Indian Journal of Biochemistry & Biophysics
Vol. 59, September 2022, pp. 909-917
DOI: 10.56042/ijbb.v59i9.61006

Computational studies on new Leishmanial drug targets against Quercetin

Kanagaraj Jyothi[1]*, V Sivaranjani[1], U Pavithra[1], Sridhar Jayavel[2] & Lakshmanan Muthulakshmi[1]*

[1]Department of Biotechnology, School of Bio and Chemical Engineering, Kalasalingam Academy of Research and Education, Krishnankoil-626-126, Tamil Nadu, India

[2]Department of Biotechnology, Madurai Kamaraj University, Madurai-625 021, Tamil Nadu, India

Received 02 March 2022; *revised 26 August 2022*

Leishmaniasis, a parasitic disease caused by *Leishmania* parasite which resides in the infected sand flies. Control of Leishmaniasis remains a source of grave concern worldwide. Studies on Leishmaniasis triggered because of its outbreak in tropical and subtropical regions of Asia, East Africa and South America. There is an urgent need for new therapeutic interventions such as vaccine and new drug targets as it develops resistance towards the available drugs. Quercetin, a derivative of polyphenolic flavonoid exhibits various biological activities by interacting with proteins and nucleic acids. In this study, computational analysis was performed to identify the potential drug target of Quercetin in *Leishmania* species by molecular docking. The newly predicted targets were subjected for subcellular localization prediction and determined the protein-protein interaction networks that would aid in the development of anti-Leishmanial drugs. This study helps in the identification of targets and development of anti-Leishmanial drugs.

Keywords: Anti-leishmanial drug, Leishmaniasis, Quercetin

Leishmaniasis is a vector borne disease caused by *Leishmania* parasites which are transmitted by infected female phlebotomine sand flies. It is one of the most important neglected tropical diseases among third world countries and Indian subcontinent[1]. According to the World Health Organization (www.who.int), there are about 12 million people were infected by Leishmaniasis leading to 20,000 to 30,000 annual deaths[2]. Among all parasitic diseases, Leishmaniasis stands second in mortality and third in its morbidity with children below 15 years, after Malaria and Schistosomiasis[3]. More than 20 different species of *Leishmania* causes human infections. *L. donovani*, *L. major*, *L. Braziliensis* and *L. Mexicana* are the major causative agents of Leishmaniasis. Three major clinical manifestations of Leishmaniasis include cutaneous, sub-cutaneous and visceral are caused by *L. major*, *L. Braziliensis* and *L. donovani*, respectively. The most common form is the cutaneous Leishmaniasis which forms large lesions and opened wounds in body such as arms, legs and face[4]. Current strategies to control this disease are mainly based on chemotherapy in spite of their high toxicity, long duration of treatment and severe adverse reactions. The drugs used for the treatment of Leishmaniasis include pentavalent antimonials, pentamidine isethionate, amphotericin B *etc* are known to cause adverse side effects[5].

Since there is no vaccine available for any form of Leishmaniasis, control of this disease remains a source of serious concern worldwide. As most of the methods for Leishmaniasis treatment and control are of limited effectiveness, there is now an urgent need to identify the potent metabolites having multiple drug targets in Leishmania. Quercetin is one the important metabolite involved in the traditional formulations against Leishmania[6] but their exact targets were unknown. Deciphering knowledge on the metabolites with their targets would be the most efficient way for the control of this disease[7]. Hence, we are performing the study to find the targets of Quercetin among the available Leishmania protein structures. This task can be achieved by insilico analysis using various computational tools. Docking of metabolites into the binding site of a receptor and estimating the binding affinity of the complex is an important part of finding the druggable target. Protein subcellular localization (PSL) is important indicator to elucidate protein function as proteins cooperate towards a common function in the same subcellular

*Correspondence:
E-mail: drjyothisamy@gmail.com (KJ)
mthlakshmi27@gmail.com (LM);

compartment. It is also essential to annotate genomes, to design proteomics experiments and to identify potential diagnostic, drug and vaccine targets[8]. Protein-Protein Interaction (PPI) of entire proteome or between a group of proteins can be useful in identifying precise host pathogen interaction, thereby leading to the development of potential drug targets[9]. There are numerous experimental approaches that were applied to predict PPI, among them the yeast two hybrid method and affinity purification by mass spectrometry are the most common methods. However, it may not be feasible to all proteins and resulting in errors. Thus, to overcome these limitations, a number of computational methods are being used to study protein-protein interactions.

