Crystal structure solution of phosphoglucomutase from *L.major* at 3.5 Å resolution and pharmacophore based data-mining to identify prospective anti-leishmanials

By

Barnali Waugh

LIFE05200804001

Saha Institute of Nuclear Physics

A thesis submitted to the

Board of Studies in Life Sciences

In partial fulfillment of requirements for the

Degree of

DOCTOR OF PHILOSOPHY ***of***

HOMI BHABHA NATIONAL INSTITUTE

February, 2015

Version approved during the meeting of Standing Committee of Deans held during 29-30 Nov 2013

HOMI BHABHA NATIONAL INSTITUTE[1]

RECOMMENDATIONS OF THE VIVA VOCE COMMITTEE

As members of the Viva Voce Committee, we certify that we have read the dissertation prepared by Barnali Waugh entitled: '**Crystal structure solution of phosphoglucomutase from *L.major* at 3.5 Å resolution and pharmacophore based data-mining to identify prospective anti-leishmanials**' and recommend that it may be accepted as fulfilling the thesis requirement for the award of Degree of Doctor of Philosophy.

Chairman –	Prof. Parthasarathi Mitra	P Mitra	**Date:** 16.6.15
Guide / Convener –	Prof. Rahul Banerjee	R Banerjee	**Date:** 16/6/15
Examiner –	Prof. Rajan Sankaranarayanan	R Sankaranarayanan	**Date:** 16/6/15
Member 1-	Prof. Udayaditya Sen	Udayaditya Sen	**Date:** 16/6/15
Member 2-	Prof. Partha Saha	Partha Saha	**Date:** 16/6/15
Member 3-	Prof. Dhananjay Bhattacharyya	Dhananjay Bhattacharyya	**Date:** 16/6/2015

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to HBNI.

I/We hereby certify that I/we have read this thesis prepared under my/our direction and recommend that it may be accepted as fulfilling the thesis requirement.

Date: 16/6/15

Place: SINP, Kolkata

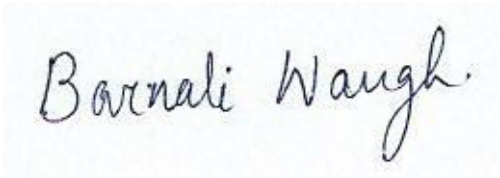
R Banerjee
Guide

[1] This page is to be included only for final submission after successful completion of viva voce.

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at Homi Bhabha National Institute (HBNI) and is deposited in the Library to be made available to borrowers under rules of the HBNI.

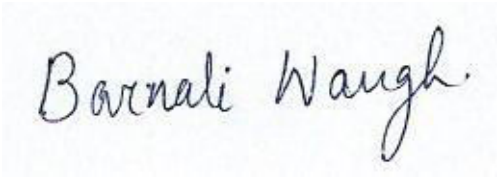
Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the Competent Authority of HBNI when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.



Barnali Waugh

DECLARATION

I, hereby declare that the investigation presented in the thesis has been carried out by me. The work is original and has not been submitted earlier as a whole or in part for a degree / diploma at this or any other Institution / University.



Barnali Waugh

List of Publications arising from the thesis

Journal

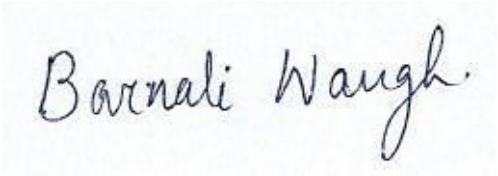
1. **In silico work flow for scaffold hopping in Leishmania**.
   Barnali Waugh, Ambarnil Ghosh, Dhananjay Bhattacharyya, Nanda Ghoshal, Rahul Banerjee. BMC Research Notes, 2014, 7: 802.

In preparation

1. **Crystal structure solution of phosphoglucomutase from *L.major* at 3.5 Å resolution**
   Barnali Waugh, Udayaditya Sen, Rahul Banerjee.

Conferences

1. **Structure of phosphoglucomutase from *L. major* at 3.5 angstrom resolution.**
   Barnali Waugh, Udayaditya Sen, Rahul Banerjee
   Annual Conference of Indian Biophysical Society entitled: Molecular Architecture, Dynamics and Assembly in Living Systems. February 7-10, 2014
2. **Drug targets in *Leishmania spp.***
   Barnali Waugh and Rahul Banerjee
   Kolkata international school cum conference on systems biology (KOLSYSBIO). December 29-31, 2012.



