Journal of Computational Intelligence in Bioinformatics
ISSN0973-385X Volume 7, Number 1 (2014) pp. 19-36

http://www.ripublication.com

In Silico Structural and Functional Analysis of Pyruvate Dehydrogenase from *Brugiapahangi*

Onkar Nath[1], Bharti Singal[1], Indrakant Kumar Singh*[1]

[1]Deshbandhu College University of Delhi, New Delhi-110007, INDIA.
**Corresponding author: indrakant@gmail.com*

Abstract

Brugiapahangiis a nematode that causes a number of serious diseases in humans like *lymphatic*filariasis and onchocerciasis. Structural and functional analysis of the vital genes required for its survival would expand the options for novel drug target to control filariasis and other diseases caused by *B. pahangi*. Thus we have selected a promising target Pyruvate dehydrogenase to study which is an important enzyme that maintains glucose level and energy level (ATP) in the body. Unfortunately not much information about its structure and functional annotation is available. Therefore in the present study, 3D structure of Pyruvate Dehydrogenase was modeled using I-TASSER server, SWISS MODEL, Modeller, Raptor X, Phyre2 and also by CPHmodels 3. 2 Server and validated with PROCHECK and VERIFY 3D. The best model was selected and energy was minimized and used to analyze structure function relationship. Prediction of secondary structure of this enzyme has been performed which can be utilized to understand its stability under natural conditions computationally. This study permits initial inferences about the unexplored 3Dstructure of the Pyruvate Dehydrogenase from *B. pahangi* and may promote in relational designing of molecules to deal with this parasite. Prediction of the active sites and cleavage sites can take this process to a next step.

Keywords:*B. pahangi*, Pyruvate Dehydrogenase, 3D Model, Homology Modelling.

1. Introduction

Several species of natural parasites of wild and domestic mammals are reported to cause chance infection in Human beings. These parasites get into the body of humans via blood sucking arthropods which feed on both animals and humans. Filarial

nematodes are one such kind of parasites. Filarial nematodes like Wuchereriabancrofti, Brugiamalayi and *Brugiapahangi*affect more than 120 million people throughout the tropical and subtropical regions. Due to this reason these organisms are studied extensively and the structural and functional analysis of their proteins becomes crucial. *Brugiapahangi* is a lymphatic filarial parasite of mammals which is closely related to Brugiamalayi (Fig. 1). Aedesaegypti and Armigeresobturbans are two of the vectors of *B. pahangi*. *B. pahangi* is majorly responsible for causing filariasis in domestic cats and dogs which can also be called its reservoir hosts [1]. It has been reported to cause lymphatic filariasis, onchocerciasis in humansand canine heartworm disease in other animals [2]. The organism exhibits anaerobic metabolism which doesn't affect its survival or motility (Wang and Saz, 1974). A relatively low activity of pyruvate dehydrogenase has been evidenced in this nematode with very less amount of pyurate available for oxidation. In addition to that, isocitrate dehydrogenase is noted in very little amount in its mitochondria which shows the activity of homolactete fermenter under in-vitro conditions [3]. It is evident that *B. pahangi* possesses highly cristated mitochondria indicating presence of some amount of aerobic metabolism [4].

Pyruvate dehydrogenase complex (PDC) is reponsible for the conversion of pyruvate into acetyl-CoA by Pyruvate decarboxylation. This Acetyl-CoA produced with the help of PDC may then be used in the Krebs cycle [3] for cellular respiration thus linking the glycolysis to the Krebs cycle and releasing energy via NADH. The blood glucose and ATP levels are maintained by PDC which is regulated by pyruvate dehydrogenase kinase (PDK) isoenzymes [3]. It has been studied that PDC is phosphorylated and inactivated by PDK2 and PDK4 in metabolically active tissues but specific mechanisms behind their transcriptional regulation is not understood clearly [5].

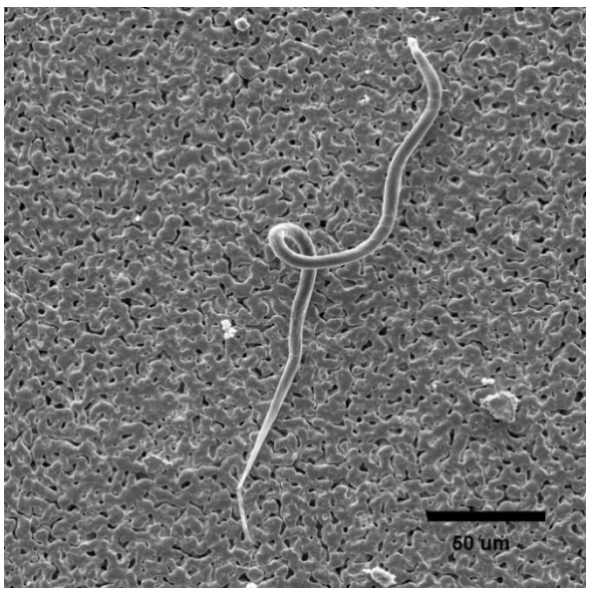

Fig. 1.Filarial worm, *Brugiapahangi*, on a silver membrane filter [59]

Hence, the three dimensional structural and functional study of pyruvate dehydrogenase in *Brugiapahangi* seems important in preventing its infection in mammals. Here, we have used homology modelling based approach to predict models for its 3D structure and comparative analysis for better structure is done using variety of tools. The reported structure can be used further for drug designing.

2. Material and Method

2. 1. Retrieval of target sequences

We have retrieved the protein sequence of pyruvate dehydrogenase of *B. pahangi* from National Center for Biotechnology information (http://www.ncbi.nlm. nih.gov/) (Accession No: ABO84944, Version: ABO84944. 1, GI: 140084473) in FASTA format and used it for further analysis.

2. 2. Template selection and alignment of the target

In homology modelling of 3D structure of the target proteins we followed a sequential order, starting from selection of templates from PDB (protein data bank) related to the target sequence using BLASTp (protein-protein alignment). BLAST (Basic Local Alignment Search Tool) was carried out against PDB database available at NCBI (National Centre for Biotechnology Information) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to select an appropriate template.

