Molecular modeling on pyruvate phosphate dikinase of *Entamoeba histolytica* and in silico virtual screening for novel inhibitors

Preyesh Stephen · Ramachandran Vijayan · Audesh Bhat · N. Subbarao · R. N. K. Bamezai

Abstract Pyruvate phosphate dikinase (PPDK) is the key enzyme essential for the glycolytic pathway in most common and perilous parasite *Entamoeba histolytica*. Inhibiting the function of this enzyme could control the wide spread of intestinal infections caused by *Entamoeba histolytica* in humans. With this objective, we modeled the three dimensional structure of the PPDK protein. We used templates with 51% identity and 67% similarity to employ homology-modeling approach. Stereo chemical quality of protein structure was validated by protein structure validation program PROCHECK and VERIFY3D. Experimental proof available in literature along with the in silico studies indicated Lys21, Arg91, Asp323, Glu325 and Gln337 to be the probable active sites in the target protein. Virtual screening was carried out using the genetic docking algorithm GOLD and a consensus scoring function X-Score to substantiate the prediction. The small molecule libraries (ChemDivision database, Diversity dataset, Kinase inhibitor database) were used for screening process. Along with the high scoring results, the interaction studies provided promising ligands for future experimental screening to inhibit the function of PPDK in *Entamoeba histolytica*. Further, the phylogeny study was carried out to assess the possibility of using the proposed ligands as inhibitors in related pathogens.

Abbreviations

ATP	Adenosine triphosphate
AMP	Adenosine monophosphate
PEP	Phospho(enol)pyruvate
iP	Inorganic phosphate
E-P	Phosphorylhistidine intermediate

P. Stephen · A. Bhat · R. N. K. Bamezai (✉)
National Centre of Applied Human Genetics, School of Life Sciences, Jawaharlal Nehru University, New Delhi 110067, India
e-mails: bamezai@ncahg.org; bamezai@hotmail.com

P. Stephen · R. Vijayan · N. Subbarao
Center for Computational Biology and Bioinformatics, Jawaharlal Nehru University, Hall No. 4, Lecture Hall Complex, New Delhi 110067, India

Introduction

Entamoeba histolytica causes amoebiasis and is the third most frequent human parasitic infection [1]. It is known to infect the mucosa of the large intestine and may damage other organs, especially the liver [2]. There are about 50 million *Entamoeba histolytica* infections per year with a fatal outcome in 1,00,000 of those infected [3]. Despite the prevalence of Metronidazole based therapeutic treatment, the parasite nevertheless persists in intestine of 40–60% of patients [4]. Moreover, recent studies have reported the in-vitro generation of strains which are resistant to Metronidazole [5]. These observations bring out the need for developing a new strategy for the treatment of amoebiasis caused by *Entamoeba histolytica*.

Various studies have been attempted on targeting the enzyme cysteine proteinases [6, 7] and the glycolytic enzyme, phosphofructokinase [8] to control the

pathogenesis of the protozoan. However, lack of characterization due to the difficulty in separating different proteinases and the problem of non-specificity has posed hurdles in drug targeting of *Entamoeba histolytica*.

Entamoeba histolytica lacks Krebs cycle and oxidative phosphorylation enzymes, and adopts the exclusive way of ATP synthesis through glycolytic pathway [9, 10]. Hence glycolytic enzymes involved in a key pathway in the energy metabolism, could result as a promising target for inhibiting the normal growth of *Entamoeba histolytica*. Little is known about kinetic and structural aspects of most of the enzymes in the pathway [11]. Pyruvate phosphate dikinase (PPDK) is a critical enzyme involved in the glycolytic pathway. The absence of PPDK in animals and it's essential role in *Entamoeba histolytica* suggests it to be the best target for designing of antiparasitic agents [12, 13].

Studies on *Clostridium symbiosum* have revealed that pyruvate phosphate dikinase (PPDK) catalyzes the interconversion of ATP, Pi, and pyruvate with AMP, PPi, and PEP [14], using two separate active sites linked by a mobile domain containing the phosphoryl group carrier His455 [15]. At the first active site, to which ATP, Pi and Mg (II) are bound, His455 attacks the -P of ATP, forming a pyrophosphorylhistidine enzyme intermediate (E-PP) and AMP. The Pi ligand then reacts with the terminal phosphoryl group of E-PP to form a phosphorylhistidine intermediate (E-P) and PPi. The phosphorylhistidine of E-P then moves to the second active site where pyruvate, Mg(II), and a monovalent cation (K^+ or NH_4^+) are bound. Upon transfer of the phosphoryl group to pyruvate to form the final product PEP, the His455 residue returns to the first active site to initiate a new catalytic cycle [16]. The study of PPDK of *Entamoeba histolytica* for the mutants of ATP-grasp domain involving: Lys-22, Arg-92, Asp-321, Glu-323, and Gln-335 were found to be inactive; Arg-337, Glu-279, Asp-280, and Arg-135 were partially active; and Thr-253 and Gln-240 were almost fully active [17].

The rationale of using PPDK as a drug target in this study was based on the already mentioned advantage of its exclusive presence in the pathogen and complete absence in the human host. Homology modeling was performed with reliable templates and virtual screening was carried out to identify novel inhibitors for PPDK of *Entamoeba histolytica*.

Materials and methods

Sequence analysis

The protein sequence of pyruvate phosphate dikinase (PPDK) of the organism *Entamoeba histolytica* was obtained from the NCBI-Protein database (accession number XP_657332). Even though PPDK homologue is not known

to be present in humans [12], the PPDK sequence was subjected to various sequence analysis programs as explained below to examine its potential as a drug target.

Pairwise sequence alignment of PPDK from *Entamoeba histolytica* was carried out against human database in BLAST [18] as well as in Ensembl [19] without any significant hit. Domain analysis (ProDom, Pfam) was carried out to check whether any domain present in the target protein could interfere in the action of the drug. ProDom, a database of protein domain families, is useful for analyzing the protein domain arrangements and helps to analyze homology relationship in modular proteins. The ProDom building procedure MKDOM2 is based on recursive PSI-BLAST searches. The source protein sequences are non-fragmentary, derived from SWISS-PROT and TrEMBL databases. Whereas, Pfam, a database of multiple alignments and HMM profiles of protein domains or conserved protein regions, is actually composed of two sets of families. Pfam-A family is based on curated multiple alignment whereas Pfam-B family is derived from ProDom, a comprehensive set of protein domain family. The screening of the database in this study proved that the domains present in the PPDK of *Entamoeba histolytica* are not ubiquitous in human.

Pyruvate kinase (PK) the substitute of pyruvate phosphate dikinase in human [16] was aligned in clustalW [20] as well as EMBOSS pairwise alignment (http://www.ebi.ac.uk/emboss/align/) and observed to have very less similarity. Structure alignment of human PK was done against the 3D model of the protein PPDK, using combinatorial extension (CE) algorithm, available from San Diego Supercomputer Center [21]. Superimposition of A-Domain (N-Terminal ATP-Grasp domain) for the protein PPDK of *Entamoeba histolytica* and PK of human was also carried out. There was no significant structural similarity between the c-alpha atom of the protein structure and no sequence conservation in the active site residue of PPDK of *Entamoeba histolytica* to PK of human. The published literature [10, 12, 13] gave more confidence for using PPDK as the potential drug target for our study.

