

Plasmodium vivax 1-deoxy-D-xylulose-5-phosphate synthase: Homology Modeling, Domain Swapping, and Virtual Screening

Divya Ramamoorthy, Sumit Handa, David J Merkler and Wayne C Guida*

Department of Chemistry, University of South Florida, USA

Abstract

Structure-based computational approaches are needed to model proteins in the absence of any crystal structures and identify protein-ligand interactions. Biochemical pathways that exist in microorganisms but absent in humans serve as excellent targets for antimicrobial drug design. The Non-Mevalonate Pathway (NMP) is one such pathway that is present in all intra-erythrocytic stages of *Plasmodium* and could serve as a target for anti-malarial drug design and development. The first enzyme of the pathway, DXS (1-deoxy-D-xylulose-5-phosphate synthase) is the rate limiting enzyme and is also important for the biosynthesis of pyridoxal and thiamine. In the absence of available crystal structures, our aim was to develop homology models for *Plasmodium* DXS, which could provide insight into the structural features of this enzyme and its likely binding to ligands. Initial models were built using the PRIME module of Schrödinger Suite 2010 and then refined using MacroModel energy minimization. Analyses were also carried out using bioinformatics tools to predict domain swapping in *Plasmodium* DXS. This study should prove useful in the design and development of novel anti-malarial therapeutics.

Keywords: *Plasmodium*; Homology modeling; Energy minimization; Active site; Domain swapping; Virtual screening

Introduction

Biochemical processes that exist in bacteria, but not in humans, provide a useful way to develop novel inhibitors to combat diseases. One such pathway is the Non-Mevalonate Pathway (NMP) exclusively present in lower animals and bacteria, which is utilized to synthesize isoprenoids for terpene biosynthesis [1]. Isopentenyl diphosphate (IPP) and Di methyl Allyl phosphate (DMAPP) are the end products of this pathway and serve as precursors for the biosynthesis of various terpenoids including cholesterol and vitamin E [2]. Since higher animals utilize the mevalonate pathway for isoprenoid biosynthesis, the NMP is an attractive biochemical pathway, the enzymes of which provide several targets for the discovery of anti-infectious therapeutic agents.

Among the infectious diseases, malaria ranks highest, affecting about 2.4 billion people worldwide [3]. Although drugs are available to treat malaria, recurring antimicrobial resistance to the currently available drugs poses a significant threat to successful treatment. Discovering new molecular targets and new class of inhibitors are essential to overcome the issue of microbial resistance to drugs. The apicoplast, a non-photosynthetic plastid found in *Plasmodium*, is nuclear encoded [1] and is necessary to the survival of the malaria parasite [4]; it harbors many pathways including the NMP [5]. Since the non-mevalonate pathway is present in all intra-erythrocytic stages of *Plasmodium*, it represents an extremely attractive target for antimalarial drug discovery and development.

Major causative agents of malaria include *Plasmodium vivax*, *Plasmodium falciparum*, *Plasmodium berghei*, *Plasmodium ovale*, *Plasmodium malariae* and *Plasmodium knowlesi*, out of which *Plasmodium falciparum* and *Plasmodium vivax* are most important for pathogenesis of the disease [6]. *Plasmodium falciparum* is deadly while *Plasmodium vivax* causes less severe complications but is more widespread, especially in temperate zones. Moreover, *Plasmodium vivax* can hibernate in the liver for months or years and can resurface, causing disease [6].

The first step involved in the NMP is formation of 1-deoxy-D-xylulose-5-phosphate (DXP) by condensation of glyceraldehyde-3-

phosphate (G3P) and pyruvate catalyzed by 1-deoxy-D-xylulose-5-phosphate synthase (DXS) [7]. The DXS enzyme has a thiamine-binding motif and requires TPP (thiamine pyrophosphate) and either a Mn^{2+} or Mg^{2+} divalent cation to manifest its activity [7]. DXS catalyzes the first step of the NMP, which is also the rate limiting step. Formation of DXP is not a committed step of the non mevalonate pathway as DXS is also essential for thiamine and pyridoxal biosynthesis [8], thus making it an alluring target to develop drugs for anti-infectious diseases. By targeting enzymes present in the NMP for inhibition, the DXS enzyme in either *P. falciparum* or could potentially lead to the development of potent curative agents for malaria and other infectious diseases.

To better understand the structural components of DXS, a crystal structure would be valuable. The only crystal structures available for this enzyme to date are from *E. coli* and *D. radiodurans*. Here, we present homology models for *P. vivax* and other *Plasmodium* species (*P. falciparum*, *P. berghei*) of DXS in order to comprehend its structural features and for use in virtual screening of compound libraries readily available from the National Cancer Institute and other sources.

Methods and Results

Comparative modeling generates a 3-D structure of an enzyme based on previously determined X-ray crystal structures. The sequence of PvDXS and other *Plasmodium* DXS were obtained from PUBMED and were aligned with the sequences of *E. coli* and *D. radiodurans*, the available crystal structures for DXS using ClustalW [9]. Two crystal

*Corresponding author: Wayne C Guida, Professor, Department of Chemistry, University of South Florida, 4202 E fowler Avenue, Tampa, FL 33620, USA, Tel: 727-452-2986; E-mail: wguida@usf.edu

Received June 25, 2014; Accepted July 30, 2014; Published August 08, 2014

Citation: Ramamoorthy D, Handa S, Merkler DJ, Guida WC (2014) *Plasmodium vivax* 1-deoxy-D-xylulose-5-phosphate synthase: Homology Modeling, Domain Swapping, and Virtual Screening. J Data Mining Genomics Proteomics S1:003. doi:10.4172/2153-0602.S1-003

Copyright: © 2014 Ramamoorthy D, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