2. 3. Construction of the models

Homology modelling was performed for pyruvate dehydrogenase protein sequence using MODELLER 9. 10 tool by Sali Lab [6][7][8][9]. The PDB sequence 2OZLB had high resolution of 1. 9 thus, selected as the best template. The structure modelling module was modified to generate 10 models. Second model generated is the best predicted structure as it had lowest dope score and highest GA341 score (Fig. 2(a)). In the first step of model building, distance, bond length, bond angle, H bond potential and dihedral angle restraints on the target sequence were derived from its alignment with the template 3D-structure

Swiss model server (http://swissmodel.expasy.org/) is a fully automated user friendly web server [10]. In the alignment mode of SWISS MODEL, a structurally known protein chain from PDB library was taken as template. The server builds a model based on selected alignment of target and template (Fig. 2(b)).

I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) is an on-line platform for protein structure and function prediction [11][12][13]. The most accurate structural and function predictions are made using state-of-the-art algorithms using Multiple-threading alignments (Fig. 2(c)).

RaptorX (http://raptorx.uchicago.edu/StructurePrediction/) is a web based server that predicts 3-state and 8-state secondary structure, solvent accessibility, disordered regions and tertiary structure for a protein sequence [14]. Raptor X excels at predicting 3D structures for protein sequences in the absence of close homologs in the PDB sequences (Fig 2(d)).

Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/) is a Protein Homology/analogy Recognition Engine [15]. Phyre is typical of many structure prediction systems, it can reliably detect up to twice as many remote homologies as standard sequence-profile searching (Fig. 2(e)).

CPHmodels 3.2 Server (http://www.cbs.dtu.dk/services/CPHmodels-3.2/) is a web based tool to predict 3D structure of protein based on profile-profile alignment [16][17]. The server predicts using scoring functions of CPHmodels-2.0 and a novel remote homology-modelling algorithm (Fig. 2(f)).

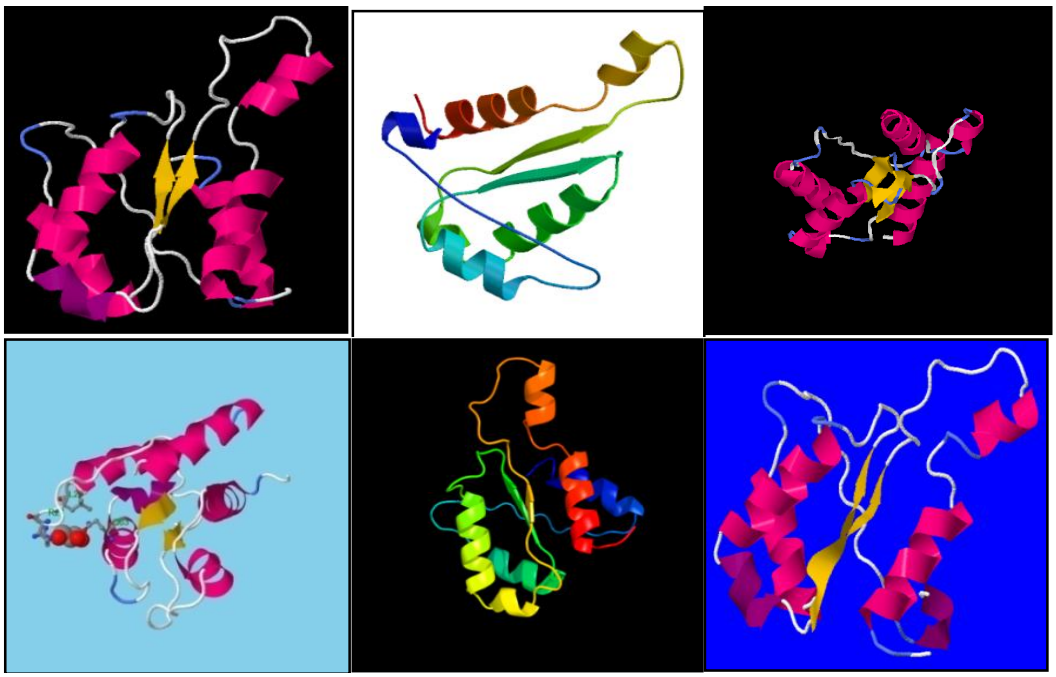

Fig.2. (a) Model build by Modeller (b) Model build by SWISS MODEL (c) Model build by I-TASSER (d) Model build by Raptor X (e) Model build by Phyre2 (f) Model build by CPHmodels 3.2 Server

2. 4. Evaluation of the constructed model

3dSS (3-Dimensional Structural Superposition) (http://cluster.physics.iisc.ernet.in/3dss/) is a web-based server that aids to superpose two or several 3D protein structures [18]. All the modelled protein 3D-structures were submitted in 3dSS server. This server superimposed the structures on the basis of their similarity (Fig. 3). VADAR (Volume, Area, Dihedral Angle Reporter) (http://vadar.wishartlab.com/) [20] is an online tool that predicts the reliability of a protein..On submission of our predicted protein it provided us with different analysis. Ramachandran plot is a plot originally developed in 1963 by G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, [19] is a way to visualize backbone dihedral angles ψ against φ of amino acid residues in protein structure. The red, yellow and green regions represent the favoured, allowed, and "generously allowed" regions as defined by VADAR [20]. SAVES server (http://nihserver.mbi.ucla.edu/SAVS/) was used for verification of the

model. ERRAT produces the evaluation of overall qualities of the protein sequence [21] whereas Q-Flipper lists the flipper scores for all asparagine and glutamine residues [22]. The protein model was evaluated using ProSA (https://prosa.services.came.sbg.ac.at/prosa.php) [23] to assess accuracy and reliability of the modelled structures.

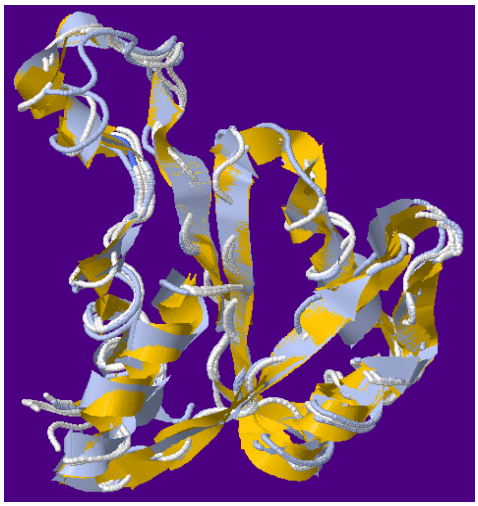

Fig. 3.Superposition of structures by 3dSS.