In this work, we investigate the interaction between Quercetin and proteins from*, L. major*, *L. braziliensis*, and *L. Mexicana* by using molecular docking studies. The viability of the metabolite interacting proteins as targets will be validated through functional protein association networks[10]. The results from this analysis were used to identify new drug targets, which can be used as potential targets for *Leishmania* drug development.

Materials and Methods

Data collection

The crystal structure of potential protein targets of various *Leishmania* species were downloaded from the protein data bank (https://www.rcsb.org./pdb). A total of 50 proteins of different *Leishmania* species were retrieved. The active site of the proteins was predicted using CASTp[11]. The 3D structure of Quercetin (CID: 5280343) was retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov).

Physicochemical properties analysis

Leishmania parasite target protein sequences were retrieved from the NCBI database in FASTA format. The collected sequence similarities, genera, and the identity of protein sequence were analysed using ProtParam and BLASTP tool. The amino acid composition, molecular weight, instability index and aliphatic index were analysed by the ProtParam software. Multiple sequence alignment and analysis were performed using Clustal Omega. The NCBI Conserved domain database was used for the identification of characteristic conserved domains among the protein sequences. The evolutionary relationship between protein sequences were confirmed by phylogenetic tree construction.

Secondary structure Prediction

The self-optimized prediction method with alignment (SOPMA) was used to predict the secondary structure of protein and related conformation. The result was confirmed based on the presence of helix motif in the amino acid sequence.

Sub cellular localization prediction (PSL)

Sub cellular localization of selected protein targets of *Leishmania major* was performed using Loctree3[12]. The improved version of LocTree3 adds a module for inferring localization information from experimentally annotated sequence homologs using PSI-BLAST. In the absence of significant PSI-BLAST hits, results from LocTree2 are used.

Homology modelling

The three dimensional structures of known proteins obtained from the Protein Data Bank (PDB) were used as templates for homology modelling of these protein determinants. Automated homology modelling of these protein determinants was performed using the SWISS-MODEL (https://swissmodel.expasy.org/). The theoretical models of these protein determinants were obtained based on the structural alignment of the amino acid sequence. The active site was predicted using CASTptool (Computed Atas of Surface Topography of proteins).

Molecular docking

To understand the protein-ligand interactions, docking studies were performed using AutoDockVina 4.2 in Ubuntu workstation. The protein and ligand were prepared using AutoDock tools. The protein preparation involves addition of polar hydrogens, united atom Kollman charge in the virtual environment. The ligand Quercetin was prepared by adding polar hydrogens, Gasteiger charge, detect flexible torsions and setting the number of torsions. The grid box was described based on the size of active site of each protein. Exhaustiveness factor of 10 is given to obtain more consistent docking result. Multiple conformations were retained and interactions were analysed manually and best pose with lowaffinity was selected. Protein-ligand docking studies were carried out based on the crystal structures of verified *Leishmania* protein drug targets.

Protein –protein interaction network (PPI Network)

A Protein-Protein Interaction network search was carried out for 31 target proteins of *Leishmania major* to retrieve functional protein association networks using STRING11.0[10].

Topological analysis of PPI network

The Protein-Protein Interaction network was visualized and analysed using Cytoscape 3.7.1 based on various parameters such as betweenness centrality, closeness centrality, degree, neighbourhood connectivity and topological coefficient. Nodes represent the protein residues and edges represent the interactions between the nodes. The degree indicates the number of edges linked to nodes; the highest degree of nodes represents significant biological functions. Betweenness centrality defines the importance of nodes based on number of shortest paths that pass through each node.

Results

Data collection

The 3D structure of Quercetin retrieved from PubChem is used in this study. A total of 50 *Leishmania* proteins were retrieved from PDB. The potential target proteins are grouped under *L. major*, *L. mexicana and L. Braziliensis* to cover all forms of Leishmaniasis. These proteins were used as a data set for further analysis. The mutant proteins and proteins with modified residues were eliminated from the dataset, resulting in a final dataset consisting of 50protein structures.

Primary structure analysis of *Leishmania* protein sequences

The physicochemical and biological properties of the proteins were computed using ProtParam tool. Table 1indicates the predicted physicochemical properties of retrieved proteins from NCBI. The observed results indicates that the stability of protein and wide range of pI values due to the variation in the presence of amino acids.

Secondary structure analysis of *Leishmania* protein sequences

Secondary structures were predicted using self-optimized prediction method (SOPMA) for the set of aligned proteins. Variations of alpha and beta helices were determined from the observed results (Table 2).

