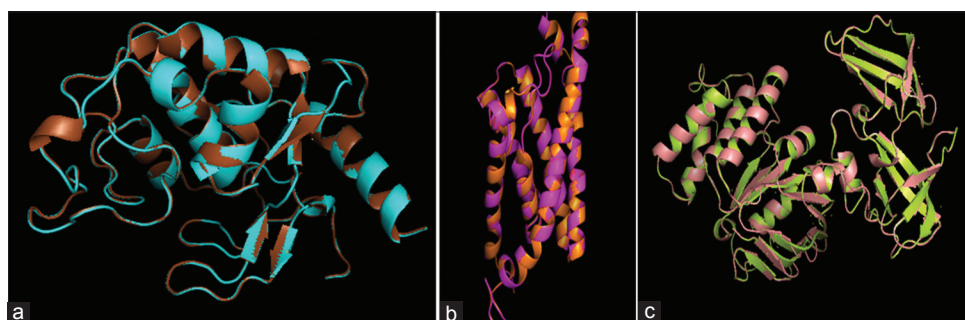HFT: *Folate transporter of H. sapiens*, FPT: *P. falciparum*, SFT: *S. anginosus*

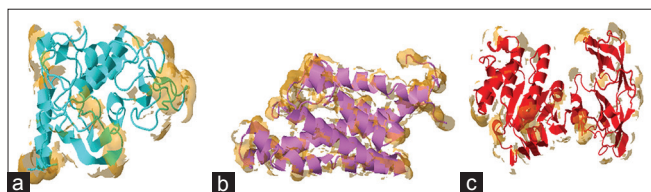**Fig. 7: Alignment of refine model over generated model by iPBA (a):** *Homo sapiens*, **(b):** *S Streptococcus auringosus,* **(c):** *Plasmodium falciparum*

Fig. 8: Solvent accessible areas highlighted at the surface by SPPIDER; (a): *Homo sapiens* **(b):** *Streptococcus auringosus* **(c):** *Plasmodium falciparum*

as compared to 33 in human and 31 in *Plasmodium* indicated the better chance of ligand binding and surpassing the molecule (Fig. 6).

Alignment of models by iPBA of human with *Plasmodium* and *Plasmodium* with *Streptococcus* folate transporter aligns only 31 number of residues as compared to 53 when alignment was done between Human and Streptococcus confirming the wide difference in HFT and PFT as compared to HFT and SFT. These differences in folate transporter proteins can prove to be the best target of drug designing (Fig. 7).

SPPIDER server helped us to distinguish the most likely interface sites at the surface of proteins. These accessible areas are well-studied considering parameters such as van der wall surfaces, dot surfaces, and mol prob surfaces (Fig. 8). The most likely interface sites were attributed to the element whose accessible areas are greater than others.

CONCLUSION

This study integrates structural and statistical approaches for modeling of folate transporters. These structure models are in agreement with other crystallographic structures. This structural comparison study will enhance our understanding about folate transporters in different species. Such differences between eukaryotic and prokaryotic folate transporters can suggest several binding and interaction sites for drug designing and help us to identify drug targets for therapeutic development.

REFERENCES

1. Bates CJ. FOLIC ACID, Properties and determination A2-caballero, Benjamin. Encyclopedia of Food Sciences and Nutrition. 2nd ed. Boston, MA: Academic Press; 2003. p. 2559-64.
2. Salcedo-Sora JE, Ochong E, Beveridge S, Johnson D, Nzila A, Biagini GA, *et al*. The molecular basis of folate salvage in *Plasmodium falciparum*: Characterisation of two folate transporters. J Biol Chem 2011;286:44659-68.
3. Shane B. Folylpolyglutamate synthesis and role in the regulation of one-carbon metabolism. Vitam Horm 1989;45:263-335.
4. Burckhardt G, Wolff NA. Structure of renal organic anion and cation transporters. Am J Physiol Rena Physiol 2000;278:F853-66.
5. Matherly LH, Goldman DI. Membrane transport of folates. Vitam Horm 2003;66:403-56.
6. Matherly LH, Hou Z, Deng Y. Human reduced folate carrier: Translation of basic biology to cancer etiology and therapy. Cancer Metastasis Rev 2007;26:111-28.
7. Desmoulin SK, Hou Z, Gangjee A, Matherly LH. The human proton-coupled folate transporter: Biology and therapeutic applications to cancer. Cancer Biol Ther 2012;13:1355-73.
8. Zhao R, Goldman ID. The molecular identity and characterization of a proton-coupled folate transporter-PCFT; biological ramifications and impact on the activity of pemetrexed. Cancer Metastasis Rev 2007;26:129-39.
9. Elnakat H, Ratnam M. Distribution, functionality and gene regulation of folate receptor isoforms: Implications in targeted therapy. Adv Drug Deliv Rev 2004;56:1067-84.
10. Rijnboutt S, Jansen G, Posthuma G, Hynes JB, Schornagel JH, Strous GJ. Endocytosis of GPI-linked membrane folate receptor-alpha. J Cell Biol 1996;132:35-47.
11. Wang L, Cherian C, Desmoulin SK, Polin L, Deng Y, Wu J, *et al*. Synthesis and antitumor activity of a novel series of 6-substituted pyrrolo[2,3-d]pyrimidine thienoyl antifolate inhibitors of purine biosynthesis with selectivity for high affinity folate receptors and the proton-coupled folate transporter over the reduced folate carrier for cellular entry. J Med Chem 2010;53:1306-8.
12. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, *et al*. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 1997;25:3389-402.
13. Schwede T, Kopp J, Guex N, Peitsch MC. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 2003;31:3381-5.
14. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK - A program to check the stereochemical quality of protein structures. J Appl Cryst 1993;6:283-91.
15. Zhang L, Skolnick J. What should the Z-score of native protein structures be? Protein Sci Publ Protein Soc 1998;7:1201-7.
16. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol 1996;7:95-9.
17. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, *et al*. Protein identification and analysis tools on the ExPASy server. In: Walker JM, editor. *The Proteomics Protocols Handbook. Totowa, NJ: Humana Press;* 2005. p. 571-607.
18. Krogh A, Larsson B, Von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 2001;305:567-80.
19. Laskowski RA, Watson JD, Thornton JM. ProFunc: A server for predicting protein function from 3D structure. Nucleic Acids Res 2005;33:W89-93.
20. Tan KP, Nguyen TB, Patel S, Varadarajan R, Madhusudhan MS. Depth: A web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pK_a of ionizable residues in proteins. Nucleic Acids Res 2013;41:314-21.
21. Pellegrini-Calace M, Maiwald T, Thornton JM. PoreWalker: A novel tool for the identification and characterization of channels in transmembrane proteins from their three-dimensional structure. PLoS Comput Biol 2009;5:1000440.
22. Ujwal R, Cascio D, Colletier JP, Faham S, Zhang J, Toro L. The crystal structure of mouse VDAC1 at 2.3Å resolution reveals mechanistic insights into metabolite gating. Proc Natl Acad Sci 2008;105:406-9.
23. Bhattacharya D, Nowotny J, Cao R, Cheng J. 3D Refine: An interactive web server for efficient protein structure refinement. Nucleic Acids Res 2016;44:406-9.
24. Gelly JC, Joseph AP, Srinivasan N, de Brevern AG. iPBA: A tool for protein structure comparison using sequence refinement. Nucleic Acids Res 2011;39:18-23.
25. Porollo A, Meller J. Prediction-based fingerprints of protein-protein interaction proteins: Structure. Funct Bioinformatics 2007;66:630-45.