Verify 3D tool is used for the improvisation of the putative 3D structures by taking its own amino acid sequence (1D) into consideration [24].

2. 5. Studying intrinsic dynamics of the protein models and visualization of the modelled protein

Biological function is governed by the structural dynamics of the protein [25]. WEBnm (http://apps.cbu.uib.no/webnma) tool [26] was used for structural dynamics studies by calculating the slowest modes and deformation energies. elNemo (http://igs-server.cnrsmrs.fr/elnemo/index.html) was used to calculate the proteins contributing to the corresponding protein movement [27]. Structure visualization of the protein was performed using Jmol [28] and SPDBV [29].

2. 6. Protein Quality Prediction

ProQ(http://www.sbc.su.se/~bjornw/ProQ/ProQ.cgi) is a neural network based predictor which is used to predict the quality of the structure [30]. ProQ is optimized to find correct models by two quality measures LG score and MaxSub. LG score is a negative logarithm of P value and MaxSub ranges from 0-1, were 0 is for insignificant and 1 is for very significant [30].

2. 7. Other predictions for overall function and analysis of impacts on biological system

Biochemical Pathway Maps (http://web.expasy.org/pathways/) was used to predict the pathways associated with the protein [31].

PSLpred (www.imtech.res.in/raghava/pslpred/*)* was used to predict the subcellular localization of the protein [32]. Peptide cutter(http://web.expasy.org/peptide_cutter/) software by EXPASY was used to find the peptides that can cleave the sequence [33]. Tools like InterProSurf, PSLpred, Isotopident, InterPro, NetAcet, PRED-TMBB etc. were used to predict the physiochemical properties and other functions related to the protein.

3. Results and Discussion

3. 1. Retrieval of target sequences

Protein sequence of pyruvate dehydrogenase of *B. pahangi* was retrieved from NCBI by using its accession number. The sequence was retrieved in FASTA format.

3. 2. Template selection and alignment of the target

BLAST search found the crystal structure of the Pyruvate Dehydrogenase S264E variant (2OZL) to be the best template with resolution 1.9 angstroms. The “B” chain of the sequence showed best similarity with the sequence.

3. 3. Construction of the model

The modelling of the 3D structure of Pyruvate Dehydogenase protein was performed using six modelling tools namely Modeller9.10, SWISS MODEL, I-TASSER, Raptor X, Phyre2 and CPHmodels 3.2 Server.

In Modeller, potential structures (PDB-ID: 1NI4, 1OLX, 1W85, 2OZL, 3DUF, 3EXE), were taken for model building out of number of hits from PDB. The templates were first aligned with each other using salign script. Using this alignment the best template and target were aligned using align2d script. On the basis of this alignment, ten comparative models of the target (pyruvate dehydrogenase) were generated by MODELLER, applying the model_multi script. The best model was decided using DOPE score and GA341 score. The model with the lowest DOPE score is expected to be the best model.

In SWISS MODEL the sequence was uploaded on the server to search the templates. The best templates were selected from the alignment and resubmitted for full 3D modelling. The models having highest GMQE and QMEAN4 values are predicted as the best models.

I-TASSER takes protein sequence in FASTA format as input. The C-score (confidence Score) of the best predicted model was 1.27. The accuracy of the model is analyzed by TM-score and RMSD value. The estimated TM-value is 0.89±0.07 and RMSD 1.9±1.5Å. It also finds the template proteins with similar binding sites. The PDB hits with similar binding sites are 3exfD and 1w85D. Similar binding site means similar response to the molecules that interact with that site.

The sequence was submitted to Raptor X input field for analysis. The result contained the 3D structure of the protein as well as analysis of the binding site. Two binding domains were predicted and the ligands GOL and EDO were expected to bind to the predicted pockets. The secondary structure was predicted to contain 53% helix, 11% Beta-sheet and 35% coils. The solvent accessed is expected to contain 40%

Buried, 29% Medium and 29% Exposed regions.

The Structure modelled by Phyre2 has a high confidence level of (>90%) and it modelled 99% of the residues. This server also predicts that the sequence contains 47% alpha helix and 13% beta strand.

The sequence was submitted to CPHmodel server in FASTA format. The best template considered was 2OZL-B with query coverage of 90.4. The E-value of the model was 4e-25.

3. 4. Secondary Structure analysis

The CFSSP (Chou &Fasman Secondary Structure Prediction Server) [34][35] predicts that the sequence contains 89 residues that may be part of helix, similarly 74 residues for sheet and 14 residues for turn. The sequence contains 35.65% alpha helix, 19.13% Extended strand and 45. 22% Random coil as estimated by GOR4 tool [36].

3. 5. Evaluation of refined models

3. 5. 1. ProSA

ProSA is a tool that is used regularly to detect possible errors in 3D models of protein structures [23]. The Z score of Pyruvate Dehydrogenase protein made by MODELLER was-4.67, for SWISS MODEL model-4.11,-4.53 for I-Tasser model,-4.79 for Raptor X model,-4.55 for Phyre2 model and-3. 96 for CPHModel model (Fig. 4(a)). The web plot of residue scores displayed local model quality using plotting energy as a function of amino acid sequence location [23]. The models were plotted in the allowed region, indication that all the models prepared by different programs predicted a reliable model. The model prepared by Raptor X has the lowest Z-Score indicating that this model is the most reliable model out of all the models. The energy plot is showing that most of the regions have high negative values for Pyruvate Dehydrogenase protein (Fig. 4(b)). Thus we interpreted; all models were good and reliable.