BLAST P analysis

Protein BLAST results determine the similarities of sequence and number of alignments expected by chance was identified. Table 3 shows the significant alignment with obtained E-Value. Among these the first two proteins are having high E-value with low significance comparatively other four protein sequence is closer to zero value indicates the more significant value obtained.

Table 1 — Predicted physicochemical properties of proteins

Properties	Protein ID					
	XP_001680815.1	1XTP_A	XP_001683287.1	4JZB_B	3KFL_A	XP_001686974.1
Number of amino acids	225	254	362	361	564	437
Theoretical pl	5.79	5.91	5.51	5.43	6.10	5.72
Stability index value	52.48	29.30	25.59	25.26	30.14	26.66
Aliphatic index value	73.78	76.81	85.97	86.20	80.35	88.40
GRAVY value	-0.506	-0.433	-0.090	-0.084	-0.320	-0.152

Table 2 — Predicted secondary structure

Protein Sequence ID	Number of β-strands (%)	Number of α-helices (%)	Others (%)
XP_001680815.1	20(8.93%)	48(21.43%)	156(69.64%)
1XTP_A	15(5.91%)	116(45.67%)	123(48.43%)
XP_001683287.1	10(2.76%)	237(65.47%)	115(31.77%)
4JZB_B	10(2.76%)	233(64.36%)	119(32.87%)
3KFL_A	31(5.50%)	260(41.10%)	273(48.4%)
XP_001686974.1	31(7.09%)	194(44.39%)	212(40.51%)

Table 3 — Blastp Analysis

Protein Sequence ID	Number of β-strands (%)	Number of α-helices (%)	Others (%)
XP_001680815.1	20(8.93%)	48(21.43%)	156(69.64%)
1XTP_A	15(5.91%)	116(45.67%)	123(48.43%)
XP_001683287.1	10(2.76%)	237(65.47%)	115(31.77%)
4JZB_B	10(2.76%)	233(64.36%)	119(32.87%)
3KFL_A	31(5.50%)	260(41.10%)	273(48.4%)
XP_001686974.1	31(7.09%)	194(44.39%)	212(40.51%)

Phylogenetic tree

Evolutionary relationship between the selected proteins was identified using phylogenetic tree construction. Figure 1 shows the constructed phylogenetic tree. The relationships between the related proteins are highlighted with the constructed tree.

Multiple sequence alignment

Multiple sequence alignment and Motif analysis of protein sequences was performed using Clustal Omega. The comparative structure of sequences was identified from the observed multiple sequence alignment. The selected proteins are aligned in a similar length represented in (Fig. 2).

Molecular docking

The 3D structure of Quercetin was docked to the active siteof target protein and their affinity was observed on the basis of their binding energy. Thus, Quercetin ligand was docked against 65 proteins of *Leishmania* species using AutodockVina version 4.2. The distribution of ligand binding affinity to various protein targets is provided as (Fig. 3A & B). The top twelve proteins with high affinity are listed in (Table 4). Proteins having binding energy within the range of −11 to −10 kcal/mol are further investigated and three protein targets from each *Leishmania* species were consider for protein-ligand interaction studies.

Protein-ligand interaction study

The PDB Id. 3M3I is a putative zinc finger protein family member, from *Leishmania major*. This protein is predicted to be involved in nucleotide metabolism and metabolic shuttling. Binding site residues of a protein are catalytic residues which play an important role in the function of a protein[13]. It gives the information about the protein-ligand interaction. This protein is a homo2-mer protein in which Quercetin binds to the A chain of the protein. His 25 is the predicted hydrogen bond forming residue. Molecular docking result of 3M3I was obtained as −11.6 kcal/mole. Quercetin binds to the hypothetical protein (3M3I) and the docked confirmation (Fig. 3C) has the lowest energy with high affinity and hence stable.

Subcellular localization prediction

LocTree analysis predicted the protein residing organelle within the cell. The reliability scores are ranging from 0 to 100 from LocTree indicates the accuracy of prediction. A protein with a score of 100 means most dependable projection for subcellular localization. The results of localization prediction of proteins are shown in (Table 5). The major residing organelles include Cytoplasm, Mitochondria, Peroxisome, Golgi apparatus, Nucleus, Secreted proteins and others as shown in (Fig. 4). As per the prediction majority of the target proteins (47%) are found to be localized in Cytoplasm. Another 17% of proteins are found to be expressed in Mitochondria.

