

Metabolic pathway analysis and molecular docking analysis for identification of putative drug targets in *Toxoplasma gondii*: novel approach

Budhayash Gautam^{1*}, Gurmit Singh², Gulshan Wadhwa³, Rohit Farmer¹, Satendra Singh¹, Atul Kumar Singh³, Prashant Ankur Jain¹, Pramod Kumar Yadav¹

¹Department of Computational Biology & Bioinformatics, Jacob School of Biotechnology and Bioengineering, Sam Higginbottom Institute of Agriculture, Technology and Sciences-Deemed to be University, Allahabad-211007, India; ²Department of Computer Science and Information Technology, Shepherd School of Engineering and Technology, Sam Higginbottom Institute of Agriculture, Technology and Sciences-Deemed to be University, Allahabad-211007, India; ³Apex Bioinformatics Centre, Department of Biotechnology, Ministry of Science and Technology, CGO complex, Lodhi Road, New Delhi - 110 003, India; Budhayash Gautam - Email: budhayash.gautam@shiats.edu.in; Phone: 91-532-3202133, +91-9451427316; *Corresponding author

Received January 09, 2012; Accepted January 17, 2012; Published February 03, 2012

Abstract:

Toxoplasma gondii is an obligate intracellular apicomplexan parasite that can infect a wide range of warm-blooded animals including humans. In humans and other intermediate hosts, *toxoplasma* develops into chronic infection that cannot be eliminated by host's immune response or by currently used drugs. In most cases, chronic infections are largely asymptomatic unless the host becomes immune compromised. Thus, *toxoplasma* is a global health problem and the situation has become more precarious due to the advent of HIV infections and poor toleration of drugs used to treat *toxoplasma* infection, having severe side effects and also resistance have been developed to the current generation of drugs. The emergence of these drug resistant varieties of *T. gondii* has led to a search for novel drug targets. We have performed a comparative analysis of metabolic pathways of the host *Homo sapiens* and the pathogen *T. gondii*. The enzymes in the unique pathways of *T. gondii*, which do not show similarity to any protein from the host, represent attractive potential drug targets. We have listed out 11 such potential drug targets which are playing some important work in more than one pathway. Out of these, one important target is Glutamate dehydrogenase enzyme; it plays crucial part in oxidation reduction, metabolic process and amino acid metabolic process. As this is also present in the targets of tropical diseases of TDR (Tropical disease related Drug) target database and no PDB and MODBASE 3D structural model is available, homology models for Glutamate dehydrogenase enzyme were generated using MODELLER9v6. The model was further explored for the molecular dynamics simulation study with GROMACS, virtual screening and docking studies with suitable inhibitors against the NCI diversity subset molecules from ZINC database, by using AutoDock-Vina. The best ten docking solutions were selected (ZINC01690699, ZINC17465979, ZINC17465983, ZINC18141294_03, ZINC05462670, ZINC01572309, ZINC18055497_01, ZINC18141294, ZINC05462674 and ZINC13152284_01). Further the Complexes were analyzed through LIGPLOT. On the basis of Complex scoring and binding ability it is deciphered that these NCI diversity set II compounds, specifically ZINC01690699 (as it has minimum energy score and one of the highest number of interactions with the active site residue), could be promising inhibitors for *T. gondii* using Glutamate dehydrogenase as Drug target.

Key words: Homology modeling, Molecular dynamics, Docking, Metabolic Pathway Analysis, Glutamate dehydrogenase, *Toxoplasma gondii*, Structural biology, Drug targets, KEGG.

Background:

Toxoplasma gondii is an obligate intracellular Apicomplexan parasite that can infect a wide range of warm-

blooded animals including humans [1]. *Toxoplasma gondii* was initially discovered by accident, in 1908, by a scientist, Charles Nicolle, who was working in North Africa and searching for a

reservoir of *Leishmania* in a native rodent, *Ctenodactylus gundi*. The gundis live in the foothills and mountains of Southern Tunisia and were commonly used to study *Leishmania* at the Pasteur Institute in Tunis. The name *Toxoplasma* means 'arc form' in Greek and was named according to the crescent-shaped morphology of the tachyzoite and bradyzoite stages of the organism observed by the scientists. At about the same time, Alfonso Splendore working in Sao Paulo discovered a similar parasite in rabbits. This pathogen is one of the most common in humans due to many contributing factors that include: (1) its complex life cycle allows it to be transmitted both sexually via felid fecal matter and asexually via carnivorousness. (2) *Toxoplasma* has an extremely wide host cell tropism that includes most nucleated cells. (3) In humans and other intermediate hosts, *Toxoplasma* develops into a chronic infection that cannot be eliminated by the host's immune response or by currently used drugs. In most cases, chronic infections are largely asymptomatic unless the host becomes immune compromised. Together, these and other properties have allowed *Toxoplasma* to achieve infection rates that range from ~23% in the USA [2] to 50-70% in France [3].

There are two major reasons that new drugs are needed to treat *Toxoplasma* infections. First, the drugs currently used to treat *Toxoplasma* infections are poorly tolerated, have severe side effects, and cannot act against bradyzoites [4]. Second, there are reports that *Toxoplasma* is developing resistance to the current generation of drugs [5, 6]. How resistance to these drugs has developed is not known but is critical to understand because it will lead to improved drug design and will increase our understanding of the biological functions of these drug targets. One way to understanding mechanisms of resistance is to compare the transcriptional profiles of wild-type and resistant parasites grown in the absence or presence of the drug. Such studies in bacterial resistance have demonstrated that pathogen responses to antibiotics are multifactorial and complex [7]. Whether the same will be true in *Toxoplasma* is unclear, but data from these types of experiments will likely impact new anti-*Toxoplasma* drug design.