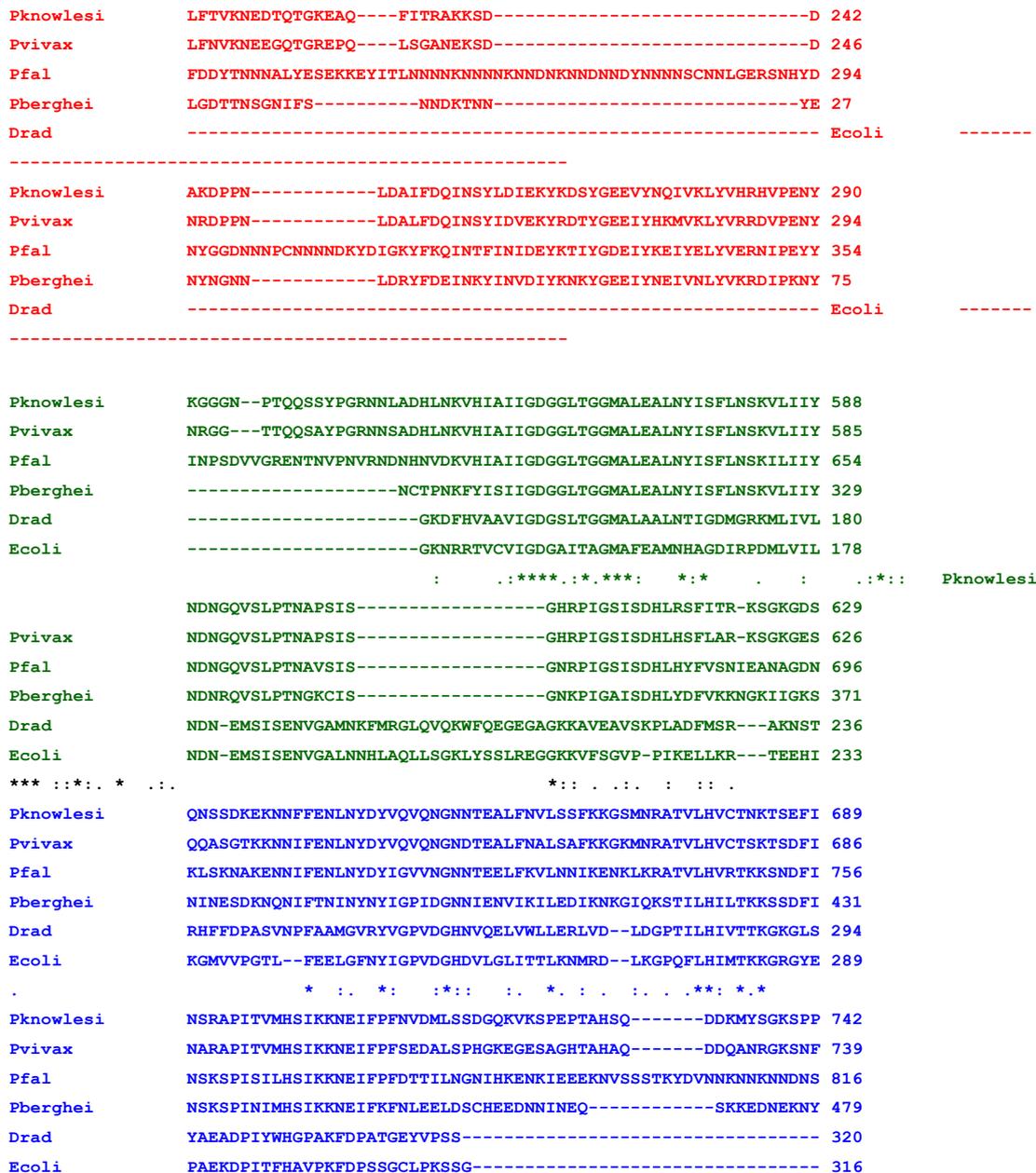


Figure 1: Multiple alignment of DXS of *P.knowlesi*, *P.vivax*, *P.falciparum*, *P.berghei*, *D.radiodurans*, and *E.coli*, using ClustalW in the predicted domain linker regions; the consensus sequence characterizes identical residues as '*', strongly similar residues as ':' and weakly similar residues as '.'.

structures of DXS (from *E. coli* and *D. radiodurans*) have been reported to date [10]. The aligned structures were manually inspected for any anomalies and corrected to reflect the structurally conserved regions. Important regions from the alignment file are shown in (Figure 1). The regions shown in (Figure 1) represent the segment that connects the transit peptide to domain I, domain I - domain II linker region, and the domain II - domain III linker region. Alignment scores between the different species of DXS are shown in (Table 1).

Homology model of PvDXS

The sequence of DXS *P. vivax* was obtained from NCBI protein sequence database, (<http://www.ncbi.nlm.nih.gov>), and it consists

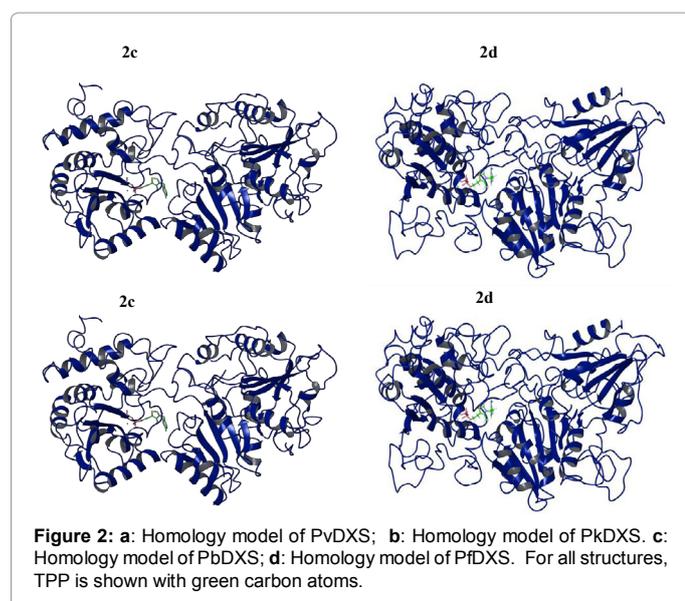
of 1111 residues. The sequence of PvDXS was then aligned with the sequences of previously resolved crystal structures for DXS, *E. coli* and *D. radiodurans*, and the alignment scores shown in (Table 1).