Protein – Protein Interaction network:

Using an interaction score threshold of 0.4, the STRINGPPI analysis (Fig.4) hasyielded a clustered

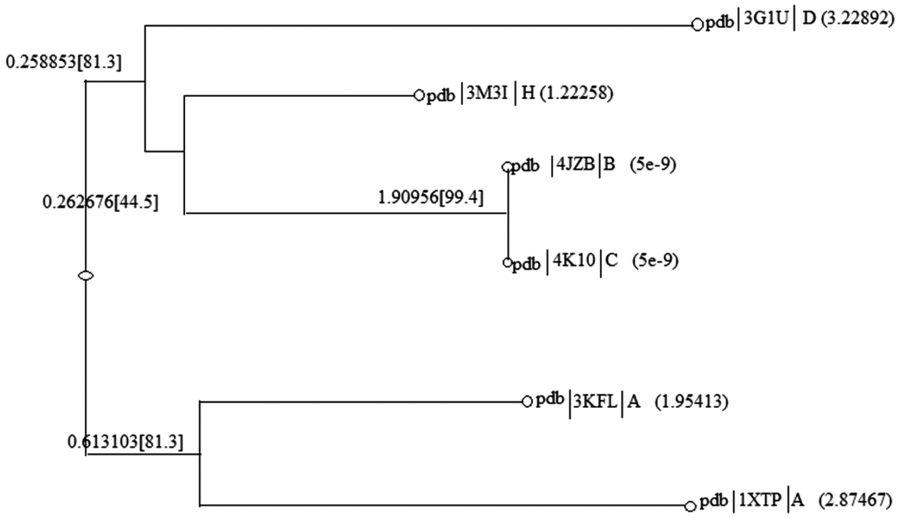

Fig. 1 — Phylogenetic tree relationship of *Leishmania* sp

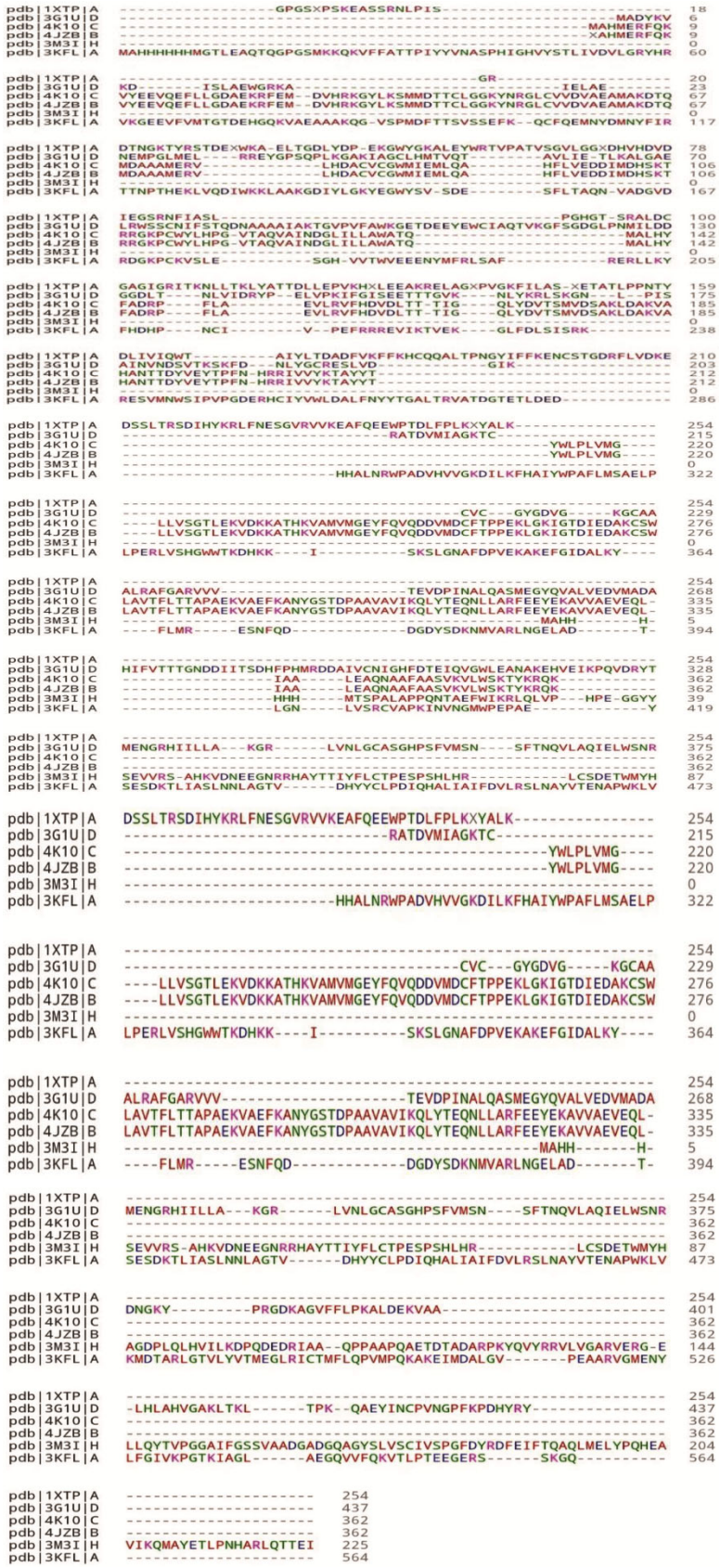

Fig. 2 — Multiple sequence analysis of *Leishmania* sp

network with clustering coefficient 0.308 containing 31 nodes with 14 edges (expected number of edges:4), indicating significantly more interactions than expected (PPI enrichment p value less than 0.001)[14]. Among the 31 proteins, 12 proteins have strong interactions with an average interaction score of above 0.4. Interestingly, half of the proteins having strong interactions were found to reside in the cytoplasm as shown in (Table 6). Hence, targeting the cytoplasmic proteins directly inhibits the survival of *Leishmania.*

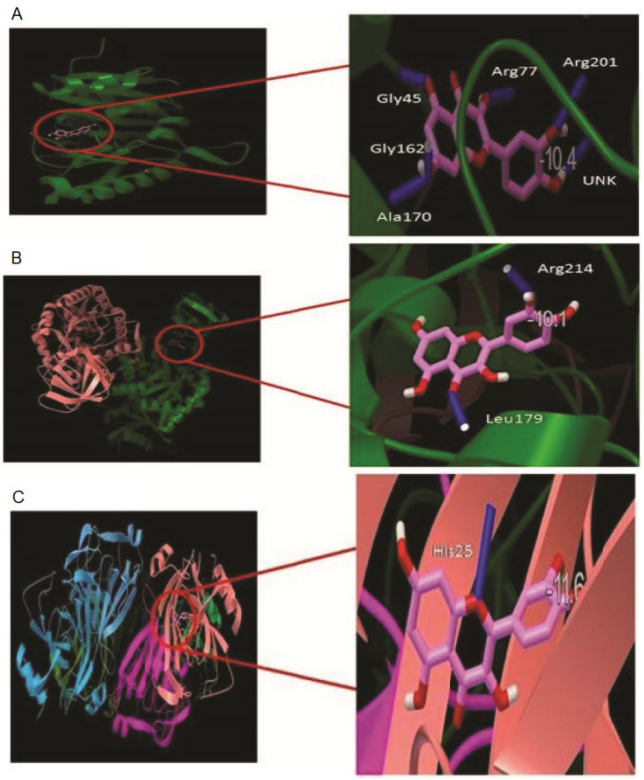