Over the last decade, complete genome sequences of several pathogens have been determined, and many more such projects are currently under way. While, these data potentially contain all the determinants of host-pathogen interactions and possible drug targets, Development of effective therapies for intracellular eukaryotic pathogens is a serious challenge, given the protected location of these pathogens and the similarity of their biology to that of the host [8]. Genomics approach can be applied to evaluate the suitability of potential targets using two criteria, i.e. "essentiality" and "selectivity" [9]. The target must be essential for the growth, replication, viability or survival of the microorganism, i.e. encoded by genes critical for pathogenic life-stages. The microbial target for treatment should not have any well-conserved homolog in the host, in order to address cytotoxicity issues. This can help to avoid expensive dead-ends when a lead target is identified and investigated in great detail only to discover at a later stage that all its inhibitors are invariably toxic to the host. Genes that are conserved in different genomes often turn out to be essential [10]. A gene is deemed to be essential if the cell cannot tolerate its inactivation by mutation, and its status is confirmed using

conditional lethal mutants. Identifying cellular processes that are unique to the parasite is therefore a crucial step towards defining appropriate drug targets. Detection of genes that are non-homologous to human genes, and are essential for the survival of the pathogen represents a promising means of identifying novel drug targets [9].

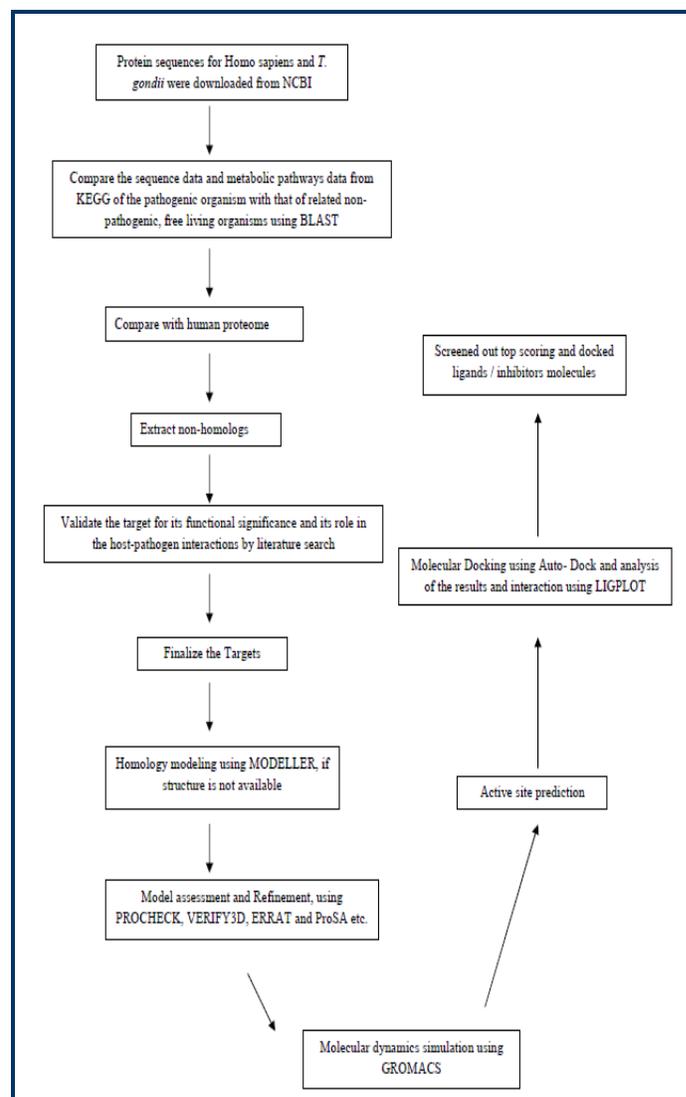


Figure 1: A workflow for complete drug target identification using metabolic pathways analysis and molecular modeling, docking studies.

We have performed a comparative analysis of metabolic pathways of the host *Homo sapiens* and the pathogen *T. gondii*. Enzymes from the biochemical pathways of *T. gondii* from the KEGG metabolic pathway database were compared with proteins from the host *H. sapiens*, by performing a BLASTp search against the non-redundant database restricted to the *H. sapiens* subset. Enzymes, which do not show similarity to any of the host proteins, below the threshold, were filtered out as potential drug targets [11]. The *T. gondii* genes for the identified enzymes were also retrieved for BLASTp search to identify any homologous protein in humans. We have listed out such potential drug targets which are playing some

important work in more than one pathway. Out of these, one important target is Glutamate dehydrogenase enzyme, which is having several activities like catalytic activity, oxidoreductase activity and binding activity. Glutamate dehydrogenase is also involved in oxidation process, metabolic process and amino acid metabolic process. During metabolic pathways analysis it was found that, it is a common enzyme in three metabolic pathways Alanine, Aspartate and Glutamate Metabolism, Arginine and Proline Metabolism and Nitrogen Metabolism. As this is also present in the targets of tropical diseases of TDR (Tropical disease related Drug) target database and no PDB and MODBASE 3D structural model is available. Thus, as a case study, we have built homology models and validated by using various online servers. The model was further explored for the molecular dynamics simulation study, virtual screening and molecular docking studies with suitable inhibitors. The three dimensional model of glutamate dehydrogenase presented here would be helpful in guiding both enzymatic studies as well as design of specific inhibitors. A complete drug target identification using metabolic pathways analysis and molecular modeling, docking studies workflow is followed in this work and given in **Figure 1**.