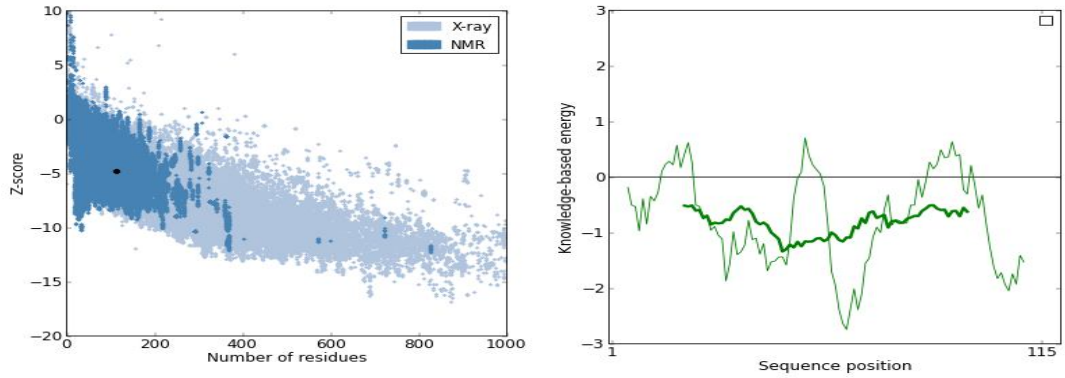

Fig.4. (a) Z-Score plot by ProSA (b) energy plot by ProSA

3. 5. 2. ERRAT for validating 3D model

ERRAT is a protein structure verification algorithm that is suitable for estimating the crystallographic model [25]. The program worked by analyzing the statistics of non-

bonded interactions between different atom types [37]. ERRAT analysis revealed the overall quality factor of pyruvate dehydrogenase protein from *B. pahangi* was 74.766 for MODELLER, 93.258 for SWISS MODEL model, 74.766 for I-Tasser model, 81.308 for Raptor X model, 73.832 for Phyre2 model and 98.936 for CPHModel model (Fig.5). These results inferred that the overall quality of pyruvate dehydrogenase protein model made by CPHModel was very good in comparison to the other models which were fairly good.

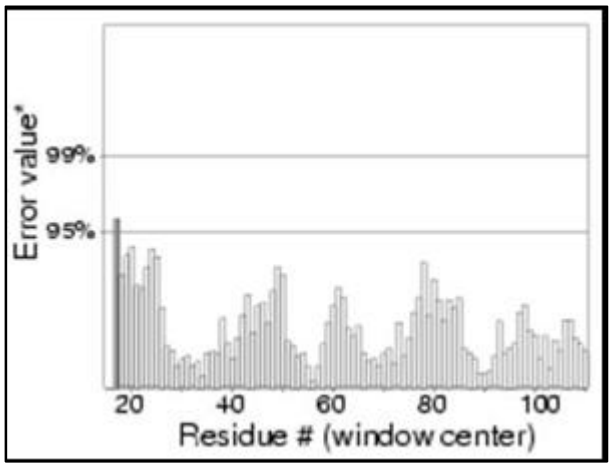

Fig. 5.Plot by ERRAT for validating 3D model from Raptor X

3. 5. 3. VERIFY 3D for validating 3D model

VERIFY3D was used to validate the predicted models. Verify3D analyzes the compatibility of an atomic model (3D) with its own amino acid sequence (1D) [24]. VERIFY 3D reveals that model predicted by Raptor X had 3D-1D score >-0.03, >-0.09 for modeller model, >-0.07 for SWISS MODEL model, >-0.10 for I-Tasser model, >-0.11 for Phyre2 model and >0.08 for CPHModel model.

3. 5. 4. Q-Flipper for validating 3D model

Q-Flipper is computed using the complete crystal symmetry, but do not include interactions with non-protein atoms (nucleic acids, ions, cofactors, etc.) [22].The structure predicted by MODELLER contains 3(60%) flips, similarly 1(20%) for model SWISS MODEL, 1(20%) for I-Tasser model, 3(60%) for Raptor X model, 1(20%) for Phyre2 model and 0(0%) for model CPHModel.

3. 5. 5. VADAR for 3D model validation

VADAR tool is used for the analysis and assessment of peptide and protein structures. The output page contains plots of Ramachandran plot (Fig. 6), Functional Accessibility Surface Area, Fractional Residue volume, Stereo/Packing Quality Index, 3D Profile Quality Index and output files of Main-Chain Table, Side-Chain Table, H-Bond Table and Statistics. The analysis showed that the sequence contains approximately 46-57 residues as helix, 12-16 residues as sheet, 42-46 residues as coil and 8-16 residues in turn regions.

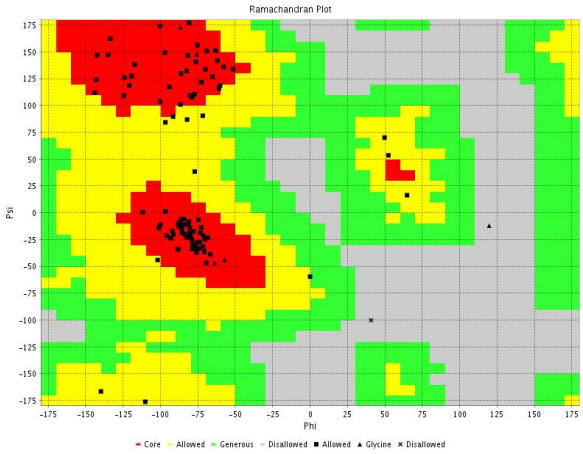

Fig.6. Ramachandran Plot of model prepared by Raptor X produced by VADAR Server

3. 6. Protein Quality Prediction

ProQ's LG score predicts the quality of the predicted model. LG score of pyruvate dehydrogenase was 3. 336 and thus can be considered as good model. The model validated by ProQ is shown in (Fig. 7).

Fig.7. 3D structure of pyruvate dehydrogenase of *B. pahangi*.