Fig. 3 — The Interaction between A chain of (A) 3M3I; (B) 3CSS; and (C) 3M3I (cartoon form) and Quercetin (Stick representation). The hydrogen bonds are represented as blue cylinders showing hydrogen bond interaction in the active site of protein

Topological analysis of PPI network

The network obtained from STRING 11.0 was visualized and analysed using Cytoscape 3.7.1 Plugin Network Analyser. The PPI network for *L. major* is shown in (Fig. 5). In this study, nodes with higher degree, BC, CC, NC, and TC were used as key parameters to analyse the network. Higher the degree and betweenness value of a protein, higher is its significance. The protein, Acyl Carrier Protein (ACP) (PDB id-2M5R) has the highest degree value of 5 and betweeness centrality value of 0.62 with the highest

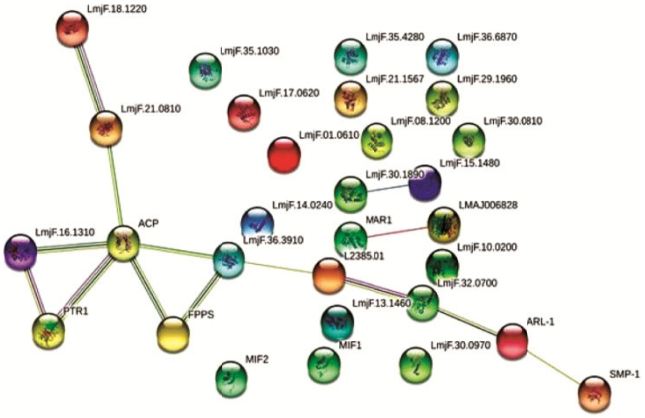

Fig. 4 — STRING PPI network connectivity of *L. major* proteins with 31 nodes, 14 edges and confidence score threshold value 0.4