Methodology:

Identification of Unique Enzymes as Drug Targets

Metabolic pathway information was obtained from the pathway database Kyoto Encyclopedia of Genes and Genomes. Enzyme commission numbers (EC) of the pathogen *Toxoplasma gondii* and the host *H. sapiens* were extracted from the KEGG database. Pathways unique to *Toxoplasma gondii* were filtered out. Out of total 84 metabolic pathways in humans 55 metabolic pathways were present in *T. gondii*. Out of these 55 metabolic pathways, 29 were such which contain unique enzymes which are only present in *T. gondii* and absent in humans (**Table 1, see supplementary material**). These are the pathways that do not appear in the host (*H. sapiens*) but are present in the pathogen. We further identified unique enzymes among shared pathways under carbohydrate metabolism, energy metabolism, lipid metabolism, nucleotide metabolism, amino acid metabolism, glycan biosynthesis and metabolism of co-factors and vitamins were obtained from the KEGG database. A total of 48 enzymes that are present in *Toxoplasma gondii* but absent in *H. sapiens* were obtained and their corresponding protein sequences were retrieved from the KEGG database. Out of these 48 enzymes, 11 enzymes were commonly present in more than one pathway (**Table 2, see supplementary material**) The protein sequences for these 48 unique enzymes were retrieved and were subjected to BLAST search against human protein sequences database at an expectation E-value cutoff of 10^{-2} to identify non-homologous genes in *Toxoplasma gondii*. Removing enzymes from the pathogen that share a similarity with the host protein ensures that the targets have nothing in common with the host proteins, thereby, eliminating undesired host protein-drug interactions [12]. The above search resulted in 48 enzymes that had “no hits” in BLAST search. Thirty of these 48 “no hits” belonged to the unique pathways set.

Comparative Homology Modeling

Out of the 11 enzymes, commonly present in more than one pathway, we have built a homology model of one of the potential drug target, Glutamate dehydrogenase enzyme (Gene I.D.: TGME49_093180 and EC No. is EC: 1.4.1.4.

Template Selection and Model Building

Template selection is a critical step in homology modeling. The amino acid sequence of NADP-specific glutamate dehydrogenase of *Toxoplasma gondii* ME49 (target) (XP_002370120.1) was retrieved from the protein database of National Center of Biotechnology Information (NCBI, <http://www.ncbi.nlm.nih.gov/>). The three-dimensional structure of the protein was not yet available in Protein Data Bank and Modbase Database. Also there is no expression information available for this gene; hence the present exercise of developing the 3D model of the glutamate dehydrogenase of *Toxoplasma gondii* was undertaken. BLASTP search was performed against Brookhaven Protein Data Bank (PDB) with the default parameters to find suitable templates for homology modeling.

The academic version of MODELLER9v6 (<http://www.salilab.org/modeller/>), was used for 3D structure generation based on the information obtained from sequence alignment. The MODELLER software employs probability density functions (PDFs) as the spatial restraints rather than energy. The 3D model of a protein is obtained by optimization of the molecular pdf such that the model violates the input restraints as possible. The molecular pdf was derived as a combination of pdfs restraining individual spatial features of the whole molecule.

Model Assessment and Refinement

Out of 6 models generated, 5 by MODELLER and one by ModWeb server, the one with the best G-score of PROCHECK and with the best VERIFY3D profile was subjected to energy minimization. VERIFY3D (a structure evaluation server) was used to check the residue profiles of the obtained three-dimensional models. In order to assess the stereo-chemical qualities of the three dimensional models, PROCHECK analysis was performed. Quality evaluation of the model for the environment profile was also predicted using ERRAT structure evaluation server. The final refined model was evaluated for its atomic contacts using the Whatif program to identify bad packing of side chain atoms or unusual residue contacts. The predicted model was also analyzed for the recognition of errors in the three dimensional structure by using the ProSA web server. This model was further subjected for identification of active site and docking study.

With the aim of evaluating the stability and folding, conformational changes and getting insights into the natural dynamics on different timescale of protein in solution, 10 nanoseconds (ns) molecular dynamics simulation were performed. The molecular dynamics (MD) simulations of modeled glutamate dehydrogenase protein were carried out with the GROMACS 4.0.6 software package by employing GROMOS 96 force field and the flexible SPC (Simple Point Charge) water model. The initial structure was immersed in a periodic water box of cubic shape (0.5 nm thick). Electrostatic energy was calculated using the particle mesh Ewald method. Cutoff distance for the calculation of the coulomb and van der Waals interaction was 1.0 Å. After energy minimization using a steepest descent for 1000 steps, the system was subjected to equilibration at 300k and normal pressure for 100 ps under the conditions of position restraints for heavy atoms. LINC constraints were performed for all bonds, keeping the whole

protein molecule fixed and allowing only the water molecule to move to equilibrate with respect to the protein structure. The system was coupled to the external bath by the Berendsen pressure and temperature coupling. The final MD calculations were performed for 10 ns under the same conditions except that the position restraints were removed. The results were analyzed using the standard software provided by the GROMACS package. An average structure was refined further using a steepest descent energy minimization.

Binding Site Prediction

Binding site was characterized by CASTp and PASS and compared by using the information of binding sites. By comparing prediction of above two algorithms, best active site was selected.

Molecular Docking

The docking of glutamate dehydrogenase was performed, against the NCI subset II molecules retrieved from ZINC database by using Autodock- vina (<http://vina.scripps.edu/>), where 1,364 molecules from the NCI diversity subset II (<http://zinc.docking.org/index.shtml>) were screened. The Python scripts in MGL tools package were used to analyze the docking results.