Barnali Waugh

DEDICATIONS

This thesis work is dedicated to my family whose constant support, guidance and love made this work possible as well as enjoyable.

ACKNOWLEDGEMENTS

The author is grateful to her supervisor, Prof. Rahul Banerjee for his kind guidance. She thankfully acknowledges Prof. Dhananjay Bhattacharyya, Dr. Ambarnil Ghosh and Prof. Nanda Ghosal, IICB. She would like to express her gratitude to Prof. Udayaditya Sen for his kind guidance and insightful suggestions. She is also grateful to Dr. R. D. Makde, Dr. B. Ghosh, Mr. A. Kumar of the BL-21 synchrotron facility, RRCAT. The author thankfully acknowledges Prof. P. Saha; Prof. D. Dasgupta; Dr A.K. Datta, IICB; Dr. S. C. Mande, NCCS; Prof S. Adak, IICB; Prof. H. Majumdar, IICB; Dr. K. M. Sinha, IMM; Prof. M. Rentala, J. N. U for their kind help.

The author is thankful to Mr. Sumit Basu (Computational Sciences Div.) for his kind help on issues ranging from installation of programs to troubleshoot hardwares/softwares and Mr. A. Bhattacharyya (C&MB Div.) for his kind help. She is grateful to the library staffs for procuring the required research articles within a short period of time besides providing books from other libraries.

Special thank goes to Dr. V. Venugopal for remaining ever approachable and for his numerous invaluable suggestions. She would also like to thank Dr. D. Bhandari, Dr. A. Maity and Ms. K. Sengupta for introducing her to the technical aspects of molecular biology.

She'd like to thank her parents for their never ending support and encouragement. Thank also goes to her sister for always being by her side, to her brother for his constant support and kind help at RRCAT and to her cousin for her encouragement. Finally, a special thank goes to her father and brother for their kind help with the transport of the delicate crystallization plates to the RRCAT synchrotron facility.

CONTENTS

SYNOPSIS

Crystal structure solution of phosphoglucomutase from *L. major* at 3.5 Å resolution and pharmacophore based data-mining to identify prospective anti-leishmanials.

Introduction

Parasitic protozoa *Leishmania spp.* are responsible for a broad spectrum of diseases collectively termed leishmaniasis, affecting humans as well as livestock. The emergence of strains resistant to the existing drugs (pentavalent antimonials) which were traditionally the first line of defense against the parasite and the high cost, toxic side effects of substitutes make it imperative to continue the search for low cost drugs with high potency against leishmaniasis.

Objective

The entire work consists of two parts (1) Data mining of drug databases by structure based pharmacophores to identify compounds which could be potent anti-leishmanials, and (2) Determination of the crystal structure of the enzyme phosphoglucomutase, from the pathogenic organism *Leishmania major* which provides structural insights into this potential drug target.

Both parts of the work have been discussed in greater detail below.

1. **Chapter 1:** In silico work flow for scaffold hopping in *leishmania*

With the emergence of strains resistant to antimonials (especially in Bihar state of India) [1] second line drugs such as amphotericin-B and miltefosine are being extensively used [2]. However, both these drugs are more expensive than antimonials, toxic with reportedly severe

side effects. Pentamidine and paromomycin are other drugs currently in use though their ready availability in endemic areas appears to be limited [3]. It is thus clear that there is an urgent need to search for and identify therapeutic alternatives to combat the disease.

In this work, the whole set of annotated coding sequences of human and the *L. major* have been compared to identify a set of proteins unique to the parasite. Crystal structures of these proteins or those of closely related homologues have been extracted from the Protein Data Bank (PDB) and the literature has been extensively surveyed to identify their specific high affinity inhibitors. Crystal structures of the protein (target) - inhibitor complexes have then been utilized to generate structure based pharmacophores. In case the crystal structures of the ligand bound targets were not available in the PDB, the inhibitors were computationally docked into the active sites of their receptors. Finally, the ZINC database and Drug Bank has been searched utilizing this set of pharmacophores to generate a set of compounds which could serve as a library in the search for prospective antileishmanial drugs.

Materials and methods

The annotated coding sequences (CDS) from the genomes of *L. major* (8316 CDS), *L. donovani* (8032 genes), *L. mexicana* (8249 genes), *L. braziliensis* (8056 genes), *L. infantum* (8227 genes), *T. Brucei* (9962 genes) and human (~41961 CDS) were downloaded from the NCBI genome database. To compare the annotated protein sequences the CDS's from *L. major* were first fed in as a query against the human CDS's database, followed by a second run where the human sequences were the 'query' and the *L. major* proteins the database. Proteins which simultaneously passed identical filters in both the runs of BLASTp were considered for the second step in the pipeline, with the sole exception of phosphofructokinase (PFK), being a

confirmed target [4,5,6]. All alignments with an E-value less than one were output from the program and default options were used for all other parameters. The software to analyze the BLAST outputs was developed locally in C/C++, Perl.