Blastp [18] against PDB databank [22] gave four significant hits: 1KBL, 1GGO, 1JDE and 1DIK. These proteins were different forms of pyruvate phosphate dikinase from *Clostridium symbiosum* and had given a hit of 51% identity, 67% positive scores and 2% gaps with an expect value of 0. Again these sequences were aligned using clustalW [20], obtaining a score of 50. This provided the broadly studied templates to model the target protein.

Homology model construction

The three dimensional structure of the protein was constructed using MODELLER8V2 [23, 24]. After confirming

the template proteins (1KBL, 1GGO, 1JDE and 1DIK) alignment was done in clustalW keeping the output format as PIR [25], which is the standard input format of MODELLER. Further on analyzing we noticed that some template sequences in the linear form are missing in the crystallized three dimensional structures, which could occur due to some difficulty in crystallizing portions of the protein. As the MODELLER developers suggest, we manually inserted some gaps to cope-up with the missing residues in the crystal structure of the template. The Python script was executed to calculate a model with all non-hydrogen atoms by satisfaction of spatial restraint. The model was screened for unfavorable steric contacts and remodeled using either a rotamer library database of INSIGHT II or manual rotations. Explicit hydrogens were added to the protein and was subjected to energy minimization using Discover [26] (module of INSIGHT II) using CVFF force field [27]. Energy minimization and relaxation of the loop regions was performed using 300 iterations in a simple minimization method. Again the steepest descent was carried out until the energy showed stability in the sequential repetition.

Model evaluation was performed in PROCHECK v3.4.4 [28] producing plots that were analyzed for the overall and residue-by-residue geometry. Ramachandran Plot [29] provided by the program PROCHECK assured very good confidence for the predicted protein. There were only 0.3% residues in the disallowed region and 0.9% residues in generously allowed regions. Most of these residues were in C-terminal domain. Nevertheless, PROCHECK assured the reliability of the structure and the protein was subjected to VERIFY3D [30], available from NIH MBI Laboratory Servers. RMSD (c-alpha atom) of the modeled protein with respect to the templates (1KBL, 1GGO, 1JDE and 1DIK) was calculated, using combinatorial extension (CE) algorithm available from San Diego Supercomputer Center [21].

Ligand binding site prediction

Site directed mutagenesis studies in catalytic site within the ATP-grasp domain of *Clostridium symbiosum* pyruvate phosphate dikinase have revealed that Lys-22, Arg-92, Asp-321, Glu-323, and Gln-335 are the critical residues involved in PPDK action [16]. Even though we had experimentally proven active sites of PPDK in *Clostridium symbiosum* [17], in silico prediction was done for the PPDK in *Entamoeba histolytica* using LIGSITE^csc algorithm [31] which uses connolly surface and defines surface–solvent–surface events. The algorithm proceeds as follows: the protein is projected onto a 3D grid with a step size of 1.0 Å; grid points are labeled as protein, surface, or

solvent using certain rules. A grid point is marked as protein if there is at least one atom within 1.6 Å. After the solvent excluded surface is calculated the surface vertices' coordinates are stored. A sequence of grid points, which starts and ends with surface grid points and which has solvent grid points in between, is called a surface–solvent–surface event. If the number of surface–solvent–surface events of a solvent grid exceeds a minimal threshold of 6, then this grid is marked as pocket. Finally, all pocket grid points are clustered according to their spatial proximity. The clusters are ranked by the number of grid points in the cluster. The top three clusters are retained and their centers of mass are used to represent the predicted pocket sites [31]. The binding pocket obtained by in silico studies on PPDK of *Entamoeba histolytica* was consistent with the known biochemical studies on *Clostridium symbiosum*.

Docking studies

Compound input libraries of 78,700 templates, based on structural/patent information on a target or protein domain of interest (ChemDivision data set—http://www.chemdiv.com/), 1,122 compounds specific to Kinase targets (Kinase Inhibitor Data base—http://www.lifechemicals.com/) and 1990 compounds with a rich structural and pharmacophore diversity (The National Cancer Institute Diversity Set—http://dtp.nci.nih.gov/branches/dscb/diversity_explanation.html), were chosen as the molecules in demand for virtual screening.

"ChemDivision data set" include both selective and dual kinase inhibitors, vasculature-targeting agents, "non-toxic" anti-miotics, dual and specific GPCR antagonistics, ion channels and developmental pathway modulators, which could be used for high throughput screening of lead molecules. "Kinase inhibitors" available from Life Chemicals are the compound collection that are composed using both ligand based and receptor based approaches to which Lipinsky's rule of five [32] as well as Veber criteria [33] and dissimilarity evaluation have been applied. The "Diversity set" was derived from almost 140,000 compounds available on plates and from this 71,756 compounds (due to unavailability of at least 1 g material) were then reduced to the final set using the program Chem-X (Oxford Molecular Group). The pharmacophores for the current structure were compared to the set of all pharmacophores found in structures already accepted into the diverse subset. If the current structure had more than a preset number of new pharmacophores, it was added to the diverse subset. Since the selection procedure was order dependent, the order in which the structures were considered was randomized. This procedure resulted in the selection of 1990 compounds in the diversity dataset.

The small molecule libraries (ChemDivision database, Diversity dataset, Kinase inhibitor database) which were available in 2D-strucuture data file (SDF) format were converted into 3D-MOL2 file with the program CORINA [34, 35]. These molecules were then read into INSIGHT II for further treatment such as energy minimization for 100 steps with CVFF force field [27]. During ligand preparation, we generated only one conformer, one charge state per molecule, although, GOLD generates large number of conformations for each ligands before docking. GOLD uses a genetic algorithm to explore the full range of ligand conformational flexibility with partial flexibility of the protein [36]. The workflow used for the virtual screening is depicted in Fig. 1. Library screening followed by detailed docking in the standard default setting was carried out with all the three databases. The parameters used in the library screening and standard default setting are listed in the Table 1. The binding affinity was estimated also by using the consensus scoring function X-Score V2.1 [37]. The similarity between the individual ligands against the database was done using the tool Jcsearch available in Jchem version 3.2 [38]. The properties of individual ligands were calculated with CSChem3DPro Version 5, available with the commercial suite ChemOffice [39]. OpenEye scientific software Filter 2.0.1 was used to eliminate inappropriate or undesirable compounds from the list. The parameters in the software Filter were modified to get more appropriate hits. We used relaxed parameter where the maximum numbers of heavy atoms were modified from 35 to 40, minimum number of carbons from 7 to 6, maximum number of carbons from 35 to 50 and minimum number of heteroatoms from 2 to 1. We had accepted predicted aggregators in the filtering. This software readily enables users to filter out the compounds with undesirable properties, using chemoinformatic parameters

Table 1 Annealing parameters (No. 1 and No. 2) and genetic algorithm parameters (No. 3 to No. 10) applied to the library screening and standard default screening in the program GOLD

No.	Parameters used	Library screening setting	Standard default setting
1	Van der Waals forces	10	4.0
2	Hydrogen bonding	05	2.5
3	Population size	50	100
4	No of islands	01	05
5	Niche size	02	02
6	Selection pressure	1.125	1.1
7	Migrate	0.00	2
8	No of operators	1,000	100,000
9	Mutate	100	95
10	Cross over	100	95

based on the classical rules. The ligands showing maximum interactions with the protein were plotted using the program LIGPLOT [40]. Hydrogen bond interactions were double-checked with the software GETNEARES, available with the program DOCK [41].