Aligning the sequences and building the homology model:

Alignment of sequences is a crucial step in building a homology model to create a connection between the reference and the target proteins. Sequences should be properly aligned to ensure high quality of the model. Hence it is important to thoroughly inspect the sequences, especially in the conserved regions. Multiple sequence alignment of PvDXS with sequences of *E. coli* (PDB ID: 2O1S) [10], *D. radiodurans* (PDB ID: 2O1X) [10] and *P.falciparum* obtained from ClustalW 2.0.12 [9] are shown in (Figure 1). The aligned structures were manually

Sequence 1	Sequence 2	Alignment Score
<i>D. radiodurans</i>	<i>P. berghei</i>	28
<i>D. radiodurans</i>	<i>P. knowlesi</i>	28
<i>D. radiodurans</i>	<i>P. vivax</i>	29
<i>E. coli</i>	<i>P. knowlesi</i>	29
<i>D. radiodurans</i>	<i>P. falciparum</i>	30
<i>E. coli</i>	<i>P. berghei</i>	30
<i>E. coli</i>	<i>P. vivax</i>	30
<i>E. coli</i>	<i>P. falciparum</i>	31
<i>D. radiodurans</i>	<i>E. coli</i>	45
<i>P. falciparum</i>	<i>P. knowlesi</i>	50
<i>P. falciparum</i>	<i>P. vivax</i>	50
<i>P. berghei</i>	<i>P. vivax</i>	56
<i>P. berghei</i>	<i>P. knowlesi</i>	57
<i>P. falciparum</i>	<i>P. berghei</i>	63
<i>P. knowlesi</i>	<i>P. vivax</i>	74

Table 1: Alignment scores between different organisms of DXS.



inspected and edited to reflect the structurally conserved regions in the available crystal structures. The tertiary structure of the three proteins can be considered similar, since they share a significant degree of similarity, with a sequence identity of 30% and 29% with *E. coli* and *D. radiodurans* respectively. Relatively reliable homology models can be expected when the sequence identity between the template and the query sequence is 30% or greater [11]. In PvDXS, since the C- and N-termini are not likely to contribute significantly toward

ligand binding [12], some residues in these regions were truncated while building the model (Figure 2). The model generated for PvDXS has residues starting from 375 through 1011. Since templates were not available to model the starting sequence (residues ranging from 1-374), the model starts at residue 375. If the templates had gaps of more than 20 amino acid residues, the gaps were retained in the model. Thiamine pyrophosphate (TPP) and the divalent metal ion (Mg^{2+}) were retained from the X-ray crystal structure of *D. radiodurans*, while generating the model. The initial model was generated using the PRIME software, available from Schrödinger, Inc. [13] and then

subjected to energy minimization to produce a refined model for docking studies. MacroModel, also available from Schrödinger, Inc. [14-16], was used for geometry optimization and it relied upon the limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm [13] for energy minimization with the OPLS 2005 Force Field. MacroModel minimization was performed for 500 iterations or until energy gradient convergence threshold of 0.1 kJ/mol/Å was obtained.

Homology model of *P. knowlesi*, *P. falciparum* and *P. berghei* DXS: The sequences of *P. knowlesi*, *P. falciparum* and *P. berghei* DXS were obtained from the NCBI. The sequence of *P. knowlesi* consists of 1122 residues, *P. falciparum*, 1205 residues, and *P. berghei*, 843 residues respectively. They were aligned with the sequences of *E. coli* and *D. radiodurans* as described earlier. The sequence of *P. knowlesi* has high homology to *P. vivax* than *P. falciparum* or *P. berghei* and *P. falciparum* has high homology to *P. berghei*. The models of *P. knowlesi*, *P. falciparum* and *P. berghei* were built using the aligned templates of *D. radiodurans* and *E. coli*; coordinates of TPP and the divalent metal cation, Mg^{2+} were retained from the *D. radiodurans* crystal structure.

Protein structure validation

The energy minimized structures were assessed for overall quality using PDBSum/ProCheck [17,18]. ProCheck is utilized to interrogate the stereochemical quality of a given protein structure by verifying the accuracy of parameters like bond angles, bond lengths, torsion angles and correctness of amino acid chirality. Using these criteria, the PvDXS model compares well with the available X-ray crystal structure of DrDXS and EcDXS.

An important indicator of the stereo chemical integrity of the model is the distribution of the main chain torsion angles: phi and psi, which can be examined using Ramachandran plots [17]. (Table 2) shows the Ramachandran plots and statistics of our homology models of *Plasmodium* DXS. The plots (Figure 3a) clearly show that the vast majority of the residues are in a phi-psi distribution consistent known secondary structure. The remaining residues that fall into the random conformation are small segments and are primarily present in the loop regions of the protein. For comparative purposes, Ramachandran plot statistics of *E. coli* and *D. radiodurans* DXS from the X-ray crystal structures are included in (Figure 3b). It is noteworthy that the homology model of PvDXS had 80.5% of total residues present in the favorable regions, while the disallowed regions contain about 1.4% of the residues.

Active-site identity

Active-sites of the homology models of DXS generated were compared with the active-sites in the crystal structures of *D. radiodurans* DXS and *E. coli* DXS. It was found that the active-sites of the models were quite similar to the active sites relative to the templates used. A comparison of the residues in the active sites is tabulated in (Table 3). It is evident from (Table 3) that the active sites share a high degree of sequence identity/similarity. We also analyzed the structure for PKC-phosphorylation sites and found that residues 27, 46, 75, 131, 148, 156, 162, 194, 231, 303, 438, 446, 621, 632, 679, 697, 790, 819 are prone to phosphorylation by Protein Kinase C. Residue 494 was identified to be prone to C-AMP phosphorylation.

Discussion

Structure of *D. radiodurans* and *E. coli* DXS

The crystal structure of *D. radiodurans* and *E. coli* DXS are both homo dimers, with each containing three distinct domains; domain I

Plot Statistics	PvDXS		PkDXS		PfDXS		PbDXS		DrDXS		EcDXS	
Residues in most favored regions [A,B,L]	446	80.5%	463	78.3%	486	71.1%	450	81%	1592	84.9%	1580	91.40%
Residues in additional allowed regions [a,b,l,p]	84	15.2%	99	16.8%	156	22.8%	78	14.2%	247	13.2%	132	7.60%
Residues in generously allowed regions [~a,~b,~l,~p]	16	2.9%	18	3.0%	23	3.4%	13	2.4%	19	1.0%	6	0.30%
Residues in disallowed regions	8	1.4%	11	1.9%	19	2.8%	9	1.6%	17	0.9%	11	0.60%

Table 2: Statistics from Ramachandran plots.