Template based homology modeling was performed using the MODELLER software both in the standalone mode and also implemented in Accelrys Discovery Studio 2.5. GOLD 5.2 was used to dock specific ligands onto the active site of their corresponding enzymes. LigandScout version 3.12 was used to generate the structure based pharmacophores from the crystallographic and modeled/docked complexes. Finally, the ZINC database was searched by all the structure based pharmacophores derived from the protein ligand complexes.

Results and discussions

The whole set of annotated coding sequences of *L. major* (8316 CDS) was compared against the human (41961 CDS) and those parasitic proteins (4991 among 8316 sequences) which could not align with any human gene, (**first filter**) with percentage sequence identity >35% and a simultaneous query coverage >50%, in two way reciprocal BLASTp runs, were retained. Hypothetical sequences (4407 CDS) were removed from this list (**second filter**). The Protein Databank (PDB) was searched for each of the remaining sequences and only those genes were selected (a total 90 sequences) which recorded hits in the PDB with >40 % sequence identity and a simultaneous query coverage >75% (**third filter**). Each gene from this set were then checked for homologues (pident >40% & query coverage >75%) in *L. donovani*, *L. mexicana*, *L. braziliensis*, *L. infantum* and also in the clusters of orthologous proteins in the Tritryp database (based on the OrthoMCL annotation [7]) only those proteins with homologues in all the five genomes were retained (**fourth filter**). The final set of genes consisted of a total of 86

polypeptide chains. Upon the application of more stringent criteria (pident > 25 and query coverage > 33) in the first filter, 47 out of the original 86 genes satisfied the new threshold values, of which 13 genes were retained from the first thirty proteins of the original set. The separated lists of hypothetical sequences (5135 CDS) were independently searched in the PDB and eighteen sequences satisfying the given criteria were listed.

Target Prioritization

The lists of parasitic proteins were sorted according to a "*weighted union*" scoring scheme based on the following factors: i) **essentiality** as determined from experimental studies, ii) **virulence factor** iii) **expression profile** and iv) whether the natural substrate to the protein is a ligand unique to *Leishmania spp* w.r.t human. A list of metabolites unique to *leishmania spp*. was identified by searching the leishmanial and human metabolome databases [8]. The full set of enzymes corresponding to these substrates was assembled into a blast database and run against the initial list of 86 proteins resulting in eight polypeptide chains associated with unique ligands.

The prominent amongst the list of the first 30 prioritized targets were three enzymes **trypanothione reductase**, putative **trypanothione synthetase** and **trypanothione - dependent glyoxalase I**. The trypanothione system in leishmania which replaces the ubiquitous glutathione system present in human enables the parasite to survive the high oxidative stress found in the host immune system and the presence of toxic heavy metals [9]. Both **trypanothione synthetase** and **trypanothione reductase** are attractive drug targets [10], as this system is the only pathway involved in the crucial regulation of oxidative stress in the parasites.

Three proteins, **UDP-galactopyranose mutase**, **putative UDP-Glc 4'-epimerase**, **UDP-sugar pyrophosphorylase** are either directly or indirectly responsible for the synthesis of β–

Gal*f*. β–Gal*f* is synthesized from its precursor UDP-galactose by the enzyme **UDP-galactopyranose mutase**, inhibition of which is known to regulate parasitic virulence and hence is an attractive target [11]. The other two enzymes **UDP - sugar pyrophosphorylase** and **UDP-Glc 4'-epimerase belonging** to the Leloir pathway and Isselbacher pathway respectively were found to be present on the prioritized list of enzymes. A set of five proteins implicated in purine/pyrimidine metabolism occupied fairly prominent positions: **adenylosuccinate lyase**, **putative deoxyuridine triphosphatase nucleotidohydrolase**, **nonspecific nucleoside hydrolase**, **adenine aminohydrolase** and putative **OMPDCase – OPRTase**. Unlike the host, leishmania lack the molecular machinery to synthesize purine nucleotides *de novo* and is thus dependent on a purine salvage pathway [12] rendering enzymes belonging to these pathway as important drug targets.