Phylogenetic analysis

We carried out a one step forward study to construct a phylogeny tree of closely related pathogens to *Entamoeba histolytica*. The PPDK protein sequences of homologous organisms were extracted from NCBI database and aligned using CLUSTAL X [42], keeping the output format in Phylip. A parsimony tree [43] was built using the Phylip package [44] with replicate or bootstrap of 500. Later the

Fig. 1 Workflow diagram followed in the in silico virtual screening for inhibitors of PPDK of *Entamoeba .histolytica*

Workflow Diagram of Virtual Screening
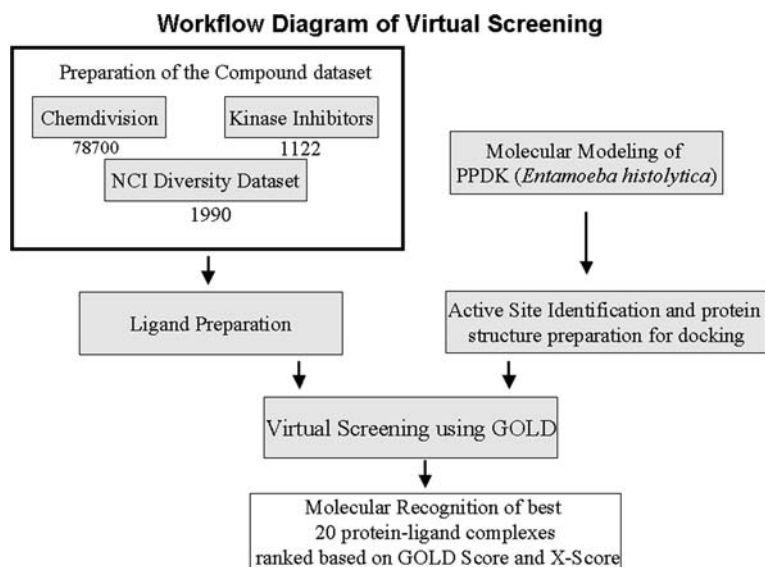
bootstrap values were manually calculated to 100 and viewed. We presume that this plot will help in drug targeting studies of pathogenic organisms, which are phylogenetically related to *Entamoeba histolytica* with reference to the protein PPDK.

Results and discussion

One major criterion of the modern drug discovery is the specificity and accuracy. The drug should never disturb any normal process of the host system. Our initial work in sequence analysis and domain study suggested to pick up PPDK, the key enzyme in the glycolytic pathway of *Entamoeba histolytica* and related pathogens involved in amoebic dysentery. Sequence as well as structural analysis was carried out to avoid any interference in the normal function of the human host system.

Pairwise sequence alignment study of PPDK, both in BLAST and Ensembl was performed against human database. We carried out domain and family analysis of PPDK in ProDom and Pfam while designing for the inhibitors of PPDK of *Entamoeba histolytica* to rule out the possibility of interfering with the identical domains of some proteins within human host. In ProDom database, there were two domains PD001289 (from position 26–372) and PD676442 (from position 28–401), similar to N-terminal part of PPDK of *Entamoeba histolytica*. There were another 13 domains in the ProDom which were similar to C-terminal region of the protein. These domains were not considered as they were not relevant to our study. However, the ProDom analysis of the N-terminal region exclusive to PPDK of *Entamoeba histolytica* indicated its involvement in the catalysis of the reversible conversion of ATP to AMP, pyrophosphate and phosphoenolpyruvate (PEP).

The pairwise sequence alignment of human pyruvate kinase isozymes, M1/M2 (P14618), with PPDK sequence of *Entamoeba histolytica*, showed 13.5% sequence identity and 24.4% sequence similarity. Whereas, with respect to human PK isozymes R/L (P30613), 12.6% identity and 22.2% sequence similarity was observed. Although, in general, sequence dissimilarity means possible dissimilar binding sites, however, exceptions where targets with significant sequence dissimilarity bind similar molecules do exist. The structural comparison of modeled protein PPDK of *Entamoeba histolytica* with human PK R/L (PDB ID: 1LIU-A chain) when carried out showed no significant structural similarity in the binding site region of PPDK as shown in Fig. 2. The RMSD value for equivalent C-alpha atoms between PPDK and PK R/L was 3.2 Å. The pairwise structural alignment had 242 equivalent residues and 223 gaps with 9.5 percentage of identity. Structural comparison carried out using CE program for A-domain of human PK

Fig. 2 Superimposed structure of modeled protein PPDK of *Entamoeba histolytica* with the Human substitute pyruvate Kinase isozyme(R/L). Active site residues of PPDK (Lys21—blue, Arg91—pink, Asp323—brown, Glu325—orange and Gln337—yellow) are depicted in the circle as dots view using Pymol

R/L (ATP-Grasp domain as classified in CATH— http://cathwww.biochem.ucl.ac.uk), consisting of two segments (from 57–85 and 429–573 residues) with A-domain of PPDK of *Entamoeba histolytica* (ATP-Grasp domain as classified by CATH for one of the best templates of PPDK in *Clostridium symbiosum*: PDB ID 1KBL used for modeling, consisting of two segments 2–109 and 199–243 residues), found only 10 residue-segment aligned. This result also reflected an absence of any significant structural similarity between the active site region of PPDK and PK R/L. The published literature added further confidence to use PPDK as the potential drug target in our study [10, 12, 13].

The atomic coordinates of PPDK for the organism *Entamoeba histolytica* were not available in Protein Data Bank [22], which necessitated developing a protein model. Homology modeling protocol was employed to predict the model of the protein. An identity of 51% with well-studied protein of *Clostridium symbiosum* (Resolution—1.94) provided a great strength for modeling the protein. Three dimensional structure prediction by comparative modeling was done by MODELLER8V2. Discover Module of INSIGHT II used in energy minimization embraces a range of force fields. Discover module is a molecular mechanics simulation environment offering energy minimization, dynamics and conformational search on molecular,

aggregate or periodic systems. Due to some unfavorable steric contacts we had to edit the raw protein manually, with out disturbing the critical residues. Simple minimization followed by a detailed minimization was done until we could find no significant variation in energy. Atomic charges were assigned using consistent valence force field (CVFF). The final model, which we took for further analysis, consisted of 884 amino acid residues.