Table 4 — List of highly interacting proteins and their binding energies against Quercetin

S. No	PDB ID	Name	Organism	Stoichiometry	Active site residues forming H- bond	Binding Energy (kcal/mol)
1	3m3i	Hypothetical protein	*L. major*	Homo 2-mer	His 25	−11.6
2	3g1u	S adenosylhomocystene hydrolase	*L. major*	Homo2-mer	Not found	−11.4
3	1xtp	SAM dependent methyl tranferase	*L. major*	Monomer	Not found	−11.3
4	4k10	Farnesyldiphosphate synthase with complex	*L. major*	Homo 2- mer	Not found	−10.8
5	4jzb	Farnesyldiphosphate synthase	*L. major*	Homo 2- mer	Tyr 211	−10.6
6	3kfl	Methionyl t-RNA synthase	*L .major*	Monomer	Trp 516	−10.5
7	3css	6 Phosphoglucanolactonase hydrolase	*L. Braziliensis*	Monomer	Ala 170,Gly 162,Arg 201,Gly 45,Arg 77	−10.4
8	4jzx	Farnesyldiphosphatesynthase with complex	*L. major*	Homo 2-mer	Lys 207, Tyr 211	−10.2
9	1pkl	Leishmania pyruvate kinase	*L. mexicana*	Homo 4-mer	Leu 179,Arg 214	−10.1
10	2r8q	Class 1 phosphodiesterase PDEB1	*L. major*	Homo 2-mer	Ser 747,His 756,Ser 759	−10.1
11	4iu5	Arginase in complex	*L. mexicana*	Homo 3-mer	Not found	−10.1
12	3ch7	6 Phosphoglucanolactonase	*L. braziliensis*	Monomer	Gly 45,Gly 162,His168,Ala 170,Arg 77	−10.1

Table 5 — Results of subcellular localization prediction of selected proteins that belongs to *L. major*

S. No	PDB ID	Organelle	Score	Expected Accuracy (%)	Annotation
1	3m3i	cytoplasmic protein	22	84	Loctree2
2	3g1u	Cytoplasmic protein	97	98	PSI-Blast
3	1xtp	Cytoplasmic protein	22	84	PSI-Blast
4	4k10	Cytoplasmic protein	29	85	PSI-Blast
5	3kfl	Mitochondrial protein	26	84	PSI-Blast
6	2r8q	Cytoplasm	17	82	PSI-Blast
7	2qt8	Mitochondrial protein	36	87	PSI-Blast
8	2m5r	Mitochondrial protein	40	88	PSI-Blast
9	3ha4	cytoplasm protein	13	81	Loctree2
10	5xil	Cytoplasmic protein	36	87	PSI-Blast
11.	1yqf	Mitochondrial protein	16	82	PSI-Blast
12	5via	Secreted protein	83	96	Loctree2
13	5l2r	Cytoplasmic protein	51	90	Loctree2
14.	5oez	Cytoplasm protein	34	87	PSI-Blast
15.	4dy9	Mitochondrial protein	83	96	PSI-Blast
16	2bfo	Peroxisome	15	81	PSI-Blast
17	5l4n	Cytoplasmic protein	15	81	PSI-Blast
18	1e92	Peroxisome	15	81	PSI-Blast
19	5oey	cytoplasm protein	34	87	PSI-Blast
20	1yf9	Cytoplasmic protein	38	88	PSI-Blast
21	5ofu	Cytoplasm protein	34	87	PSI-Blast
22	1y1x	Cytoplasmic protein	18	82	PSI-Blast
23	5yq8	Cytoplasmic protein	39	88	PSI-Blast
24	3h4v	Peroxisome	15	81	PSI-Blast
25	5ijm	-	-	-	Not found
26	1y63	Nucleus	33	86	PSI-Blast
27	2x77	Golgi apparatus	45	89	PSI-Blast
28	4ckm	Cytoplasm	6	80	PSI-Blast
29	3pg1	Nucleus	24	84	PSI-Blast
30	3fwt	Secreted	12	80	PSI-Blast
31	3fwu	Secreted	12	80	PSI –Blast
32	3ksv	Nucleus	45	89	PSI-Blast
33	3b64	Cytoplasm	14	81	PSI-Blast
34	2ar1	Nucleus	37	88	PSI-Blast
35	1xn4	Secreted protein	25	84	PSI-Blast

interaction score of 0.846 from STRING PPI network. Interestingly, the docking predicted binding affinity value for this protein falls in high affinity range.