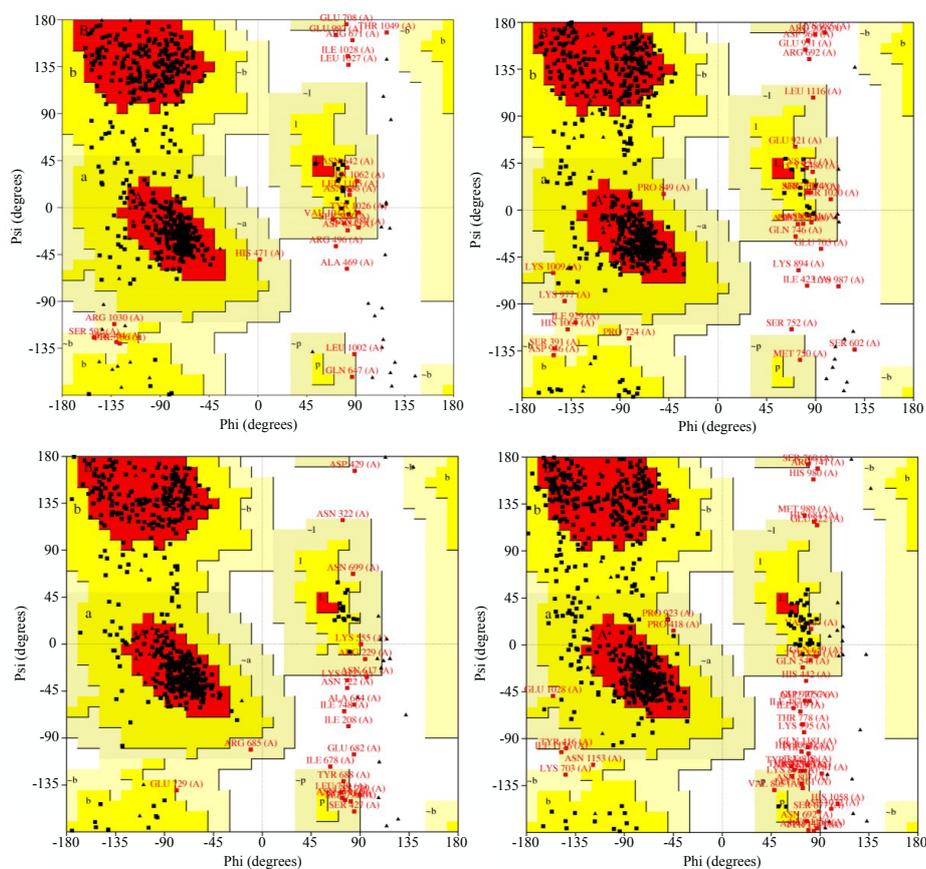


Figure 3: (Top) Ramachandran plots of PvDXS and PkDXS; (Bottom) Ramachandran plots of PbDXS and PfDXS.

is located above domain II and domain III of the same monomer such that the active site is located within the same monomer, in the region between domain I and domain II. However, both of these crystal structures have missing residues in domain I. In *E. coli* DXS, the missing residues are found in not only in domain I (referred to as 'segment I' here) but also in the linker that connects domain I to domain II (referred to as 'segment II' here); the two missing segments were apparently the result of using a fungal protease during crystallization [10]. Residues in segment I of *E. coli* are present close to the active site. The crystal structure of *D. radiodurans* DXS was crystallized without the use of a fungal protease; it also has missing residues (199-243) in domain I in the region that connects strand 4 and 5 (segment I); however, residues present near the active site are ordered in this segment. The missing residues in segment I of *D. radiodurans* DXS (in spite of not using the fungal protease) might indicate that this region in domain I might be

domain swapped such that residue 198 of one monomer is connected to residue 244 of the other monomer.

Structure of *E. coli* Transketolase

One of the closest structural homologues of DXS is *Transketolase* (TK), which catalyzes a similar biomolecular transformation. The crystal structure of TK from *E. coli* is also a homo dimer with three distinct domains. However, the arrangement of domains in TK differs substantially from DXS; domain I of one monomer is located above domain II and domain III of the other monomer such that the active site lies in the dimer interface region. The region that links domain I to domain II is larger in TK consisting of 95 residues. On the other hand, the linker region between domain I and domain II in DXS is comprised of only 20 residues [10]. The difference in domain arrangement between TK and DXS and the missing residues in DXS in the domain

Position	<i>D. radiodurans</i>	<i>E. coli</i>	<i>P. vivax</i>	<i>P. knowlesi</i>	<i>P. falciparum</i>	<i>P. berghei</i>
1	54:SER	52:SER	401:ALA	397:ALA	461:SER	182:PRO
2	80:VAL	78:VAL	469:ALA	423:ILE	487:ILE	208:ILE
3	82:HIS	80:HIS	471:HIS	425:HIS	489:HIS	210:HIS
4	83:GLN	81:GLN	472:SER	426:GLN	490:GLN	211:GLN
5	123:GLY	121:GLY	530:ARG	466:GLY	530:GLY	251:GLY
6	124:HIS	122:HIS	531:GLY	467:HIS	531:HIS	252:HIS
7	125:ALA	123:SER	532:GLY	468:SER	532:SER	253:SER
8	153:GLY	151:GLY	558:GLY	561:GLY	572:GLU	302:GLY
9	154:ASP	152:ASP	559:ASP	562:ASP	573:ARG	303:ASP
10	155:GLY	153:GLY	560:GLY	563:GLY	574:ILE	304:GLY
11	156:SER	154:ALA	561:GLY	564:GLY	575:PHE	305:GLY
12	181:ASN	179:ASN	586:ASN	589:ASN	655:ASN	330:ASN
13	183:ASN	181:ASN	588:ASN	591:ASN	657:ASN	332:ASN
14	185:MET	182:GLU	591:VAL	594:VAL	661:SER	333:ARG
15	186:SER	XXX:XXX ^a	592:SER	595:SER	661:SER	304:GLY
16	187:ILE	XXX:XXX ^a	593:LEU	596:LEU	662:LEU	XXX:XXX ^a
17	289:LYS	284:LYS	681:LYS	702:LYS	751:LYS	426:LYS
18	304:HIS	288:TYR	696:HIS	717:ASP	773:ILE	430:PHE
19	347:PRO	344:PRO	778:PRO	779:PRO	853:PRO	516:PRO
20	348:ALA	345:ALA	779:ALA	780:ALA	854:ALA	517:ALA
21	349:MET	346:MSE	780:MET	781:MET	855:MET	518:MET
22	371:ILE	368:ILE	802:ILE	803:ILE	877:ILE	308:GLY
23	373:GLU	370:GLU	804:GLU	805:GLU	879:GLU	541:GLU
24	398:PHE	395:PHE	830:PHE	831:PHE	905:PHE	567:PHE
25	401:ARG	398:ARG	833:ARG	834:ARG	908:ARG	570:ARG
26	434:HIS	431:HIS	866:HIS	867:HIS	941:HIS	603:HIS

Table 3: Comparison of residues in the active-sites of DXS homology models and templates.