We used both PROCHECK and the VERIFY3D softwares to check the quality of the modeled protein. Ramachandran Plot obtained from the program PRO-CHECK, which checks the stereochemical quality of a protein structures, producing a number of postscript plots, analyzing its overall and residue-by-residue geometry, assured the reliability of the modeled protein with 91.1% residues in most allowed region and 7.7% in additional allowed region. There were only 0.3% residues in disallowed region and 0.9% in generously allowed region. The residues in the disallowed region were, Gln417 and Glu663. The assessment with VERIFY3D, which derives a "3D-1D" profile based on the local environment of each residue, described by the statistical preferences for: the area of the residue that is buried, the fraction of side-chain area that is covered by polar atoms (oxygen and nitrogen), and the local secondary structure, also substantiated the reliability of the three dimensional structure. The residues that deviated from the standard conformational angles of Ramachandran plot were the members of C terminal domain of the protein. This was an ignorable condition since the C-terminal end was not critical in our study. The distance of these residues to the active site residues also were found to be more than 10 Å, which suggested that those residues would interfere little with the binding of ligands in the active site region of PPDK. The three domains were clearly differentiable as N-terminal domain, Central domain and C-terminal domain (Fig. 3). RMSD (c-alpha atom) of the modeled protein with respect to the templates (1KBL, 1GGO, 1JDE and 1DIK) were calculated using combinatorial extension (CE) algorithm available from San Diego Supercomputer Center [21] and found to have RMSD values of: 0.6, 0.6, 0.7 and 0.8 Å, for each of the templates, respectively.

Active site was identified with reference to the studies done on pyruvate phosphate dikinase in *Clostridium symbiosum* [16]. We carried out in silico studies to confirm these active sites, using LIGSITE^{csc} algorithm. The output from the LIGSITE^{csc} program showed coherent active sites for the target protein as documented in literatures [16]. Figure 4 shows that all five active site residues were located in the N terminal domain and the phosphate carrier histidine in the central domain. This helped us to define the binding pocket as 10 Å around Glu325, which included all other active site residues also. A multiple sequence
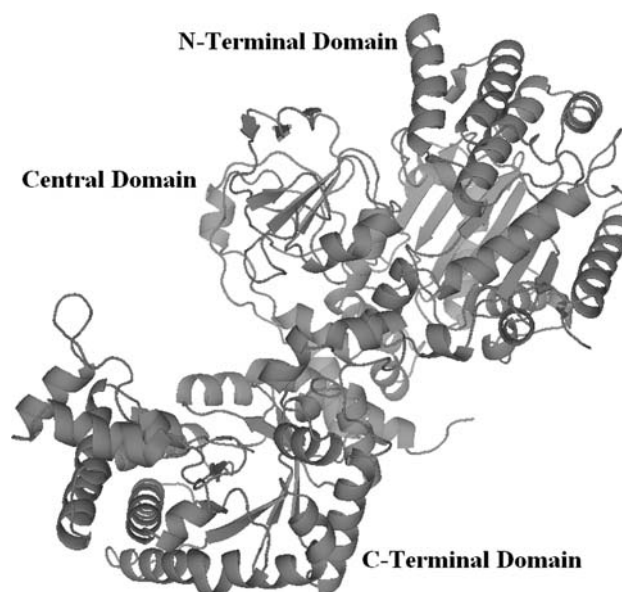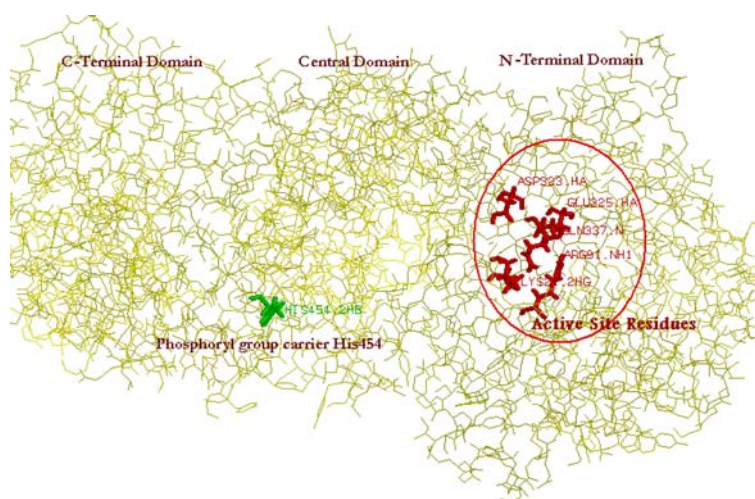

Fig. 3 Modeled three-dimensional structure of protein pyruvate phosphate dikinase of *Entamoeba histolytica*

alignment against the target sequence, highlighting the conserved active site region, is shown in the Fig. 5. The figure shows that in the course of evolution, the conserved regions, Lys22, Arg92, Asp321, Glu323, Gln335, in *Clostridium symbiosum* were replaced with Lys21, Arg91, Asp323, Glu325 and Gln337 in *Entamoeba histolytica*, respectively.

The best 20 ligands were chosen based on the scores obtained from the genetic docking algorithm GOLD and the consensus scoring program X-Score. The standard methodology accepted for virtual screening with the program GOLD involves, applying the parameters of library screening followed by the standard default setting. Library screening was done to filter the non-docked compounds from the three databases that we used for virtual screening. After filtering non-docked compounds, remaining compounds were used for detailed docking. Fifty genetic algorithm runs with standard default parameter settings were performed without early termination, after which 10 best solutions were kept for each ligand. Docking procedure consisted of three interrelated components; (a) identification of binding site, (b) a search algorithm to effectively sample the search space (the set of possible ligand positions and conformations on the protein surface) and (c) a scoring function. The GOLD fitness function consisted of four components: (a) protein–ligand hydrogen bond energy (external H-bond); (b) protein–ligand van der Waals (vdw) energy (external vdw); (c) ligand internal vdw energy (internal vdw); (d) ligand torsional strain energy (internal torsion). On the other hand, the scoring schema used in the software X-Score computes a binding score for

Fig. 4 Active site residues (red) in the N-terminal domain of pyruvate phosphate dikinase (*Entamoeba histolytica*)

Fig. 5 Multiple Sequence Alignment of protein pyruvate phosphate dikinase of *Entamoeba histolytica* with the templates (PDBIDs: 1KBL, 1DIK, 1JDE and 1GGO) used in the present study. The conserved residues in the active site region are represented in boxes, with corresponding residue number in *Entamoeba histolytica*

a given protein–ligand complex structure, and this binding score correlates to experimental binding constants well. It takes into account van der Waals interactions, hydrogen bonding, deformation penalty, and hydrophobic effects between the receptor and the ligand. X-Score does not perform the molecular docking process and it is applied in combination with docking programs [37] such as GOLD, FlexX, AutoDock and DOCK.

The best 20 ligands chosen with GOLD Fitness score proved their reliability in the X-Score (Table 2). From these ligands, top scoring three ligands predicted by GOLD docking program and corresponding X-Score are illustrated further in this paper. Out of the three databases screened, all the top 20 ligands happened to be the members of ChemDivision database. Further in the paper, name of the ligands are indicated by the ligand ID's given in the database.

Filter-the OpenEye scientific software, a molecular screening and selection tool that uses a combination of physical property calculations and functional-group knowledge to assess compound collections and quickly removes compounds with undesirable elements, functional groups, or physical properties, was used with the modified parameters (as mentioned in "Materials and methods"). This indicated the drug-likeness in the 16 of 20 predicted ligands, unlike in four such compounds, m13336, m25476, m34014 and m4711. The ligand, m13336, failed to pass the filtering due to the presence of undesirable functional group beta carbonyl quaternary nitrogen. The ligand, m25476, had maximum atom count of 41 instead of 40. The ligand, m34014, failed because of violation of Lipinski rule from 1 to 2. The ligand, m4711, exceeded the default enamine number of 0 to 1.