Analyzing the Docking Results

The search for the best ways to fit ligand molecules from the NCI diversity subset II, into glutamate dehydrogenase modeled structure, using Autodock- vina resulted in docking files that contained detailed records of docking. These log files were read into ADT (Auto Dock Tool) to analyze the results of docking. The similarity of docked structures was measured by computing the root mean square deviation (RMSD) between the coordinates of the atoms and creating clustering of the conformations based on the RMSD values. In most cases the first cluster was also the largest cluster found. The lowest binding energy conformation in the first cluster was considered as the most favorable docking pose. Binding energies that are reported represent the sum of the total intermolecular energy, total internal energy and torsional free energy minus the energy of the unbound system. The top ten ligands were selected based on the energy score after virtual screening **Table 3 (see supplementary material)** and were further analyzed by the program LIGPLOT. The Ligplot represents the hydrogen and hydrophobic interactions between ligand and active site residues.

Discussion:

Metabolic pathways analysis

A total of 48 enzymes that are found present in *Toxoplasma gondii* but absent in *H. sapiens* were listed out as potential drug targets. Out of these 48 enzymes, some are already reported in literature and often used as target in treatment of toxoplasmosis e.g. (i) Pyrimethamine is a competitive inhibitor of parasite DHFR with narrow (~ six-fold) selectivity for the *T. gondii* enzyme compared with the mammalian enzyme [13, 14], (ii) DHPS is the target of sulfa drugs [15 - 17], (iii) Acetyl-CoA, which mediates the formation of malonyl -CoA [18] etc..

Template search and Model Building

Based on the maximum identity with high score and lower e-value crystal structure of Glutamate dehydrogenase (NADP+)

of *Plasmodium falciparum* (PDB code: 2bma_A) was selected as template. The sequence identity and similarity between the target and template are 58% and 73%, respectively. E- Value was 0.0. The sequence alignment of glutamate dehydrogenase of *Toxoplasma gondii* and *Plasmodium falciparum* was carried out using the CLUSTALW (<http://www.ebi.ac.uk/clustalw/>) program.

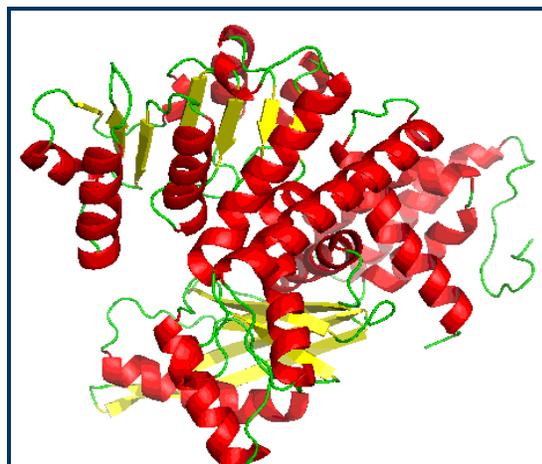


Figure 2: Ribbon representation of modeled glutamate dehydrogenase protein. The α -helices & β -sheets are shown as helices and ribbons respectively & rest are shown as loops. (Pymol Model)

Model Assessment and Refinement

All the six glutamate dehydrogenase models obtained were validated by using the SAVS and PROSA server. Ramachandran plots were drawn and the structures were analyzed by PROCHECK, a well-known protein structure checking program. By comparison of the results for all the models second model was judged as best among six (**Figure 2**). In the case of second model, it was found that the phi/psi angles of 91.3% of the residues fell in the most favored regions, 8.0% of the residues fell in the additional allowed regions, 0.7% fell in the generously allowed regions, and only 0.0 % of the residues fell in the disallowed regions. All these findings suggest a stereo-chemically very good model. The overall PROCHECK G factor for the homology-modeled structure was 0.00. The energetic architecture as predicted by PROSA score was negative (-9.62) for the modeled protein. The value is close to that of template (-9.60), which indicates the reliability of the model. The final structure was validated by an ERRAT graph. The quality factor of 77.106 indicated good quality, as scores of >50 are acceptable for a reasonable model. High quality of model is also confirmed from VERIFY 3D server as 76.00% of residues of modeled protein showed a score higher than 0.02 thus the model showed satisfactory 3D-1D scores for all the residues in the sequence. Results of WHAT_CHECK also indicate about the correctness of the modeled structure. These results indicate that the obtained structure has reasonably good quality. This model was used for virtual screening of the potential inhibitors for glutamate dehydrogenase.

Molecular Dynamics Simulation

The predicted Model of glutamate dehydrogenase has shown good accuracy of the structure. But in order to use this model

for virtual screening predicted model should have stable molecular dynamic behavior. The molecular dynamic stability has been established by performing the molecular dynamic simulation study using GROMACS 4.0.6 software package. The root mean square deviation (RMSD) during the simulation was increasing in the beginning but after 220 ps it becomes almost constant for rest of the duration of the simulation. This suggests that the glutamate dehydrogenase model has very less RMSD for the backbone and it also has less flexibility, indicating that model has a stable dynamic behavior structure. Molecular dynamic simulation study showed that the energy of the molecule was found to be constant throughout the time period of simulation. This suggests that the molecule has a stable structure as required for the virtual screening and drug designing. The root mean square (RMS) fluctuations were very less. Most the atoms were free from the RMS fluctuations. Very few atoms have shown RMS fluctuation at C and N terminal due to the loop region. This suggests the glutamate dehydrogenase model has an accurate and stable structure which can be used for virtual screening. The simulation studies also indicated that radius of gyration was increasing in the beginning but after 400 ps it decreases upto 420 ps and finally became almost constant for rest of the duration of the simulation. This suggests that the glutamate dehydrogenase model has a compact structure which provides the required stability to the molecule.