Discussion

The observed GRAVY values denote the hydrophobicity value of peptide present in the sequence. The obtained GRAVY value mostly occurs in negative, which is indicates the hydrophilic nature of proteins. Based on the ligand affinity value, the proteins are classified in to three groups. The ligand affinity value ranging from −12 to −8 kcal/mol is defined as high affinity. The ligand affinity value ranging from −7.9 to −5 kcal/mol is defined as moderate affinity and the ligand affinity value less than −5 kcal/mol is considered as low affinity. The 6-phosphoglucanolactonase protein (PDB ID: 3CSS) from *Leishmania braziliensis*, which involves in carbohydrate metabolic process and pentose phosphate pathway. The protein may also participate in multiple interactions since it possesses multiple binding residues[15]. This is a monomer protein where Quercetin binds to the A chain of the protein. The hydrogen bond forming residues are Ala 170, Gly 162, Arg 201, Gly 45 Arg 77. The molecular docking result of 3CSS was obtained as −10.4 kcal/mole. The binding of quercetin to 6-phospho glucanolactonase active site is shown in the result. The docked confirmation has high affinity. The protein Pyruvate kinase (PDB id. 1PKL,) from *Leishmania mexicana*,

Table 6 — List of *L. major* drug targets found in PPI that are residing in cytoplasm

S. No	PDB ID	Protein Name	Interaction Score	Interacted Protein
1	3g1u	Adenosylhomocysteinase	0.607, 0.557, 0.505.	FPPS, 3-oxoacyl-acyl carrier protein synthase ii, putative, FBP protein.
2	4k10	Farnesyl pyrophosphate synthase	0.780, 0.607	ACP, Adenosylhomocysteinase
3	2r8q	cAMP specific phosphodiesterase, putative	0.65	Adenylate kinase isoenzyme 6 homolog
4	5xil	Bifunctionalaminoacyl-tRNAsynthetase, putative	0.744	methionyl-tRNAsynthetase
5	5oey	FBP protein	0.682, 0.557	Ubiquitin carrier protein 4, putative, Adenosylhomocysteinase
6	1yf9	Ubiquitin carrier protein 4, putative	0.682, 0.507	FBP Protein, ADP-ribosylation factor-like protein.

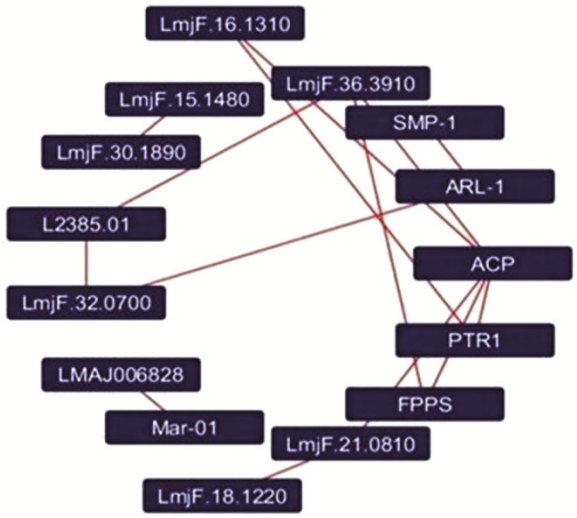

Fig. 5 — Visualization of protein network using Cytoscape 3.7.1 (Group attribute layout- Number of undirected edges)

involves in glycolysis metabolic pathway[16]. This is a homo-4 mer and quercetin binds to the E chain of the protein with low binding energy of -10.1kcal/mole. The hydrogen bond forming residues of Pyruvate kinase are Leu 179, Arg 214. The *Leishmania* protein targets investigated in this study are not known for their cell localization. So, it is very important to predict the subcellular localization of these target protein. The prediction of protein subcellular localization (PSL) focuses on determining localization sites of unknown proteins in a cell. This study of PSL helps in elucidating protein functions involved in various cellular processes and identifying drug targets[17].