The conserved active site residues in the N-terminal domain of the protein established a binding pocket for ligands. Figures 6a, 7a show that the top scoring ligands were completely buried into the active site pocket of the protein. The geometry of protein binding with the ligand has a significant role to judge the specificity of ligands. The substantial number of hydrogen bonding and hydrophobic interactions (Figs. 6b, 7b) must enhance the binding potency of the inhibitor.

Table 2 Database ID number, chemical structure, IUPAC names and fitness scores of top scored ligands docked with pyruvate phosphate dikinase of *Entamoeba histolytica* using docking program GOLD
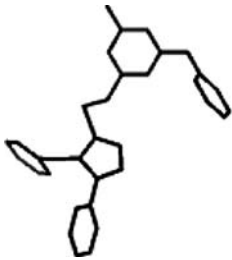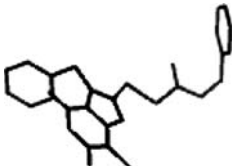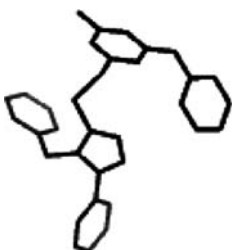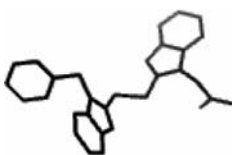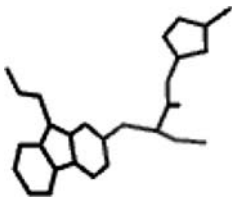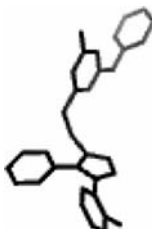
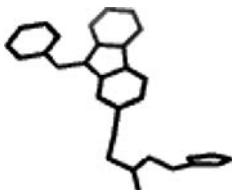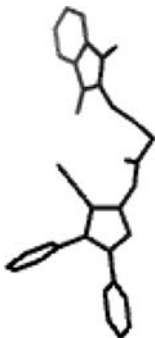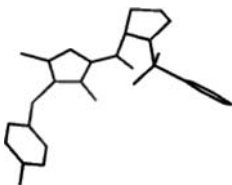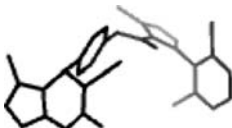| Lig ID no.	Ligand structure	Ligand name	GOLD Score	X-Score
m659*		[3-Methyl-7-(3-methyl-benzyl)-2,6-dioxo-2,3,6,7-tetrahydro-1H-purin-8-ylsulfanyl]-acetic acid benzyl ester	71.44	−7.83
m31721*		(7-Benzyl-3-methyl-2,6-dioxo-2,3,6,7-tetrahydro-1H-purin-8-ylsulfanyl)-acetic acid benzyl ester	70.28	−7.37
m3798^		6-(4-Benzyl-5-phenyl-4H-[1,2,4]triazol-3-ylsulfanylmethyl)-N-phenyl-[1,3,5]triazine-2,4-diamine	69.59	−7.11
m24061		2-[2-(1-Benzyl-1H-benzoimidazol-2-ylsulfanylmethyl)-benzoimidazol-1-yl]-acetamide	68.43	−7.64
M32392		2-(9-Allyl-9H-1,3,4,9-tetraaza-fluoren-2-ylsulfanyl)-N-(5-methyl-isoxazol-3-yl)-butyramide	67.62	−7.68
m3803^		6-[5-(2-Chloro-phenyl)-4-phenyl-4H-[1,2,4]triazol-3-ylsulfanylmethyl]-N-phenyl-[1,3,5] triazine-2,4-diamine	67.59	−8.27

Table 2 continued
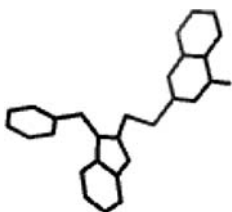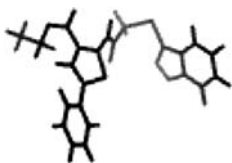
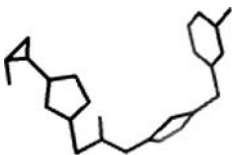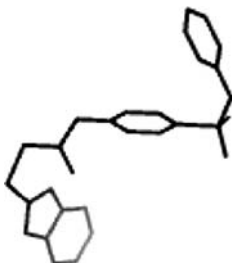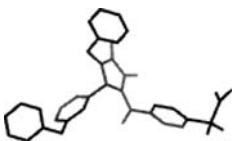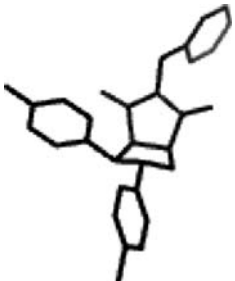| Lig ID no.	Ligand structure	Ligand name	GOLD Score	X-Score
M9237		2-(9-Benzyl-9H-1,3,4,9-tetraaza-fluoren-2-ylsulfanyl)- N-furan-2-ylmethyl-acetamide	66.80	−7.41
M26177[@]		N-(3-Cyano-4,5-diphenyl-furan-2-yl)-4-(1,3-dioxo- 1,3-dihydro-isoindol-2-yl)-butyramide	66.61	−8.25
M31469		(1-Benzenesulfonyl-pyrrolidin-2-yl)-(3,5-dimethyl-4- p-tolylsulfanyl-pyrazol-1-yl)-methanone	66.59	−7.94

m4711		3-(2,6-Dichloro-phenyl)-5-methyl-isoxazole-4-carboxylic acid
		    4-(6-amino-5-cyano-3-methyl-1,4-dihydro-pyrano[2,3-c]
		    pyrazol-4-yl)-phenyl ester
			66.52	−8.76

m33677		2-(1-Benzyl-1H-benzoimidazol-2-ylmethylsulfanyl)- 3H-quinazolin-4-one	66.33	−7.87
m19088		5-Phenyl-2-[2-([1,2,4]triazolo[4,3-a]pyridin-3-ylsulfanyl)- acetylamino]-thiophene-3-carboxylic acid ethyl ester	66.32	−7.60