Active Site Identification

After getting the final model, the possible binding sites of modeled structure were searched using the CASTp server and PASS program. Ten possible binding sites were obtained. These sites were compared with active site of the template and it was found that pocket1 is highly conserved with the template. Since the glutamate dehydrogenase protein of *Toxoplasma gondii* and *Plasmodium falciparum* are well conserved in both sequence and structure, it is predicted that their biological function may be identical. From the structure-structure comparison of template and final refined models of glutamate dehydrogenase using SPDBV program, it was found that the residues in the site1 Phe294, Pro295, Lys292, Gly293, Gly275, Ala272, Leu296, Ala276, Lys277, Arg417 and Asp421 are highly conserved within the active site of template. In this study, site1 is chosen as the binding site to dock with the NCI diversity set molecules.

Virtual screening

The top ten ligand molecules having minimum energy were screened out as the possible inhibitors for glutamate dehydrogenase [19] **Table 3 (see supplementary material)**. All selected molecules were having energy score as follows: Ligand ZINC01690699 is having minimum energy score -9.1 Kcal/Mol. Ligand ZINC17465979 are having -8.1 Kcal/Mol. Four ligands, ZINC17465983, ZINC18141294_03, ZINC05462670 and ZINC01572309 are having energy score -7.9 Kcal/Mol, Four ligands, ZINC18055497_01, ZINC18141294, ZINC05462674 and ZINC13152284_01 are having energy score - 7.8. With the help of the Ligplot study we have selected ZINC01690699 as the possible inhibitor lead molecule, as it has minimum energy score and one of the highest number of interactions with the active site residue, it has 09 hydrophobic and 2 hydrogen interactions. **Figure 3 (a) to (e)**, represent interactions of the top five ligands drawn by ligplot according to the energy score.

Correlation coefficient analysis was performed between energy score calculated for all selected ligands molecules; log P value and molecular weight **Table 3 (see supplementary material)**. The correlation coefficient is a statistical calculation that is used to examine the relationship between two sets of value. The Correlation coefficient between energy score and log P value was -0.101603358 and between energy score and molecular weight was -0.236816724. These values suggest a clear negative correlation between energy score of the ligands and their molecular weight and log P value. Thus the energy score of the ligands is independent of their molecular weight and log P value. The value of the correlation coefficient tells us about the strength and the nature of the relationship. Correlation coefficient values can range between +1.00 to -1.00. The negative correlation suggests that energy score may depend on interactions or the conformation of ligands and active site residues [20].

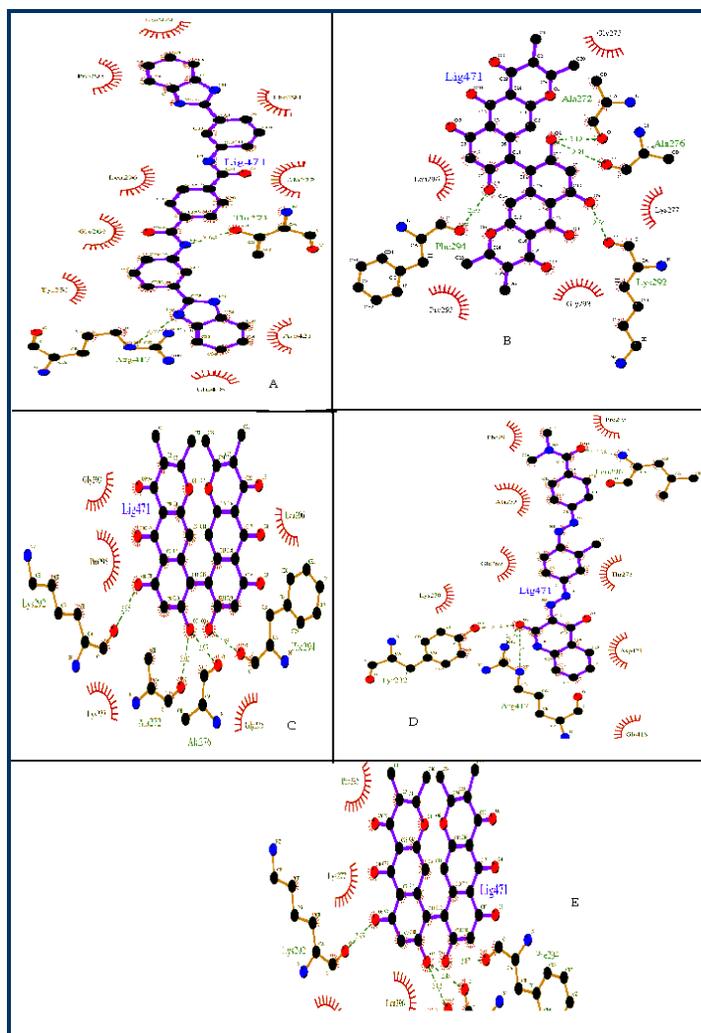


Figure 3: Ligplot showing the protein-ligand interactions of top five ligands, based on energy score (hydrogen bonding and hydrophobic), generated by Ligplot program. (A) ZINC01690699; (B) ZINC17465979; (C) ZINC17465983; (D) ZINC18141294_03 and (E) ZINC05462670.

It was found the inhibitor ZINC01690699 has minimum energy score which reveals higher binding affinity towards the glutamate dehydrogenase and the inhibitor was also showing one of the best interactions with residues of the active site. The other important drug like properties like molecular weight and logP value were also found within the limits of drug like molecules. For the wet laboratory validation of present study, samples of *T. gondii* can be collected from the pathological laboratories and the molecule ZINC01690699 can be procured from the Pubchem database. Growth inhibition of *T. gondii* by inhibitor ZINC01690699 can be established by the qualitative and quantitative microbiological techniques. Thus, the three dimensional model of glutamate dehydrogenase presented here would be helpful in guiding both enzymatic studies as well as design of specific inhibitors.