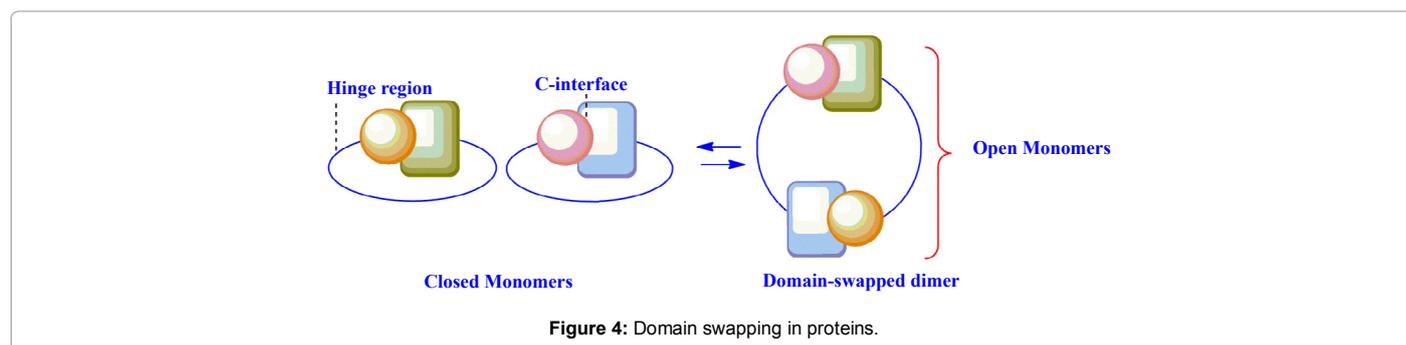


Figure 4: Domain swapping in proteins.

linker region might imply domain swapping in DXS where domain I of one monomer swaps partners with domain II of the other monomer such that the active site is shifted from the domain interface.

Domain swapping

Domain swapping is a phenomenon by which a bond is created between two or more protein molecules as their identical domains are exchanged to form an intertwined dimer or oligomer (Figure 4) [19]. Factors such as salt bridges and pH reveal important information regarding domain swapping [19]. The ability of a protein to domain swap can also be determined from the sequence by analyzing the

sequence derived structure entropies of the residues. Usually domain swapping occurs in a hinge region of the protein, and often proline residues are present in the hinge regions that might increase the rigidity of the loop associated with domain swapping [20]. Domain swapping favors the entangling of polypeptide chains, which might change the native-state properties of the protein, like proteinase-resistance and toxicity [21]. However, domain swapping is often observed at unusually low pH (~4) and not generally at physiological pH [22].

In order to derive information about possible domain swapping in PvDXS, we analyzed the sequence for hinge regions, where swapping could potentially occur. We studied the sequence derived structure

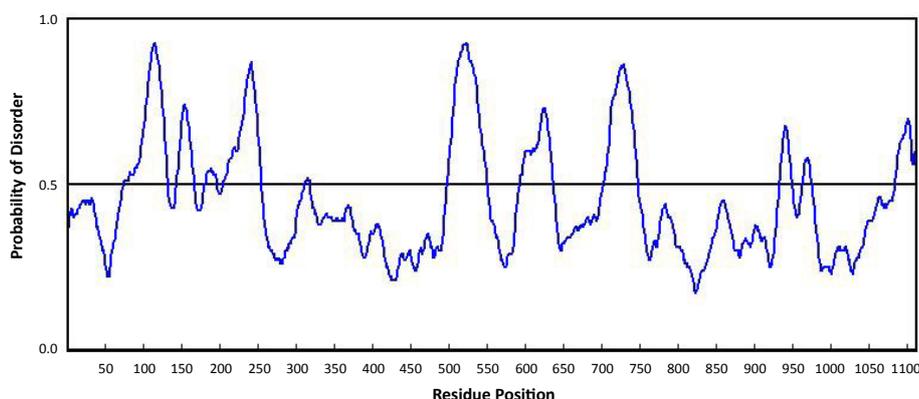


Figure 5: Disorder probability of PvDXS, derived from sequence.



Figure 6: Homology model of PvDXS showing the residues in the predicted region of domain swapping. TPP shown with green carbon atoms.

entropy for PvDXS. The sequence derived structure entropy (SDSE) values for the protein sequences were calculated using tri-peptides and tetra-peptides from different non-redundant databases as the standard using the Molecular Bioinformatics Center server [23]. The hinge regions were detected using DOMPRED [24] and DisEMBL [25], which indicated disorder in the regions: residues 1-250, residues 500-550, 605-650 and around residue 727. The disordered regions based on loops/coils in the secondary structure [26] were checked on the DisEMBL server [25]. Our results indicate that residues 1-250 are highly disordered, which suggest that this part of the protein could be the signal/transit peptide [12]. Likewise, the sequence was checked for the presence of a signal peptide using the signalP server [27], which indicates that residues 1-25 form the signal peptide. Based on the above studies, PvDXS is predicted to have three domains, Domain I consisting of residues 250-582, domain linker region between residues 583-609, domain II consisting of residues 609-724 and domain III consisting of residues 725-1111. We predict that domain swapping could occur in PvDXS between residues 499-528 of domain I, which form the linker between strand 4 and 5 of domain I (Figure 5).

We used the SDSE tool [23] from the Molecular Bioinformatics Center to generate the entropy based on sequence. This was performed to identify the plausible domain swap region. To confirm the region of domain swapping, the proline, histidine and isoleucine residues that

lie in the disordered region of domain I and in the domain I - domain II linker region were mutated to alanine and the sequence derived structure entropy change was calculated; the results of which are shown in (Table 3). There was a huge gain in entropy for the Histidine residue His499 upon mutation to alanine, in the predicted swap region. As expected, the proline residues present in the hinge region had a loss of entropy when mutated to alanine, which suggests that it provide the rigidity associated with domain swap. Based on the results, it is evident that if domain swapping occurs, it most likely occurs in the region 499-594. Similarly, for *P. knowlesi*, the domain swap is predicted to occur between residues 586-612. For *P. falciparum*, the domain swap is predicted to occur between residues 653-678 and for *P. berghei*, between residues 327-353 respectively (Figure 6).