Table 2 continued
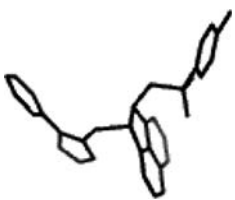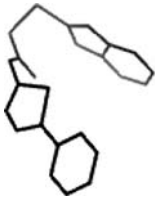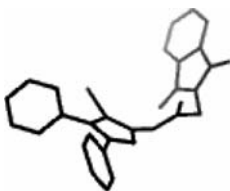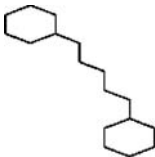
Lig ID no.	Ligand structure	Ligand name	GOLD Score	X-Score
m47425		N-[5-(3-Chloro-benzyl)-thiazol-2-yl]-3-[5-(2-methyl-cyclopropyl)-furan-2-yl]-propionamide	66.10	−7.83
m48312		2-(1H-Benzoimidazol-2-ylsulfanyl)-N-[4-(pyridin-2-ylsulfamoyl)-phenyl]-acetamide	65.74	−7.95
m25476		4-[4-Hydroxy-5-oxo-2-(3-phenoxy-phenyl)-1-pyridin-3-ylmethyl-2,5-dihydro-1H-pyrrole-3-carbonyl]-N,N-dimethyl-benzenesulfonamide	65.50	−8.62
m13336		5-Benzyl-2-(4-chloro-phenyl)-3-(4-fluoro-phenyl)-tetrahydro-pyrrolo[3,4-d]isoxazole-4,6-dione	65.16	−7.18
m34014		4-Chloro-N-[2-(1-phenyl-1H-tetrazol-5-ylsulfanyl)-acenaphthen-1-yl]-benzenesulfonamide	65.14	−8.88
m37394		2-(1H-Benzoimidazol-2-ylsulfanyl)-N-(5-phenyl-[1,3,4]thiadiazol-2-yl)-acetamide	65.12	−7.97

Table 2 continued

Lig ID no.	Ligand structure	Ligand name	GOLD Score	X-Score
m26175@		*N*-(3-Cyano-4,5-diphenyl-furan-2-yl)-2-(1,3-dioxo-1,3-dihydro-isoindol-2-yl)-acetamide	65.11	−7.67
m25578		2-Benzylsulfanyl-*N*-[5-(3,4-dichloro-benzyl)-[1,3,4]thiadiazol-2-yl]-acetamide	64.89	−8.38

The calculated binding energies using X-Score are given in separate column. Structurally similar molecules that are predicted by Jcsearch from Jchem version 3.2 are marked with identical symbols

The ligand m659 was a ring structure having a molecular weight of 450.521 amu and a log P value of 3.87. It showed the highest value in GOLD Fitness Score (Table 2). X-Score value also was reasonably good for this compound. There were three hydrogen bond interactions and various hydrophobic interactions (Table 3, Fig. 6b) established between the ligand and the protein. The atom zeta nitrogen of basic polar amino acid Lys21 formed hydrogen bond with the atom O13 of the ligand m659. The same way, the atom NH2 of basic polar amino acid Arg339 made hydrogen bond with the atom O12 of the m659. Meanwhile, the atom oxygen at epsilon-1 position of the acidic polar residue Glu325 formed a hydrogen bond with the atom N4 of m659. In the obtained hydrogen bond interactions, two hydrogen bonds were with the active sites residue of PPDK: i.e., Lys21 and Glu325, as proposed by us in this paper. The ligand, m31721, had a molecular weight of 436.494 amu with a second highest GOLD Fitness Score and X-Score of −7.37 kcal/mol (Table 3) in our study. The log P value of this compound was 3.38. Even though LIGPLOT did not calculate all hydrogen bond interactions (Fig. 7b) for m31721, the program GET-NEARRES could estimate four hydrogen bonds as listed in Table 3. In case of m31721, Lys21 and Glu325, hydrogen bond was formed as in the case of m659. In addition, neutral polar amino acid Asp323 with its delta oxygen formed hydrogen bond with atom N8 of ligand m31721; and basic polar amino acid Arg339 with the atom NH2 formed hydrogen bond with the atom N8 of the ligand m31721. The three of the residues (Lys21, Asp323, Glu325) involved in hydrogen bond formation have already been proposed by us in this paper to act as active sites. The ligand m3798 had a molecular weight 466.573 amu and the

third highest GOLD Fitness Score and a X-Score of −7.11 kcal/mol (Table 2). The log P of m3798 was 7.05. It formed three hydrogen bonds (Table 3) in the active site cleft, which included NH1 atom of basic polar amino acid Arg92 with N9 atom of the ligand m3798, delta sulfur atom of neutral polar amino acid Met242 with the N17 atom of the ligand m3798 and oxygen at epsilon position of acidic polar amino acid Glu325 with N1 atom of the ligand m3798. Although Sulfur is not a strong Hbond acceptor, to identify the interacting atoms between the protein and ligands, we obtained a reasonable distance to assume the formation of a Hbond with Sulfur atom of Met242 with N17 of ligand m3798, using the program GETNEARES. The two of the residues (Arg91 and Glu325) involved in hydrogen bond formation with ligand m3798 have already been proposed by us in this paper to act as active sites.

Interaction studies were carried out for more ligands which showed best scores in the program GOLD and X-Score, however, the molecules that showed higher X-Score than the already discussed ligands and did not pass through the OpenEye scientific software Filter is not documented and discussed in this paper. Table 3, nevertheless, provides information about the highest X-Score ligands (m34014, m4711, m25476) for reference.

Among the top ranking molecules how many ligand molecules are distinctly different (i.e., diversity set of molecules) as repeat testing of similar molecule is a waste of resource and time. Jcsearch available with Jchem software performs substructure, superstructure, exact, exact fragment, similarity and perfect searches as well as match counts on the specified query and target molecules. Similarity concept of Jcsearch in Jchem is based on hashed binary chemical fingerprints with Tanimoto metrics.
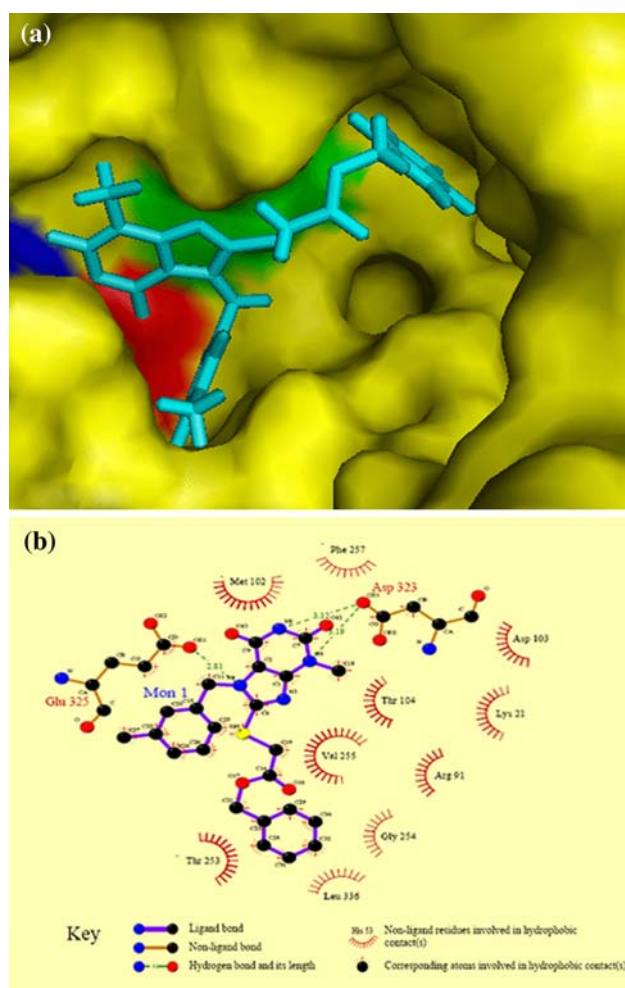
Fig. 6 Docked complex of ligand m659 in the active site of pyruvate phosphate dikinase of *Entamoeba histolytica*. (**a**) The protein is depicted in surface view and ligand m659 as stick in the binding pocket (Lys21 is colored red, Glu325 green and Arg339 as blue). (**b**) Schematic drawing of types of interactions of the ligands generated by the program Ligplot