Conclusion:

Our approach of comparative metabolic pathway analysis and molecular docking analysis resulted in the identification of potential drug targets. For the first time, the availability of complete genome sequences of many bacterial species is facilitating many novel computational approaches. The complete definition of all gene products by gene identification tools exemplified here is just the first step. The data presented here demonstrates that stepwise prioritization of genome open reading frames using simple biological criteria can be an effective way of rapidly reducing the number of genes of interest to an experimentally manageable number. This process is an efficient way for enriching potential target genes, and for identifying those that are critical for normal cell function. The generation of a comprehensive essential gene list will allow an accelerated genetic dissection of traits such as metabolic flexibility and inherent drug resistance that render *Toxoplasma gondii* such a tenacious pathogen. Such a strategy will enable us to locate critical pathways and steps in pathogenesis; to target these steps by designing new drugs; and to inhibit the infectious agent of interest with new antimicrobial agents. Hence, in present study, it can be concluded that the molecule ZINC01690699 1-N, 4-N-bis [3-(1H-benzimidazol-2-yl) phenyl] benzene-1, 4-dicarboxamide has the potential to inhibit the growth of *Toxoplasma gondii* and can act as remedy for the treatment of *Toxoplasma gondii* infection.

Acknowledgement:

The authors are grateful to the Sam Higginbottom Institute of Agriculture, Technology & Sciences, Deemed to be University, Allahabad for providing the facilities and support to complete the present research work.

References:

- [1] Montova JG & Liesenfeld O, *Lancet*. 2004 **363**: 1965 [PMID: 15194258].
- [2] Jones JL *et al.* *Am J Epidemiol*. 2001 **154**: 357 [PMID: 11495859].
- [3] Tenter AM *et al.* *Int J Parasitol*. 2000 **30**: 1217 [PMID: 11113252]
- [4] McCabe R. *Cambridge University Press*. 2001 p. 319-59.
- [5] Baatz H *et al.* *Ocul Immunol Inflamm*. 2006 **14**: 185 [PMID: 16766403]
- [6] Aspinall TV *et al.* *J Infect Dis*. 2002 **185**: 1637 [PMID: 12023770]
- [7] Brazas MD & Hancock RE. *Drug Discov Today* 2005 **10**: 1245 [PMID: 16213417]
- [8] Miesel L *et al.* *Nat Rev Genet*. 2003 **4**: 442 [PMID: 12776214]
- [9] Sakharkar KR *et al.* *In Silico Biol*. 2004 **4**: 355 [PMID: 15724285]
- [10] Kobayashi K *et al.* *Proc. Natl. Acad. Sci. USA* 2003 **100**: 4678 [PMID: 12682299]
- [11] Singh *et al.* *Elixir Bio Phy*. 2011 **32**.
- [12] Singh *et al.* *Bioinformation* 2011 **6**: 31 [PMID: 21464842]
- [13] Schweitzer BI *et al.* *FASEB J*. 1990 **4**: 2441 [PMID: 2185970].
- [14] Chio LC & Queener SF, *Antimicrob. Agents Chemother*. 1993 **37**: 1914 [PMID: 8239605].
- [15] Dorsey G *et al.* *Am J Trop Med Hyg*. 2004 **71**: 758 [PMID: 15642967].
- [16] Ogutu BR *et al.* *Trop Med Int Health*. 2005 **10**: 484 [PMID: 15860096].
- [17] Omar SA *et al.* *Exp Parasitol*. 2005 **110**: 73 [PMID: 15804381].
- [18] Seeber F. *Curr Drug Targets Immune Endocr Metabol Disord*. 2003 **3**: 99 [PMID: 12769782].
- [19] Khan AU *et al.* *Bioinformation*. 2011 **5**: 331 [PMID: 21383920]
- [20] Farmer R *et al.* *Bioinformation*. 2010 **4**: 290 [PMID: 20978601]

Edited by P Kanguane

Citation: Gautam *et al.* *Bioinformation* 8(3): 134-141 (2012)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Unique enzymes, which are only present in *T. gondii* and homologs are absent in humans.