3. 7. Functional annotation and physiochemical characterization

InterProSurf is designed to predict the most likely sites on proteins which interact with other proteins, such as toxin elements, cell receptors and other proteins that make up virus capsids [38]. According to InterProSurf results the sequence has Probe radius: 1.400, POLAR area/energy = 2673.65, APOLAR area/energy = 4917.60, Total area/energy = 7591. 25, Number of surface atoms = 602 and Number of buried atoms

= 273. The predicted subcellular localization is Cytoplasmic as predicted by PSLpred using Hybrid approach of prediction. The prediction had an expected accuracy of ~98.1% [32]. Pathways in which this molecule appears are pyruvate dehydrogenase (cytochrome) (enzyme: 1.2.2.2), pyruvate dehydrogenase (lipoamide) (enzyme: 1.2.4.1), pyruvate dehydrogenase (lipoamide) kinase (enzyme: 2.7.1.99), and pyruvate dehydrogenase (lipoamide) phosphatase (enzyme: 3.1.3.43) as predicted by Biochemical Pathway Maps [31]. The ENZYME number corresponds to the data from ENZYME Data Bank. The computed PI (isoelectric point)/Mw (molecular weight) is 6.99/12618.67 as theoretically analyzed by Compute pi/Mw tool [33][55][56]. The sequence does not contain any O-glycosylated region as predicted by DictyOGlyc1.1 Server [39]. The Metazoa function of GPI Prediction server predicts that the sequence do not contain any GPI-modification site [40]. The GPMAW lite tools predicts the average mass of the sequence as 12,616.69 Dalton, Monoisotopic mass (12,608.64 Dalton), Molar absorbance (0.66 cm^-1), Molar extinction coefficient. At 280nm (8,370 cm^-1 M^-1), Isoelectric point 7. 45 and Hydrophobicity index as (-0.04) (Fig.8) [41]. HeliQuest tool plots the Hydrophobicity and Hydrophobic moment of the protein sequence [42].

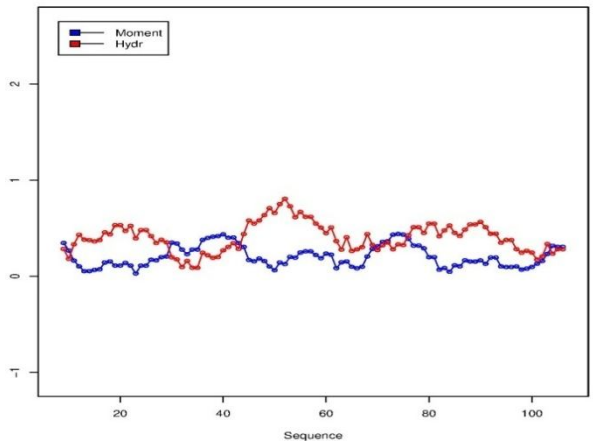

Fig.8. Hydrophobicity and Hydrophobic moment of the protein sequence as plotted by HeliQuest.

HMMTOP analyzes the Total entropy of the model to be 17.0115 [43][44]. The mass of the sequence as calculated by Isotopic calculation tool Isotopident (http://education.expasy.org/student_projects/isotopident/htdocs/) is 12778.699 amu and the elemental mass are C_{566}:6798.056, H_{922}:929. 321, N_{154}:2157.032, O_{171}:2735.897 and S_5 : 160.325. InterPro (http://www.ebi.ac.uk/interpro/) [45][46] predicts the sequence to contains two domain. The first domain Transketolase, C-terminal/Pyruvate-ferredoxinoxidoreductase, domain II (IPR009014) covers from 4th to 113th amino acid. This domain can be found in a number of different enzymes, including the C-terminal domain of the pyruvate dehydrogenase E1 component [PMID: 11955070], the C-terminal domain of branched-chain alpha-keto acid dehydrogenases [PMID: 10426958], and domain II of pyruvate-

ferredoxinoxidoreductase (PFOR)[PMID: 11752578]. This domain contains catalytic activity and also participates in metabolic process. Another domain Transketolase, C-terminal (IPR005476) is from 9th to 98th amino acid. It catalyzes the reversible transfer of a two-carbon ketol unit from xylulose 5-phosphate to an aldose receptor, such as ribose 5-phosphate, to form sedoheptulose 7-phosphate and glyceraldehyde 3-phosphate. This enzyme, together with transaldolase, provides a link between the glycolytic and pentose-phosphate pathways. Transketolase requires thiamine pyrophosphate as a cofactor. TK sequences from a variety of eukaryotic and prokaryotic sources [PMID: 1567394, PMID: 1737042] show that the enzyme has been evolutionarily conserved. 1-deoxyxylulose-5-phosphate synthase (DXP synthase) [PMID: 9371765] is an enzyme so far found in bacteria (gene dxs) and plants (gene CLA1) which catalyzes the thiamine pyrophosphoate-dependent acyloin condensation reaction between carbon atoms 2 and 3 of pyruvate and glyceraldehyde 3-phosphate to yield 1-deoxy-D-xylulose-5-phosphate (dxp), a precursor in the biosynthetic pathway to isoprenoids, thiamine (vitamin B1), and pyridoxol (vitamin B6). It participates in metabolic process and catalytic activity. Myristoylation is an irreversible, protein lipidation modification where a myristoyl group, derived from myristic acid, is covalently attached by an amide bond to the alpha-amino group of an N-terminal glycine residue [47]. NMT tool (http://mendel.imp.ac.at/myristate/SUPLpredictor.htm) [48] predicts the sequence not to contain any potential Myristoylation site. NetAcet(http://www.cbs.dtu.dk/services/NetAcet/) analyzes that the sequence do not contain any Acetylation site [49]. NetGlycate (http://www.cbs.dtu.dk/services/NetGlycate/) [50] predicts the sequence to contains three Glycation sites at position 10, 106 and 109, thus the sequence contains typically covalent bonding of a protein or lipid molecule with a sugar molecule, such as fructose or glucose, without the controlling action of an enzyme. The sequence contains seven predicted kinase site-site (T-8, S-37, T-41, S-67, S-67, S-90, S-90), Kinase (PKC, PKC, PKC, CKII, cdc2, CLII, cdc2) and Score (0.57, 0.93, 0.87, 0.54, 0.51, 0.54, 0.53) respectively as predicted by GPS2.1 server (http://gps.biocuckoo.org/) [51]. The predicted site at position 37 has the highest score, thus can be considered as best predicted site. PRED-TMBB (http://biophysics.biol.uoa.gr/PRED-TMBB/) [52] predicts the sequence to contain beta-barrel outer membrane protein. The triple-state (Residue positions) in which a residue can be are: [in] for periplasmic space (1-11 and 52-115), [tm] for transmembrane strand (12-20 and 41-51) and [out] for extracellular space (21-40). The protein is expected to belong to EC 2.4: Transferases–Glycosyltransferases, Zinc-binding, All lipid-binding proteins, or EC 4.1: Lyases-Carbon-Carbon Lyases families, with decreasing probability in the order. The most probable family is Transferases-Glycosyltransferases as predicted by SVMProt (http://jing.cz3. nus.edu.sg/cgi-bin/svmprot.cgi) [53].