Conclusion

In this present work, we identified the potential drug targets in *Leishmania* for Quercetin using molecular docking. Docking studies revealed that the protein 3M3I (Putative zinc finger protein family member) of *L. major*, 3CSS (6-phospho glucanolactonase) of *L. brazilliensis* and 1PKL (Pyruvate kinase) of *L. Mexicana* have the highest binding affinities and interaction. The protein - protein interaction network results indicated that the Acyl Carrier Protein (ACP) has the highest interaction score which forms super central hub by connecting other proteins. It also indicates that majority of the interacting proteins of PPI network were residing in cytoplasm. Hence by targeting the network of proteins in cytoplasm, the metabolism of *Leishmania major* can be arrested. Overall, this study revealed new drug targets of Quercetin along with the inhibited networks from *Leishmania* that can be developed as potential drug targets.

Acknowledgment

The authors are grateful to Kalasalingam Academy of Research and Education India for providing facilities and support to carry out the work.JS thank RUSA phase 2.0 for sponsoring “Centre of Excellence in Computational Intelligence and Data Analytics” in Madurai Kamaraj University, Madurai.

Conflict of interest

All authors declare no conflict of interest.

References

1 Diksha K, Summaya P, Rashmi S & Kuljith S, Advancement in leishmaniasis diagnosis and therapeutics: An update. *Eur J Pharmacol*, 910 (2021) 174436.

2 Gize A, Workineh A & Hailu T, A trend prevalence of visceral Leishmaniasis in West Armachiho District, Amhara Region, Northwest Ethiopia. *Trop Dis Travel Med Vaccines*, 6 (2020) 1.

3 Verjee MA, Schistosomiasis: Still a cause of significant morbidity and mortality. *Res Rep Trop Med*, 10 (2019) 153.

4 Abamor ES, A new approach to the treatment of leishmaniasis: Quercetin-loaded polycaprolactone nanoparticles. *J Turkish Chem Soc*, 5 (2018) 1071.

5 Brindha J, Balamurali MM & Kaushik C, An overview on the therapeutics of neglected infectious diseases-Leishmaniasis and Chagas diseases. *Front Chem*, 9 (2021) 286.

6 Bahare S, Laura M, Lianet M, Javad SR, Shahira ME, Mohamed AS, Rana MM, Nihal MEI, Ceyda SK, Oksana S, Mehdi SR, Farukh S, Natália M, Miquel M & William CC, Therapeutic potential of quercetin: new insights and perspectives for human health. *ACS Omega*, 5 (2020) 11849.

7 Rebolledo GAG, Jonas SD & Arellanes MAJ, Natural compounds and extracts from Mexican medicinal plants with anti-leishmaniasis activity: An update. *Asian Pac J Trop Med*, 10 (2017) 1105.

8 Hajialibeigi A, Amani J & Gargari SLM, Identification and evaluation of novel vaccine candidates against *Shigella flexneri* through reverse vaccinology approach. *Appl Microbiol Biotechnol*, 105 (2021) 1159.

9 Farooq QA, Shaukat Z, Aiman S, & Li CH, Protein-protein interactions: Methods, databases, and applications in virus-host study. *World J Virol*, 10 (2021) 288.

10 Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julein P, Roth A, Simonovic M & Bork P, STRING 8-a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res*, 37 (2009) D412.

11 Tain W, Chen C, Lei X, Zhao J & Liang J, CASTp 3.0: computed atlas of topography of proteins. *Nucleic Acids Res,* 2 (2018) W363.

12 Almagro Armenteros JJ, Sonderby CK, Sonderby SK, Nielsen H & Winther O, DeepLoc: Prediction of protein subcellular localization using deep learning. *Bioinformatics*, 33 (2017) 3387.

13 Tseng YY & Li WH, Evolutionary approach to predicting the binding site residues of a protein from its primary sequences. *Proc Natl Acad Sci*, 108 (2011) 5313.

14 Vykoukal J, Sun N, Bonavides CA,Katayama H,Tanaka I, Fahrmann JF, Capello M, Fujimoto J, Aguilar M, Wistuba II, Taguchi A, Ostrin EJ & Hanash SM, Plasma derived extracellular vesicle proteins as a source of biomarkers for lung adenocarcinoma. *Oncotarget*, 8 (2017) 95466.

15 Horvath A, Miskei M, Ambrus V, Vendruscolo M & Fuxreiter M, Sequence-based prediction of protein binding mode landscapes. *PLoS Comput Biol*, 16 (2020) 1.

16 Rigden D J, Phillips SE, Michels PA & Gilmore LAF, The structure of pyruvate kinase from Leishmania mexicana reveals details of the allosteric transition and unusual effector specificity. *J Mol Biol*, 291 (199) 745.

17 Anubha D & Usha C, Subcellular localization of proteins. *Arch Appl Sci Res*, 3 (2011) 392.