S. No.	Metabolic Pathway	Enzyme	Enzyme Detail	<i>T. gondii</i> Gene ID.
1.	Glycolysis Gluconeogenesis	[EC:4.1.1.49]	Phosphoenolpyruvate carboxykinase	TGME49_089930
2.	Citrate Cycle (TCA cycle)	[EC:4.1.1.49]	Phosphoenolpyruvate carboxykinase	TGME49_089650
3.	Fatty Acid Biosynthesis	[EC:2.3.1.180] [EC: 4.2.1.-] [EC:1.3.1.9]	3-Oxoacyl-(acyl-carrier-protein) synthase III family protein; (3R)-Hydroxymyristoyl ACP dehydratase; Enoyl-acyl carrier reductase	TGME49_031890 TGME49_121570 TGME49_051930
4.	Oxidative Phosphorylation	[EC:1.6.99.3] [EC:3.6.3.6]	Mitochondrial alternative NADH dehydrogenase 1; Pyridine nucleotide-disulphide oxidoreductase; Plasma-membrane H ⁺ -ATPase	TGME49_009150 TGME49_088830 TGME49_052640 TGME49_084600
5.	Purine Metabolism	(EC:3.2.1.3) [EC:2.7.7.4]	SIN-like domain-containing protein; Sulfate adenyltransferase-adenylsulfate kinase	TGME49_064460 TGME49_082230
6.	Pyrimidine Metabolism	[EC:3.5.2.3] (EC:3.2.1.3); [EC:2.1.1.451.5.1.3]	Dihydroorotase protein; SIN-like domain-containing protein; Bifunctional dihydrofolate reductase / thymidylate synthase	TGME49_093610 TGME49_064460 TGME49_049180
7.	Alanine, Aspartate and Glutamate Metabolism	[EC:2.1.3.2] [EC:1.4.1.2] [EC:1.4.1.4]	Aspartate carbamoyltransferase; NAD-specific glutamate dehydrogenase	TGME49_091640 TGME49_049390 TGME49_093180
8.	Glycine, Serine and Threonine Metabolism	[EC:2.7.2.4] [EC:1.2.1.11] [EC:2.7.1.39] [EC:4.2.3.1]	Aspartate kinase; Aspartate-semialdehyde dehydrogenase; Homoserine kinase; Threonine synthase	TGME49_027090 TGME49_005420 TGME49_016640 TGME49_020840
9.	Cysteine and Methionine Metabolism	[EC:2.5.1.47] [EC:2.1.1.10] [EC:2.7.2.4] [EC:1.2.1.11] [EC:2.3.1.31]	O-acetylserine (thiol) lyase; Cysteine synthase A; Homocysteine S-methyltransferase-1; Aspartate kinase; Aspartate-semialdehyde dehydrogenase; Homoserine O-acetyltransferase	TGME49_078910 TGME49_057750 TGME49_027090 TGME49_005420 TGME49_120730
10.	Lysine Biosynthesis	[EC:2.7.2.4] [EC:1.2.1.11]	Aspartate kinase; Aspartate-semialdehyde dehydrogenase	TGME49_027090 TGME49_005420
11.	Arginine and Proline Metabolism	[EC:1.4.1.4]	NADP-specific glutamate dehydrogenase	TGME49_093180
12.	Phenylalanine, Tyrosine and Tryptophan Biosynthesis	[EC:4.2.3.5] [EC:2.5.1.54]	Chorismate synthase; DAHP synthetase	TGME49_001380 TGME49_021260
13.	Beta-Alanine Metabolism	[EC:6.3.2.1]	Pantoate--beta-alanine ligase	TGME49_065870
14.	Seleno-compound Metabolism	[EC:2.7.7.4]	Sulfate adenyltransferase-adenylsulfate kinase	TGME49_082230
15.	Starch and Sucrose Metabolism	[EC:2.4.1.34] [EC:2.4.1.15] [EC:2.4.1.21]	1,3-beta-glucan synthase; Trehalose-6-phosphate synthase domain-containing protein; Glycogen synthase	TGME49_078110 TGME49_097720 TGME49_022720
16.	Amino Sugar and Nucleotide Sugar Metabolism	[EC:2.7.7.64]	UDP-N-acetylglucosamine pyrophosphorylase	TGME49_018200
17.	Glycosphingolipid Biosynthesis - globo series	[EC: 2.3.1.-]	Acetyl-CoA transporter	TGME49_015940
18.	Pyruvate Metabolism	[EC:4.4.1.5] [EC:4.1.1.49]	Lactoylglutathione lyase; Phosphoenolpyruvate carboxykinase	TGME49_048400 TGME49_089930 TGME49_089650
19.	Butanoate Metabolism	[EC:1.1.1.157]	3-Hydroxybutyryl-CoA dehydrogenase	TGME49_032090
20.	Vitamin B6 Metabolism	[EC:2.7.1.35] [EC:4.2.3.1] [EC: 4.-.-] [EC: 2.6.-.-]	Hypothetical protein; Threonine synthase; Ethylene inducible protein; Glutamine amidotransferase	TGME49_097080 TGME49_020840 TGME49_037140 TGME49_081490
21.	Nicotinate and Nicotinamide Metabolism	[EC:2.7.1.23] [EC:1.6.1.1]	ATP-NAD kinase domain-containing protein; Transhydrogenase	TGME49_044700 TGME49_118650
22.	Pantothenate and CoA-Biosynthesis	[EC:6.3.2.1]	Pantoate--beta-alanine ligase	TGME49_065870
23.	Lipoic acid Metabolism	[EC:2.7.7.63]	Lipoate-protein ligase A	TGME49_071820
24.	Folate Biosynthesis	[EC:2.5.1.15] [EC:2.7.6.3] [EC:2.6.1.85]	Hydroxymethylidihydropterin; Pyrophosphokinase-dihydropteroate synthase; Para-aminobenzoate synthase	TGME49_059550 TGME49_002920
25.	Porphyrin and Chlorophyll Metabolism	[EC:1.3.99.22] [EC:4.4.1.17]	Radical SAM domain containing protein; Coproporphyrinogen III oxidase; Cytochrome c heme lyase;	TGME49_088640 TGME49_114040

26.	Terpenoid Biosynthesis	Backbone	[EC:2.2.1.7] [EC:1.1.1.267] [EC:4.6.1.12] [EC:1.17.1.2]	1-deoxy-D-xylulose 5-phosphate synthase; 1-deoxy-D-xylulose 5-phosphate reductoisomerase; 2C-methyl-D-erythritol 2; lytB domain-containing protein, Hydroxy-3-methylbut-2-enyl diphosphate reductase	TGME49_008820 TGME49_014850 TGME49_055690 TGME49_027420
27.	Nitrogen Metabolism		[EC:1.4.1.4]	NADP-specific glutamate dehydrogenase	TGME49_093180
28.	Sulfur Metabolism		[EC:2.3.1.31] [EC:2.5.1.47]	Homoserine O-acetyltransferase; O-acetylserine (thiol) lyase	TGME49_120730 TGME49_078910
29.	Biosynthesis of Unsaturated Fatty Acids		[EC:1.1.1.100]	Oxoacyl-ACP reductase	TGME49_017740

Table 2: Enzymes, which are present in more than one pathway in *T. gondii* and absent in humans.