3. 6. Active site analysis

According to APD2 (http://aps.unmc.edu/AP/prediction/prediction_main.php) (Antimicrobial Peptide Predictor) [54], the molecular weight is 12582.694 dal, total hydrophobic ratio is 42% and Protein-binding Potential (Boman index) is: 1.28 kcal/mol. The estimated half-life is: 100 hours (mammalian reticulocytes, in vitro)>20

hours (yeast, in vivo)>10 hours (Escherichia coli, in vivo). The expected PI is 6.99 as predicted using Protparam (http://web.expasy.org/protparam/) [33]. The enzymes that can cleave this protein are Arg-C proteinase, Asp-N endopeptidase + N-termialGlu, BNPS-Skatole, CNBr, Chymotrypsin-high specificity (C-term to [FYW], not before P), Chymotrypsin-low specificity (C-term to [FYWML]not before P), Clostripain, Formic acid, Glutamylendopeptidase, Iodosobenzoic acid, LysC, LysN, NTCB (2-nitro-5-thiocyanobenzoic acid), Pepsin (pH1.3), Pepsin (pH>2), Proline-endopeptidase, Proteinase K, Staphylococcal peptidase I, Thermolysin, Trypsin according to peptide cutter software (http://web.expasy.org/peptide_cutter/) by EXPASY (Table.1) [33]. Pocket-Finder (http://www.bioinformatics.leeds.ac.uk/pocketfinder) [57] is a pocket detection algorithm that works by scanning a probe of radius 1.6 angstroms along all gridlines of grid resolution 0.9 angstroms surrounding the protein (Fig.9). The probe also scans cubic diagonals [58].

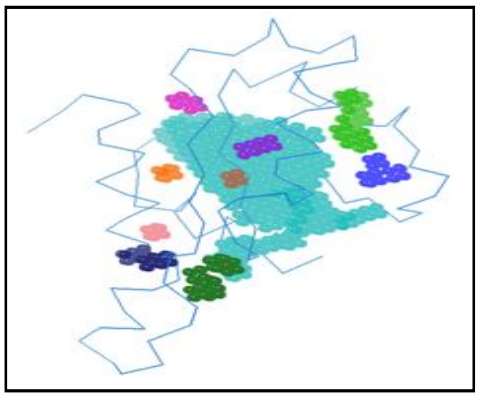

Fig.9. Active sites predicted by Pocket Finder (http://www.modelling.leeds.ac.uk/cgi-bin/pocketfinder/pfmage.cgi) contains 10 active sites.

Table.1. List of enzymes, number of cleavage sites and positions of cleavage sites analyzed by PeptideCutter

Name of enzyme	No. Of cleavages	Positions of cleavage sites
Arg-C proteinase	5	2 24 27 43 79
Asp-N endopeptidase	7	5 29 67 70 73 83 101
Asp-N endopeptidase + N-terminal Glu	15	5 8 18 29 31 47 58 65 67 70 73 83 90 93 101
BNPS-Skatole	1	51
CNBr	3	64 87 98
Chymotrypsin-high specificity (C-term to [FYW], not before P)	5	5 53 70 89 96

Chymotrypsin-low specificity (C-term to [FYWML], not before P)	18	5 11 14 17 20 23 26 29 42 44 53 64 70 73 89 96 103 112
Clostripain	5	2 24 27 43 79
Formic acid	7	6 30 68 71 74 84 102
Glutamylendopeptidase	8	9 19 32 48 59 66 91 94
Iodosobenzoic acid	1	51
LysC	10	10 13 35 36 39 40 100 106 109 110
LysN	10	9 12 34 35 38 39 99 105 108 109
NTCB (2-nitro-5-thiocyanobenzoic acid)	2	32 53
Pepsin (pH1. 3)	15	5 10 11 13 14 20 22 23 25 28 52 69 70 72 73
Pepsin (pH>2)	19	5 10 11 13 14 20 22 23 25 28 51 52 69 70 72 73 88 95 96
Proline-endopeptidase	1	28
Proteinase K	59	1 4 5 8 9 11 12 14 16 18 19 20 21 23 25 26 29 32 34 38 41 44 45 46 47 48 49 51 53 56 58 59 60 62 65 66 69 70 73 75 77 80 81 83 85 89 91 92 93 94 95 96 104 105 107 108 111 112 114
Staphylococcal peptidase I	8	9 19 32 48 59 66 91 94
Thermolysin	35	3 4 10 11 13 15 17 20 22 25 28 33 37 43 44 46 52 55 57 61 63 64 69 72 76 79 82 92 103 104 106 107 110 111 113
Trypsin	14	2 10 13 24 35 36 39 40 43 79 100 106 109 110

4. Conclusion

The 3D structure of the protein plays a very important role in molecular functions analysis. Structural analysis of pyruvate dehydrogenase in the *B. pahangi* was done using six molecular modelling programs. Secondary structure analysis was performed using CFSSP and GOR4. Ramachandran plot, Z-Score plot, Energy plot, minimized energy value, ERRAT, verify 3D, Q-Flipper, proSA results revealed the rigidity and quality of the model. ProQ was used to optimize the protein model and to find best model. Different other tools were used to predict the basic properties like polarity, hydrophobicity, energy, surface area etc. Some others were used to find the domain related properties and restriction sites. Active site was found using Pocket Finder. These sites help in drug discovery process indicating the binding positions. Different peptides were tried for cleavage site and cleaving enzyme detection. The 3D-structure of the protein needs to be studied in the laboratory conditions for further structural information and confirmation of the predicted model. Further conformational and computational analysis of pyruvate dehydrogenase, need to be studied to understand its functional mechanisms.