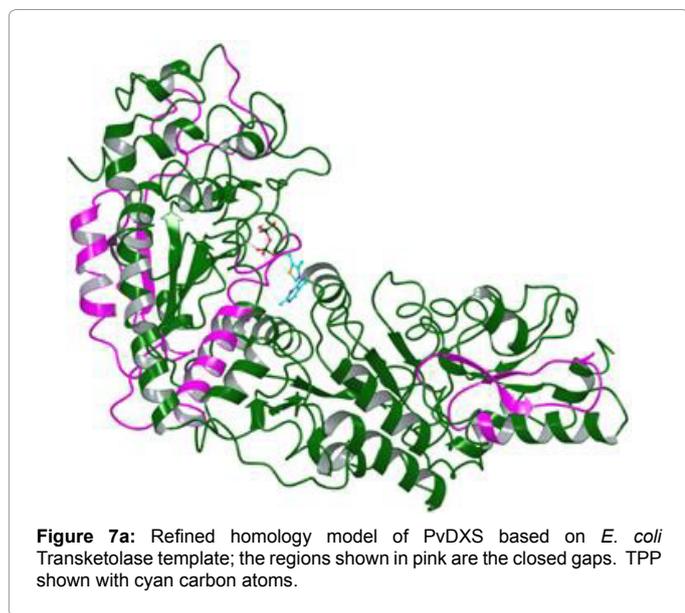
Residues His499 and Pro500 are specific to *P. knowlesi* and *P. vivax* but absent in other *Plasmodium* species. Residue His510 is specific to PvDXS and is absent in all other *Plasmodium* species, indicating that this residue might play an important role as a gatekeeper in domain swapping. Salt bridges reveal important information regarding domain swapping [19]. Salt bridges were analyzed for the dimeric PvDXS structure using VMD [28]. Analysis revealed three salt bridges occurring in segment I of the dimer, connecting chain A and chain B; Asn486-His527, Asn518-His572 and Asn547-Lys706 respectively. The first salt bridge formed between Asn486 of Chain A and His527 of Chain B is interesting since this segment lies in the predicted domain swapped region. This salt bridge could be retained in the domain swapped dimer such that Asn486 of Chain A is connected to His 527 of the same chain after domain I is swapped between the residues 499-528, with Pro528 providing the necessary rigidity associated with domain swap. Similar observations are noted for *P. knowlesi*, *P. berghei* and *P. falciparum* (Table 4).

Structure of domain swapped PvDXS

To determine the structure of domain swapped PvDXS, comparative modeling was performed using *E. coli* Transketolase (PDB: 2R8O) as an additional template (along with *E. coli* DXS and *D. radiodurans* DXS); it had only 9% sequence identity with the sequence of PvDXS. The model was corrected for bond orders, bond angles, protonation states and then energy minimized using Schrodinger's IMPACT module in the Protein Preparation Wizard. Later the prepared protein model was energy minimized employing MacroModel, using the LBFGS algorithm and the OPLS 2005 force field. The refined structure was validated using ProCheck. The refined model after energy minimization is shown in

Residue	Entropy			Mutant	Entropy			ΔS
	H499		0.87006	H499A		1.28247	0.412407	
P500		1.39877		P500A		1.47198	0.073211	
H510		1.38436		H510A		1.34371	-0.04065	
P511		1.52913		P511A		1.40586	-0.12326	
H	527	1	.38635	H	527A	1.45683	0.070476	
	P528		1.4274	P528A		1.43547	0.008071	
P540		1.44851		P540A		1.39533	-0.05318	
I583		1.09052		I583A		1.1942	0.103685	
I584		1.12959		I584A		1.29053	0.16094	
P594		1.27088		P594A		1.45006	0.179181	
P598		1.42308		P598A		1.39244	-0.03064	
P605		1.28438		P605A		1.26844	-0.01594	
I606		1.29282		I606A		1.4603	0.167478	
I609		1.35446		I609A		1.47025	0.115785	
H612		1.55882		H612A		1.40138	-0.15745	
H614		1.39783		H614A		1.24602	-0.15181	

Table 4: Sequence derived structure entropies for residues in the predicted domain swap region of PvDXS.



(Figure 7a). The refined structure had a score of 77.8% for the residues in the most favorable regions, 19.2% for residues in the additionally allowed regions, 1.7% for residues in generously allowed regions and 1.2% for residues in the disallowed regions (Figure 7c). In comparison to the unrefined structure, the residues in the most favorable regions have changed from 81.6% (Figure 7b) to 77.8% in the refined model (Figure 7c), but the residues in disallowed regions have changed from 2.2% (13 residues) to 1.2% (7 residues). Moreover, the residues in the disallowed regions were manually inspected for their vicinity to the active-site. It was found that these residues lie in the regions remote from

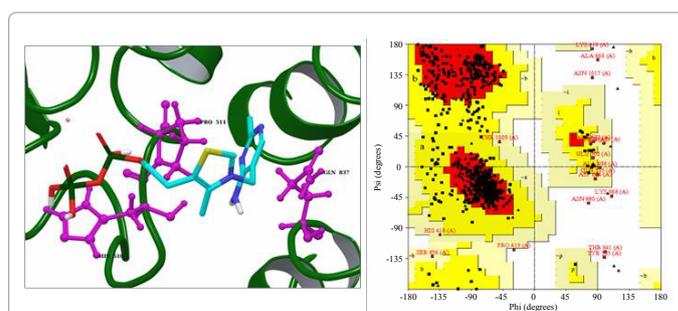
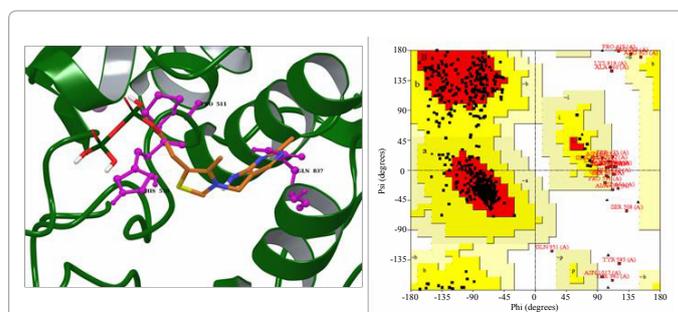
the active-site. The above results suggest that the refined model may be a better one compared to the previous models to perform docking studies. As seen in (Figure 7b,7c) the phenyl ring of TPP is seen to occupy the region close to His510 and Pro511. Note that the side chains of the residues were slightly shifted using the transform tool of Maestro 8.5.2 prior to energy minimization, such that the residues no longer overlap with TPP