S. No.	<i>T. gondii</i> Gene I.D.	Enzyme	Metabolic Pathway	Function
1.	TGME49_089930	Phosphoenolpyruvate Carboxykinase	1. Glycolysis / Gluconeogenesis 2. Pyruvate Metabolism	Phosphoenolpyruvate carboxykinase (PEPCK) is an enzyme in the lyase family used in the metabolic pathway of gluconeogenesis. It converts oxaloacetate into phosphoenolpyruvate and carbon dioxide.
2.	TGME49_089650	Phosphoenolpyruvate Carboxykinase	1. Citrate cycle (TCA cycle) 2. Pyruvate Metabolism	Phosphoenolpyruvate carboxykinase (PEPCK) is an enzyme in the lyase family used in the metabolic pathway of gluconeogenesis. It converts oxaloacetate into phosphoenolpyruvate and carbon dioxide.
3.	TGME49_093180	NAD-specific glutamate dehydrogenase	1. Alanine, Aspartate and Glutamate Metabolism 2. Arginine and Proline Metabolism 3. Nitrogen Metabolism	(GLDH) is an enzyme, present in most microbes and the mitochondria of eukaryotes, required for urea synthesis, that converts glutamate to α-Ketoglutarate, and vice versa.
4.	TGME49_027090	Aspartate kinase	1. Glycine, Serine and threonine Metabolism 2. Cysteine and Methionine Metabolism 3. Lysine Biosynthesis	Catalyzes the phosphorylation of the amino acid aspartate. This reaction is the first step in the biosynthesis of three essential amino acids: methionine, lysine, and threonine, known as the "aspartate family".
5.	TGME49_005420	Aspartate-semialdehyde dehydrogenase	1. Glycine, Serine and Threonine Metabolism 2. Cysteine and Methionine Metabolism 3. Lysine Biosynthesis	It forms an early branch point in the metabolic pathway forming lysine, methionine, leucine and isoleucine from aspartate. This enzyme belongs to the family of oxidoreductases, specifically those acting on the aldehyde or oxo group of donor with NAD+ or NADP+ as acceptor.
6.	TGME49_020840	Threonine synthase	1. Glycine, Serine and Threonine Metabolism 2. Vitamin B6 Metabolism	This enzyme belongs to the family of lyases, specifically those carbon-oxygen lyases acting on phosphates.
7.	TGME49_078910	O-acetylserine (thiol) lyase	1. Cysteine and Methionine Metabolism 2. Sulfur Metabolism	The enzyme O-acetylserine (thiol)-lyase, using sulfide sources, converts O-acetylserine ester into cysteine, releasing acetate
8.	TGME49_120730	Homoserine O-acetyltransferase	1. Cysteine and Methionine Metabolism 2. Sulfur Metabolism	This enzyme belongs to the family of transferases, specifically those acyltransferases transferring groups other than aminoacyl groups. The two substrates of this enzyme are acetyl-CoA and L-homoserine, whereas its two products are CoA and O-acetyl-L-homoserine.
9.	TGME49_065870	Pantoate--beta-alanine ligase	1. Beta-Alanine Metabolism 2. Pantothenate and CoA biosynthesis	This enzyme belongs to the family of ligases, specifically those forming carbon-nitrogen bonds as acid-D-amino-acid ligases (peptide synthases). The 3 substrates of this enzyme are ATP, (R)- pantoate, and beta-alanine, whereas its 3 products are AMP, diphosphate, and (R)- pantothenate.
10.	TGME49_082230	Sulfate adenyltransferase-adenylsulfate kinase	1. Purine Metabolism 2. Selenocompound Metabolism	This enzyme belongs to the family of transferases, specifically those transferring phosphorus-containing nucleotide groups (nucleotidyltransferases). The two substrates of this enzyme are ATP and sulfate, whereas its two products are diphosphate and adenylsulfate.
11.	TGME49_064460	SIN-like domain-containing protein	1. Purine Metabolism 2. Pyrimidine Metabolism	Utilizes a DNA template, i.e. the catalysis of DNA-template-directed extension of the 3'-end of an RNA strand by one nucleotide at a time.

Table 3: Top ten ligand molecules identified from NCI diversity subset II of the ZINC database after virtual screening

S. No.	ZINC ID of the screened molecules	Energy score (Kcal/Mol)	No. of Hydrogen Interaction of ligand with active site residues	No. Hydrophobic Interaction of ligand with active site residues	Molecular Weight (g/mol)	Log P value
1.	ZINC01690699	-9.1	2	9	548.59336	6.0
2.	ZINC17465979	-8.1	5	5	546.52144	5.9
3.	ZINC17465983	-7.9	4	5	546.52144	5.9
4.	ZINC18141294_03	-7.9	3	8	453.47264	5.3
5.	ZINC05462670	-7.9	4	5	546.52144	5.9
6.	ZINC01572309	-7.9	3	5	474.55802	5.0
7.	ZINC18055497_01	-7.8	1	7	479.67574	4.7
8.	ZINC18141294	-7.8	1	12	453.47264	5.3
9.	ZINC05462674	-7.8	4	6	546.52144	5.9
10.	ZINC13152284_01	-7.8	1	7	392.4492	6.7