5. Acknowledgement
We would like to gratefully acknowledge the help of Dr. Ajay K. Arora, The Principal, Deshbandhu College, University of Delhi, Kalkaji, NewDelhi-110007 for providing the computational facility and infrastructure required for the work. The whole work was a fruitful result of efforts made in Molecular Biology Research Lab of Deshbandhu College, University of Delhi.

6. References

[1] Muslim A, Fong MY, Mahmud R, Lau YL and Sivanandam S., 2013, "Armigeressubalbatus incriminated as a vector of zoonotic *Brugiapahangi*filariasis in suburban Kuala Lumpur, Peninsular Malaysia." Parasites & Vectors 6, pp. 219.

[2] Chirgwin SR, Coleman SU, Porthouse KH, Nowling JM, Punkosdy GA, Klei TR., 2003, "Removal of Wolbachia from *Brugiapahangi* is closely linked to worm death and fecundity but does not result in altered lymphatic lesion formation in Mongolian gerbils (Merionesunguiculatus)" Infect Immun. 71, pp. 6986–6994.

[3] Kumar S, Nath O, Govil S, Pathak AN., 2014, " Computational 3D Structure Prediction, Evaluation and Analysis of Pyruvate Dehydrogenase an Effective Target for Filarial Infection by *Brugiapahangi* Using Homology Modeling Approach" International JPSDR. 6(2), pp. 120-123.

[4] Middleton KR, Saz HJ., 1979, "Comparative utilization of Pyruvate by BrugiaPahangi, DipetalonemaViteae, and LitomosoidesCarinii" The Journal of Parasitology. 65(1), pp. 1-7.

[5] Jeong JY, Jeoung NH, Park KG, Lee IK., 2012, "Transcriptional regulation of pyruvate dehydrogenase kinase" Diabetes & Metabolism Journal. 36(5), pp. 328-335.

[6] Eswar N, Marti-Renom MA, Webb B, Madhusudhan MS, Eramian D, Shen M, Pieper U, Sali A., 2006, "Comparative Protein Structure Modeling With MODELLER" Current Protocols in Bioinformatics. John Wiley & Sons, Inc. Supplement 15, pp. 5. 6. 1-5. 6. 30.

[7] Marti-Renom MA, Stuart A, Fiser A, Sánchez R, Melo F, Sali A., 2000, "Comparative protein structure modeling of genes and genomes" Annu. Rev. Biophys. Biomol. Struct. 29, pp. 291-325.

[8] SaliA& Blundell TL., 1993, "Comparative protein modelling by satisfaction of spatial restraints" J. Mol. Biol. 234, pp. 779-815.

[9] Fiser A, Do RK, &Sali A., 2000, "Modeling of loops in protein structures" Protein Science. 9, pp. 1753-1773.

[10] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T., 2014, "SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information" Nucleic Acid Res.

[11] Zhang Y., 008, "I-TASSER server for protein 3D structure prediction" BMC Bioinformatics. 9, pp. 40.
[12] Roy A, Kucukural A, Zhang Y., 2010, "I-TASSER: a unified platform for automated protein structure and function prediction" Nature Protocols. 5, pp. 725-738.
[13] Roy A, Yang J, Zhang Y., 2012, "COFACTOR: an accurate comparative algorithm for structure-based protein function annotation" Nucleic Acids Research. 40, pp. W471-W477.
[14] Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H &Xu J., 2012, " Template-based protein structure modeling using the RaptorX web server, Nature Protocols. 7, pp. 1511–1522.
[15] Kelley LA and Sternberg MJE., 2009, "Protein structure prediction on the web: a case study using the Phyre server" Nature Protocols. 4, pp. 363-371.
[16] Nielsen M., Lundegaard C., Lund O., Petersen TN., 2010, "CPHmodels-3. 0-Remote homology modeling using structure guided sequence profiles" Nucleic Acids Research. 38, pp. W576-81.
[17] Lund O, Nielsen M, Lundegaard C, Worning P., 2002, "CPHmodels 2. 0: X3M a Computer Program to Extract 3D Models" Abstract at the CASP5 conference. December 1-5; Asilomar, CA, USA. p. A102.
[18] Sumathi K, Ananthalakshmi P, Roshan MNA Md., and Sekar K., 2006, "3dSS: 3D structural superposition" Nucl. Acids Res. 34(2), pp. W128-W132.
[19] Ramachandran GN, Ramakrishnan C and Sasisekharan V., 1963, "Stereochemistry of polypeptide chain configurations" Journal of Molecular Biology. 7, pp. 95–99.
[20] Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, and Wishart DS., 2003, "VADAR: a web server for quantitative evaluation of protein structure quality" Nucleic Acids Res. 31 (13), pp. 3316. 3319.
[21] Colovos C and Yeates TO., 1993, "Verification of protein structures: patterns of non bonded atomic interactions"Protein Science. 2(9), pp. 1511-1519.
[22] Weichenberger CX and SipplMJ.,2006, "NQ-Flipper: Validation and Correction of Asparagine/Glutamine Amide Rotamers in Protein Crystal Structures" Bioinformatics. 2006;22, pp. 1397-1398.
[23] Wiederstein&Sippl., 2007, "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins" Nucleic Acids Research. 35, pp. W407-W410.
[24] Eisenberg D, Luthy R and Bowie JU., 1997, "VERIFY3D: assessment of protein models with three-dimensional profiles" Methods in Enzymology. 277, pp. 396–404.
[25] Yang LW, Rader AJ, Liu X, Jursa CJ, Chen SC, Karimi HA and Bahar I., 2006, "oGNM: online computation of structural dynamics using the Gaussian Network Model" Nucleic Acids Research. 34, pp. 24–31.
[26] Hollup SM, Salensminde G and Reuter N., 2005, "WEBnm@: a web application for normal mode analysis of proteins" BioMed Central Bioinformatics. 6, pp. 1–8.