Active site identity: For the domain swapped PvDXS structure, the active site residues were compared with *E. coli* TK, *D. radiodurans* DXS and *E. coli* DXS. It was found that most of the residues are identical or similar with the active site of DXS as opposed to Transketolase (TK), as shown in (Table 5).

Docking studies

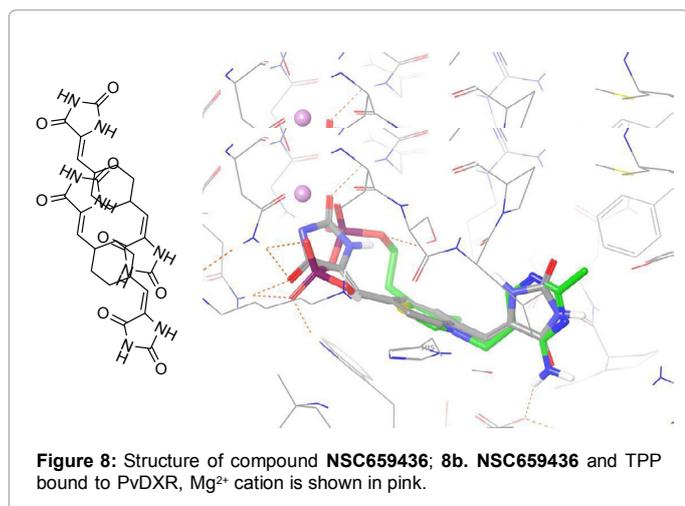
Docking of TPP and the NCI Diversity Set compounds was performed using Schrodinger's GLIDE software in order to validate the homology model(s) and the docked compounds were compared to the estimated absolute binding free energies of TPP and the NCI diversity set compounds docked to *D. radiodurans* DXS (reported here as GLIDE docking scores; only the lowest energy pose is reported).

After docking the NCI Diversity Set compounds, compound **NSC659436** (5Z)-5-[[4-[(E)-(2,5-dioximidazolidin-4-ylidene)methyl]phenyl]methylidene]imidazolidine-2,4-dione was found to occupy putative TPP site in PvDXS and have a very favorable GLIDE docking score of -13.3 kcal/mol. Similarly **NSC659436** when docked to DrDXS had a docking score of -12.9 kcal/mol. TPP docked to PvDXS possessed a docking score of -14.1 kcal/mol and TPP when self-docked to DrDXS (PDB ID: 2O1X), exhibited a docking score of -13.4 kcal/mol. The GLIDE docking scores and poses (Figure 8) of



Position	<i>D. radiodurans</i> DXS	<i>P. vivax</i> domain swapped DXS	<i>E. coli</i> DXS	<i>E. coli</i> TK
1	54:SER	401:ALA	52:SER	119:PRO
2	80:VAL	469:ALA	78:VAL	XXX:XXX ^a
3	82:HIS	471:HIS	80:HIS	121:VAL
4	83:GLN	472:SER	81:GLN	121:VAL
5	109:PHE	516:PRO	107:PHE	9:ASN
6	123:GLY	530:ARG	121:GLY	XXX:XXX ^a
7	124:HIS	531:GLY	122:HIS	XXX:XXX ^a
8	125:ALA	532:GLY	123:SER	XXX:XXX ^a
9	153:GLY	558:GLY	151:GLY	33:ILE
10	154:ASP	559:ASP	152:ASP	32:LEU
11	155:GLY	560:GLY	153:GLY	54:ILE
12	156:SER	561:GLY	154:ALA	50:SER
13	157:LEU	562:LEU	155:ILE	49:SER
14	181:ASN	586:ASN	179:ASN	35:TYR
15	183:ASN	588:ASN	181:ASN	31:LEU
16	185:MET	591:VAL	182:GLU	29:ALA
17	186:SER	592:SER	XXX:XXX ^a	55:SER
18	187:ILE	593:LEU	XXX:XXX ^a	55:SER
19	188:SER	594:PRO	XXX:XXX ^a	56:GLY
20	289:LYS	681:LYS	284:LYS	27:ALA
21	347:PRO	778:PRO	344:PRO	XXX:XXX ^a
22	348:ALA	779:ALA	345:ALA	56:GLY
23	349:MET	780:MET	346:MSE	XXX:XXX ^a
24	371:ILE	802:ILE	368:ILE	53:ALA
25	373:GLU	804:GLU	370:GLU	XXX:XXX ^a
26	395:TYR	827:TYR	392:TYR	XXX:XXX ^a
27	398:PHE	830:PHE	395:PHE	XXX:XXX ^a
28	401:ARG	833:ARG	398:ARG	XXX:XXX ^a
29	434:HIS	866:HIS	431:HIS	XXX:XXX ^a

Table 5: Comparison of residues in the active-sites of DXS homology models and templates.



TPP and NSC659436 docked to PvDXS (vs. DrDXS) suggest that our homology model can serve as a viable and predictive one for virtual screening. On the other hand, the docked poses of TPP and NSC659426 in the domain swapped model yielded docking

scores for TPP and NSC659436 -5.51 and -5.38 kcal/mol respectively suggesting that our domain swapped structure, while interesting, may not be the biologically active form.

Conclusion

In conclusion, we have generated homology models for *P. vivax* DXS and other *Plasmodium* species. We have predicted the probable regions of domain swapping in *P. vivax* DXS and have attempted to build the homology model based on the template of Transketolase for the putative domain swapped *P. vivax* structure. We have used energy minimization to refine our homology models. To validate our homology model, we have docked compounds from the NCI Diversity Set, from which NSC659436 (5Z)-5-[[4-[(E)-(2,5-dioximidazolidin-4-ylidene)methyl]phenyl]methylidene]imidazolidine-2,4-dione has a very favorable glide score of -13.7 kcal/mol. The docked pose of our virtual hit, NSC659436, is similar to the pose of the substrate (TPP), and exhibits interactions with PvDXS and DrDXS that are similar to the interactions observed for TPP.