Fig. 7 Interactions of ligand m31721 with in the active site of pyruvate phosphate dikinase of *Entamoeba histolytica*. (**a**) The protein is depicted in surface view and ligand m31721 as stick in the binding pocket (Lys21 colored red, Asp323 white, Glu325 green and Arg339 as blue). (**b**) Schematic drawing of types of interactions of the ligands with the protein generated by the program Ligplot

Similarity searching finds molecules that are similar to the query structure. The calculation applies the Tanimoto coefficient, which has two arguments—fingerprint of the query structure and the fingerprint of the molecule in the database. The Tanimoto coefficient is calculated by the formula: NA&B/(NA+NB-NA&B), where NA and NB are the number of bits set in molecule A and B, respectively. NA&B is the number of bits that are set in both. The similarity threshold specifies a lower limit for the Tanimoto coefficients. If a Tanimoto value is greater than the threshold, then the query structure and the given database structure are considered similar. Jcsearch program was used to check the similarity between each ligand against all the other top ranking molecules. In the listed ligands, m659 and m31721 showed 95% similarity. Ligands, m3798 and m3803, showed 80% similarity while m26177 and m26175 showed 90% similarity.
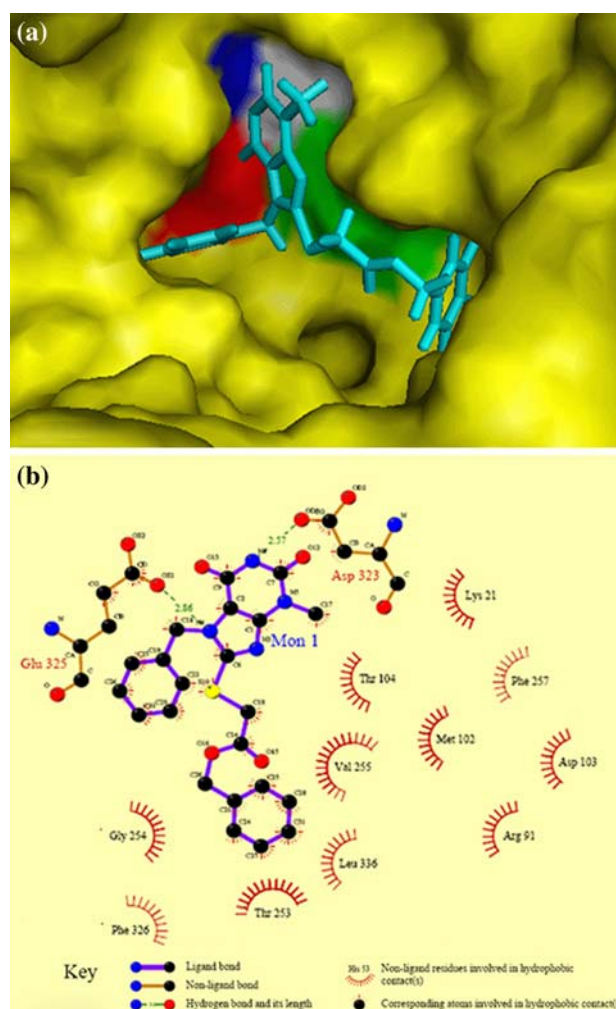
In this paper we used GOLD scoring function to estimate the putative ligands. In spite of high efficiency, GOLD has been blamed for poorly correlated docking scores and experimental binding affinities [45, 46] achieving R^2 values just over 0.5 [44, 47]. To substantiate the estimations done by the GOLD program, we used consensus scoring technique [47, 48]. The approach involved docking the protein with some search engine and primary score function (here GOLD algorithm and GOLD Fitness Score) and then re-scoring the output list with a secondary score function (here X-Score).

Since computational screenings always demand experimental testing in order to confirm the accurate drug molecule(s), the proposed LEAD molecules need to be optimized in further studies. The significance of this work is in providing a relatively inexpensive approach to screen

Table 3 Calculated Hydrogen bond interactions for top three ligands based fitness score of GOLD and consensus score of X-Score
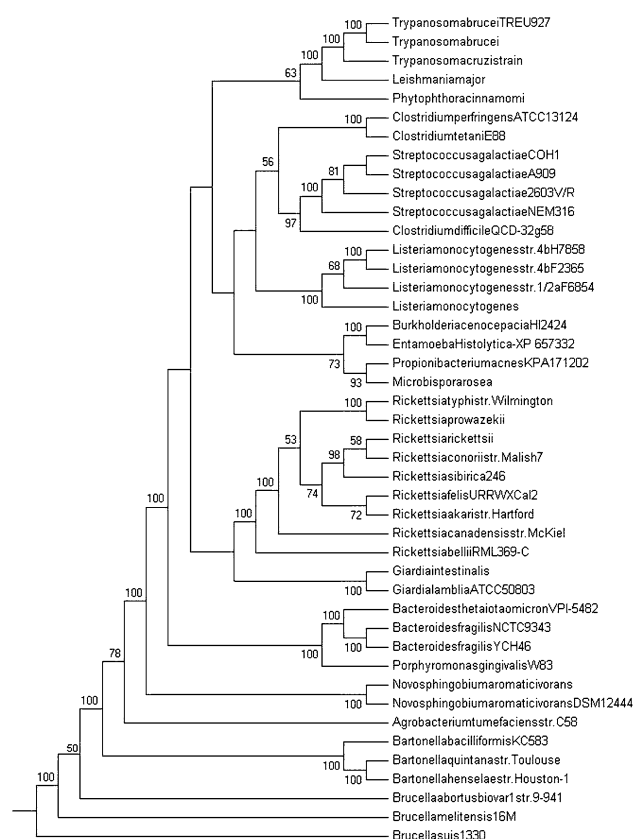
Ligands	A.A residues/number/atom name	Ligand atom no.	Distance (Å)
m659	Lys21 NZ	O13	2.41
	Glu325 OE1	N4	2.81
	Arg339 NH2	O12	2.81
m31721	Lys21 NZ	O13	2.27
	Asp323 OD1	N8	2.57
	Glu325 OE1	N4	2.86
	Arg339 NH2	N8	3.13
m3798	Arg91 NH1	N9	2.99
	Met242 SD	N17	3.15
	Glu325 OE1	N1	3.20
m13336	Ser92N	F26	2.65
	Asp103O	F26	2.48
	Thr104OG1	O12	2.67
	Arg339 NH2	Cl25	2.62
m4711	Arg91 NH1	O19	2.65
	Met102 O	N12	2.29
	Thr104 N	N13	3.02
	Gln277 OE1	Cl29	3.11
	Asp323 OD1	N18	2.99
	Phe326 N	Cl28	3.12
m25476	Arg91 NH1	O18	2.27
	Thr104 OG1	O17	2.63
	Gln277 OE1	O12	3.27
	Glu325 OE1	O15	2.12

Fig. 8 Phylogeny Tree based on pyruvate phosphate dikinase from different organisms generated using Phylip3.2: The numbers at the nodes indicate bootstrap values (in percentage) retrieved from 500 replicates

compounds that are likely to inhibit the action of PPDK in *Entamoeba histolytica*.