[27] Suhre K and Sanejouand YH., 2004, “ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement” Nucleic Acids Research. 32, pp. 610–614.
[28] Jmol: an open-source Java viewer for chemical structures in 3D. http://www. jmol. org/.
[29] Kaplan W and Littlejohn TG.,2001, Swiss-PDB Viewer (Deep View) Brief Bioinform. 2, pp. 195-197.
[30] Wallner B &Elofsson A., 2003, “Can correct protein models be identified” Protein Science. 12, pp. 1073-1086.
[31] (http://web. expasy. org/pathways/) digitized version of the Roche "Biochemical Pathways" wall chart.
[32] Bhasin M, Garg A and Raghava GPS., 2005, “PSLpred: prediction of subcellular localization of bacterial proteins” Bioinformatics. 21(10), pp. 2522-4.
[33] Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A., 2005, “Protein Identification and Analysis Tools on the ExPASy Server” (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. pp. 571-607.
[34] Chou PY and Fasman GD., 1974, “Prediction of protein conformation. Biochemistry”13(2), pp. 222–245.
[35] Chou PY and Fasman GD., 1974, “Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins” Biochemistry. 13(2), pp. 211-222.
[36] Garnier J, Gibrat JF, Robson B., 1996, “GOR secondary structure prediction method version IV”Methods in Enzymology. R. F. Doolittle Ed. 266, pp. 540-553.
[37] Koteswara RG, Kaleswara RM, Nagamalleswara RK and Gyana RS., 2010, “Comparative modeling of Undecaprenyl pyrophosphate phosphatase in Clostridium botulinum-a potent target of Botulism” International Journal of Systems Biology. 2(2), pp. 10-15.
[38] Li BW, Rush AC, Mitreva M, Yin Y, Spiro D, Ghedin E, Weil GJ., 2009, “Transcriptomes and pathways associated with infectivity, survival and immunogenicity inBrugia malayiL3” BMC Genomics 10, pp. 267.
[39] Gupta R, Jung E, Gooley AA, Williams KL, Brunak S, Hansen., 1999, “Scanning the available Dictyosteliumdiscoideum proteome for O-linked GlcNAc glycosylation sites using neural networks” Glycobiology. 9(10), pp. 1009-22.
[40] Eisenhaber B, Bork P, Yuan Y, Loeffler G, Eisenhaber F., 2000, “Automated annotation of GPI anchor sites” case study C. elegans. TIBS. 25(7), pp. 340-341.
[41] http://www.alphalyse.com/gpmaw_lite. html
[42] Gautier R, Douguet D, Antonny B and Drin G., 2008, “HELIQUEST: a web server to screen sequences with specific α-helical properties” Bioinformatics. 24(18), pp. 2101-2.

[43] Tusnády GE, Simon I., 1998, "Principlesgoverning amino acid composition of integral membrane proteins: application to topology prediction" J Mol Biol. 283(2), pp. 489-506.

[44] Tusnády GE, Simon I., 2001, "The HMMTOP transmembrane topology prediction server" Bioinformatics. 17(9), pp. 849-50.

[45] Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, Castro ED, Coggill P, Corbett M, Das U, DaughertyL, Duquenne L, Finn RD, Fraser M, Gough J, HaftD, Hulo N, Kahn D, KellyE, Letunic I, Lonsdale D, Lopez R, MaderaM, MaslenJ, McAnulla C, McDowallJ, McMenaminC, MiH, Mutowo-Muellenet P, MulderN, NataleD, OrengoC, Pesseat S, Punta M, QuinnAF, RivoireC, Sangrador-VegasA, Selengut JD, Sigrist CJA, Scheremetjew M, TateJ, Thimmajanarthanan M, Thomas PD, Wu CH, Yeats C, Yong S., 2012, InterPro in 2011"new developments in the family and domain prediction database" Nucleic Acids Research. 40(D1), pp. D306–D312.

[46] Jones P, Binns D, Chang H, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S, Lopez R, and HunterS., 2014, "InterProScan 5: genome-scale protein function classification" Bioinformatics. 30(9), pp. 1236–1240.

[47] Lehninger A, Cox MC, Nelson DL.,2005, Lehninger principles of biochemistry (4th ed. ed.). New York: W. H. Freeman.

[48] Maurer-Stroh S, Eisenhaber B, Eisenhaber F. N-terminal., 2002, "N-myristoylation of proteins: prediction of substrate proteins from amino acid sequence" J Mol Biol. 317(4), pp. 541-57.

[49] Kiemer L, Bendtsen JD, Blom N., 2005, "NetAcet: prediction of N-terminal acetylation sites" Bioinformatics. 21(7), pp. 1269-70.

[50] Johansen MB, Kiemer L and Brunak S., 2006, "Analysis and prediction of mammalian protein glycation" Glycobiology. 16(9), pp. 844-53.

[51] Miller ML, Blom N., 2009, "Kinase-specific prediction of protein phosphorylation sites" Methods Mol Biol. 527, pp. 299-310.

[52] Bagos PG, Liakopoulos TD, Spyropoulos IC and Hamodrakas SJ., 2004, "PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins" Nucleic Acids Res. 32, pp. W400-4.

[53] Citation: C. Z. Cai, L. Y. Han, Z. L. Ji, X. Chen, Y. Z. Chen., 2003, "SVM-Prot: Web-Based Support Vector Machine Software for Functional Classification of a Protein from Its Primary Sequence" Nucleic Acids Res., 31, pp. 3692-3697.

[54] Wang GLiX and Wang Z., 2009, "APD2: the updated antimicrobial peptide database and its application in peptide design" Nucleic Acids Research. 37, pp. D933-D937.

[55] Bjellqvist, B., Hughes, G. J., Pasquali, Ch., Paquet, N., Ravier, F., Sanchez, J.-Ch., Frutiger, S. &Hochstrasser, D. F., 1993 "The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences" Electrophoresis. 14, pp. 1023-1031.

[56] Bjellqvist, B., Basse, B., Olsen, E. and Celis, J. E., 1994, “Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions” Electrophoresis. 15, pp. 529-539.

[57] Laurie ATR and Jackson RM., 2005, “Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites” Bioinformatics. 21(9), pp. 1908-1916.

[58] Hendlich M, Rippmann F, Barnickel G., 1997, “LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins” J Mol Graph Model. 15(6), pp. 359-63, 389.

[59] Michalski ML, Erickson SM, Bartholomay LC, and Christensen BM., 2010, “Midgut Barrier Imparts Selective Resistance to Filarial Worm Infection in Culxpipienspipiens” PLoSNegl Trop Dis. 4(11), pp. e875.