Acknowledgement

Support from the Graduate Multidisciplinary Scholar (GMS) award by the Florida Center of Excellence for Bio molecular Identification and Targeted therapeutics (FCoE-BITT) to Sumit Handa is highly acknowledged. We thank the Florida Center of Excellence for Biomolecular Identification and Targeted therapeutics for the seed grant, No. GALS018, awarded to Dr. Wayne C. Guida and Dr. David J. Merkler.

References

- Jomaa H, Wiesner J, Sanderbrand S, Altinciecek B, Hintz M et al. (1999) Inhibitors of the Nonmevalonate Pathway of Isoprenoid Biosynthesis as Antimalarial Drugs. *Science* 285: 1573-1576.
- Eisenreich W, Bacher A, Arigoni D (2004) Biosynthesis of isoprenoids via the non-mevalonate pathway. *Cellular and Molecular Life Sciences* 61:1401-1426.
- Oldfield E (2010) Targeting Isoprenoid Biosynthesis for Drug Discovery: Bench to Bedside. *Accounts of Chemical Research* 43:1216-1226.
- Cilingir G, Broschat SL (2012) ApicoAP: The First Computational Model for Identifying Apicomplast-Targeted Proteins in Multiple Species of Apicomplexa. *PLoS One* 7(5): e36598.
- Stuart A Ralph, Giel G van Dooren, Ross F Waller, Michael J Crawford, Martin J Fraunholz, et al. (2004) Metabolic maps and functions of the *Plasmodium falciparum* apicomplast. *Nature Reviews Microbiology* 2: 203-216.
- Louis H Miller, Dror I Baruch, Kevin Marsh (2002) The pathogenic basis of malaria. *Nature* 415: 673-679.
- Kuzuyama T, Takagi M, Takahashi S, Seto H (2000) Cloning and Characterization of 1-Deoxy-D-Xylulose 5- Phosphate Synthase from *Streptomyces* sp. Strain CL190, Which Uses both the Mevalonate and Nonmevalonate Pathways for Isopentenyl Diphosphate Biosynthesis. *J. Bacteriol.* 182: 891-897.
- Lois LM, Campos N, Putra SR, Danielsen K, Rohmer M, et al (1998) Cloning and characterization of a gene from *Escherichia coli* encoding a transketolase-like enzyme that catalyzes the synthesis of D-1-deoxyxylulose 5- phosphate, a common precursor for isoprenoid, thiamin, and pyridoxol biosynthesis. *Proc Natl Acad Sci USA* 95: 2105-2110.
- Thompson JD, Higgins DG (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673-4680.
- Song Xiang, Gerlinde Usunow, Gudrun Lange, Marco Busch, Liang Tong (2007) Crystal Structure of 1-Deoxy-d-xylulose 5-Phosphate Synthase, a Crucial Enzyme for Isoprenoids Biosynthesis. *Journal of Biological Chemistry* 282: 2676-2682.
- Eswar N, Sali A (2009) Protein Structure Modeling in From Molecules to Medicines, J. L. Sussman and P. Spadon, Eds., Springer Netherlandspp.
- Handa S, Ramamoorthy D, Spradling TJ, Guida WC, Adams JH, et al. (2013) Production of recombinant 1-deoxy-d-xylulose-5-phosphate synthase from *Plasmodium vivax* in *Escherichia coli*. *FEBS Open Bio* 3: 124-129.

13. XIANG YUAN YA (1991) A Modified BFGS Algorithm for Unconstrained Optimization. IMA Journal of Numerical Analysis 11: 325-332.
14. Mohamadi F, Richard NGJ, Guida WC, Liskamp R, Lipton M, et al. (1990) MacroModel - an Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. J. Comput. Chem. 11: 440-467.
15. Banks JL, Beard HS, Cao Y, Cho AE, Damm W, et al. (2005) Integrated Modeling Program, Applied Chemical Theory (IMPACT). Journal of Computational Chemistry. 26:1752-80.
16. Jacobson MP, Pincus DL, Rapp CS, Day TJF, Honig B, et al. (2004) A Hierarchical Approach to All-Atom Protein Loop Prediction. Proteins 55: 351-367.
17. Laskowski RA (2001) PDBsum: summaries and analyses of PDB structures. Nucleic Acids Research 29: 221-222.
18. Laskowski RA (1993) PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography 26: 283-291.
19. Kishan KVR, Gulliot SD (2001) Effect of pH and salt bridges on structural assembly: Molecular structures of the monomer and intertwined dimer of the Eps8 SH3 domain. Protein Science 10: 1046-1055.
20. Schymkowitz JW, Wilkinson H, Friedler A (2001) Observation of signal transduction in three-dimensional domain swapping. Nat Struct Mol Biol 8: 888-892.
21. Laurine E, Gregoire C, Fandrich M, Engemann S, Marchal S, et al. (2003) Lithostathine Quadruple-helical Filaments Form Proteinase K-resistant Deposits in Creutzfeldt-Jakob Disease. J Biol Chem 278: 51770-51778.
22. Liu Y, Eisenberg D (2002) 3D domain swapping: As domains continue to swap. Protein Science 11: 1285-1299.
23. Huang SW, Hwang JK (2005) Computation of conformational entropy from protein sequences using the machine-learning method - Application to the study of the relationship between structural conservation and local structural stability. Proteins 59: 802-809.
24. Marsden RL, Jones DT (2002) Rapid protein domain assignment from amino acid sequence using predicted secondary structure. Protein Science 11: 2814-2824.
25. Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, et al. (2003) Protein Disorder Prediction: Implications for Structural Proteomics. Structure 11: 1453-1459.
26. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577-2637.
27. Petersen TN, Henrik N (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8: 785-786.
28. Humphrey W, Dalke A (1996) VMD: Visual molecular dynamics. Journal of Molecular Graphics 14: 33-38.

This article was originally published in a special issue, [Algorithm Evaluation and Validation in Proteomics](#) handled by Editor(s). Dr. Tamanna Sultana, University of Pittsburgh, USA