In addition, this study also searched for the pathogens with evolutionary proximity and genetic relatedness with respect to PPDK protein, whose growth could be controlled by the ligands proposed in this paper for *Entamoeba histolytica*. We propose that a similar study in silico could be designed for the pathogens within, Burkholderiacenocepacia, Propionibacteriumacnes and Microbiosporarosea, which cluster with *Entamoeba histolytica* (Fig. 8) with high confidence interval, to assess the feasibility of using the inhibitors defined here against their PPDK. Further, a detailed experimental study would be interesting to validate the utility of the proposed growth inhibitors as drug molecules for the human pathogen, *Entamoeba histolytica*.

Conclusion

In this paper, we propose probable chemical compounds, which could be tested to devise drug molecules to retard the hazardous proliferation of *Entamoeba histolytica* outbreaks in various parts of the world. Extensive literature study and sequence analyses were carried out to project PPDK as the potential drug target. In the absence of crystal structure, we used homology-modeling protocol to predict the three dimensional structure of the protein with very good template structures. Virtual screening was carried out, using the GOLD program, followed by a consensus scoring algorithm X-Score. The phylogenetic tree predicted, using the software Phylip3.2, was used to generate clusters of pathogens possessing feasible targets with the predicted ligands.

The scope of this work could be to use this data to do cost-effective experimental screening. The proposed potential chemical compounds could provide the prime lead for future experimental screening.

References

1. Walsh JA (1986) Rev Infect Dis 8:228
2. Spinella S, Levavasseur E, Petek F, Rigothier MC (1999) Eur J Biochem 266:170

3. http://www.who.int/entity/vaccine_research/documents/Parasitic_Diseases.pdf

4. Petri WA Jr (2003) Trends Parasitol 19:523

5. Samarawickrema NA, Brown DM, Upcroft JA, Thammapalerd N, Upcroft P (1997) J Antimicrob Chemother 40:833

6. Que X, Brinen LS, Perkins P, Herdman S, Hirata K, Torian BE, Rubin H, McKerrow JH, Reed SL (2002) Mol Biochem Parasitol 119:23

7. de Meester F, Shaw E, Scholze H, Stolarsky T, Mirelman D (1990) Infect Immun 58:1396

8. Byington CL, Dunbrack RL Jr, Whitby FG, Cohen FE, Agabian N (1997) Exp Parasitol 87:194

9. Reeves RE (1984) Adv Parasitol 23:105

10. McLaughlin J, Aley S (1985) J Protozool 32:221

11. Saavedra E, Encalada R, Pineda E, Jasso-Chavez R, Moreno-Sanchez R (2005) FEBS J 272:1767

12. Saavedra-Lira E, Perez-Montfort R (1996) Arch Med Res 27:257

13. Cosenza LW, Bringaud F, Baltz T, Vellieux FM (2000) Acta Crystallogr D Biol Crystallogr 56:1688

14. Wood HG, O'Brien EW, Micheales G (1977) Adv Enzymol Relat Areas Mol Biol 45:85

15. Herzberg O, Chen CC, Kapadia G, McGuire M, Carroll LJ, Noh SJ, Dunaway-Mariano D (1996) Proc Natl Acad Sci USA 93:2652

16. Ye D, Wei M, McGuire M, Huang K, Kapadia G, Herzberg O, Martin BM, Dunaway-Mariano D (2001) J Biol Chem 276:37630

17. Thrall SH, Dunaway-Mariano D (1994) Biochemistry 33:1103

18. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) J Mol Biol 215:403

19. Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM, Gilbert J, Hammond M, Herrero J, Hotz H, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Kokocinsci F, London D, Longden I, McVicker G, Melsopp C, Meidl P, Potter S, Proctor G, Rae M, Rios D, Schuster M, Searle S, Severin J, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C, Birney E (2005) Nucleic Acids Res 33:D447

20. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Nucleic Acids Res 31:3497

21. Shindyalov IN, Bourne PE (1998) Protein Eng 11:739

22. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C (2002) Acta Crystallogr D Biol Crystallogr 58:899

23. Sali A, Blundell TL (1993) J Mol Biol 234:779

24. Eswar N, John B, Mirkovic N, Fiser A, Ilyin VA, Pieper U, Stuart AC, Marti-Renom MA, Madhusudhan MS, Yerkovich B, Sali A (2003) Nucleic Acids Res 31:3375

25. Wu CH, Yeh L-SL, Huang H, Arminski L, Castro-Alvear J, Chen Y, Hu Z, Kourtesis P, Ledley RS, Suzek BE, Vinayaka CR, Zhang J, Barker WC (2003) Nucleic Acids Res 31:345

26. Karplus M, Petsko GA (1990) Nature 347:631

27. Hagler AT, Huler E, Lifson S (1974) J Am Chem Soc 96:5319

28. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283

29. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) J Mol Biol 7:95

30. Eisenberg D, Luthy R, Bowie JU (1997) Methods Enzymol 277:396

31. Huang B, Schroeder M (2006) BMC Struct Biol 6:19

32. Lipinski CA, Lombardo F, Dominy B, Wand Feeney PJ (1997) Adv Drug Delivery Rev 23:3

33. Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD (2002) J Med Chem 45:2615

34. CORINA Version 3.0, Molecular Networks GmbH, Erlangen, Germany (2004)

35. Tetko I.V, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin VA, Radchenko EV, Zefirov NS, Makarenko AS, Tanchuk VY, Prokopenko VV (2005) J Comput Aided Mol Des 19:453

36. Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997) J Mol Biol 267:727

37. Wang R, Lai L, Wang S (2002) J Comput Aided Mol Des 16:11

38. Csizmadia F (2000) J Chem Inf Comput Sci 40:323

39. Buntrock RE (2002) J Chem Inf Comput Sci 42:1505

40. Wallace AC, Laskowski RA, Thornton JM (1995) Protein Eng 8:127

41. Ewing TJ, Makino S, Skillman AG, Kuntz ID (2001) J Comput Aided Mol Des 15:411

42. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) Nucleic Acids Res 25:4876

43. Kolaczkowski B, Thornton JW (2004) Nature 431:980

44. Felsenstein J (1995) In: PHYLIP (Phylogeny Inference Package) Version 3.57c. Department of Genetics, University of Washington, Seattle, WA

45. Ferrara P, Gohlke H, Price DJ, Klebe G, Brooks CL 3rd (2004) J Med Chem 47:3032

46. Mohan V, Gibbs AC, Cummings MD, Jaeger EP, DesJarlais RL (2005) Curr Pharm Des 11:323

47. Wang R, Lu Y, Wang S (2003) J Med Chem 46:2287

48. Charifson PS, Corkery JJ, Murcko MA, Walters WP (1999) J Med Chem 42:5100