As **putative OMPDCase - OPRTase (OMPDC - OPRT)** is active in the final step of pyrimidine biosynthesis; its inhibition is expected to be lethal for the parasite. The three glycolytic enzymes **2,3 - bisphosphoglycerate – independent phosphoglycerate mutase**, **ATP – dependent phosphofructokinase** and **glycerol - 3 - phosphate dehydrogenase**, including two other enzymes either upstream or downstream of the glycolytic pathway, **putative NADP-dependent alcohol dehydrogenase** and **glucokinase** appeared in the list. Since glycolysis is the only known source for ATP in leishmania these enzymes offer attractive targets, especially **phosphoglycerate mutase (i-PGM)** which is distinct in terms of structure, catalytic mechanism and whose reduced expression was also found to be lethal for cultured *T. brucei*.

Pharmacophore Design and Screening of Zinc Database

A total of 34 structure based pharmacophores were derived from their corresponding ligand

bound three dimensional structures using LigandScout version 3.12. These pharmacophores were used to search the ZINC database using ZINCPharmer. A total number of 344 hits were recorded from the ZINC database which was then used to search the Drug Bank so as to identify similar molecules actually in use as pharmaceutical products. From the 344 compounds distributed over 23 pharmacophores, 9 exhibited similarities to drugs under investigation, 319 showed similarities to experimental drugs and **16** were similar to approved drugs (in at least one country).

The structure based pharmacophore derived from methotrexate bound to pteridine reductase returned 20 small molecule compounds from the ZINC database, with the pteridine ring being the principal pharmacophoric feature. Two of them were found to be similar to approved drugs pralatrexate and triamterene. Pharmacophores from trypanothione-inhibitor complexes found small molecules from the ZINC database out of which one was similar to the approved drug primaquine. Especially, fruitful were pharmacophores due to complexes with **nonspecific nucleoside hydrolase** as they yielded acarbose; mannitol, calcium gluceptate, nelarbine, didanosine, vidarbine; kanamycin, tobramycin, neomycin, framycetin, paromomycin, gentamicin, glucosamine, netilimicin, pitavastatin and diphylline.

Conclusions

The current work demonstrates the series of computational steps beginning with the comparison of the whole set of annotated sequences, prioritization of prospective drug targets, culling or assembly of inhibitor – target complexes through template based model building and docking, generation of pharmacophores and their subsequent use for searching small molecule databases to rationally assemble a set of lead compounds for experimentally testing as potential antileishmanials. The natural appearance of paromomycin a drug currently being employed

against visceral leishmaniais, in the list of lead compounds lends some confidence to the adoption of such scaffold – hopping techniques in order to generate a library of prospective antileishmanials.

References:

1. Rojas R, Valderrama L, Valderrama M, Varona MX, Ouellette M, Saravia NG. *J Infect Dis* 2006, **193**(10):1375-1383.
2. Ouellette M, Drummelsmith J, Papadopoulou B. *Drug Resist Updat* 2004, **7**(4-5):257-266.
3. Guerin PJ, Olliaro P, Sundar S, Boelaert M, Croft SL, Desjeux P, Wasunna MK, Bryceson AD. *Lancet Infect Dis* 2002, **2**(8):494-501.
4. Lopez C, Chevalier N, Hannaert V, Rigden DJ, Michels PAM, Ramirez JL**.** *Eur.J. Biochem.* 2002, **269** : 3978-3989.
5. Nowicki MW, Tullock LB, Woralll L, McNae IW, Hannaert V, Michels PAM, Fothergill-Gilmore LA, Walkinshaw MD, Turner NJ**.** *Bioorganic & Medicinal Chemistry.* 2008,**16** : 5050–5061.
6. Verlinde CLMJ, Hannaert V, Blonski C, Willson M, Perie JJ, Fothergill-Gilmore LA, Opperdoes Fr, Gelb MH, Hol WGJ, Michels PAM. 2001,**4**, 50–65.
7. Li L, Stoeckert,Jr. CJ, Roos, DS**.** *Genome Res. 2003,* **13**(9):2178 – 2189.
8. Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, Paley S, Rhee SY, Shearer AG, Tissier C. *Nucleic acids research* 2008, **36**(suppl 1):D623-D631.
9. Krauth-Siegel RL, Meiering SK, Schmidt H. *Biological chemistry* 2003, **384**(4):539-549.
10. Olin-Sandoval V, Moreno-Sanchez R, Saavedra E. *Current drug targets* 2010, **11**(12):1614-1630.

11. Oppenheimer M, Valenciano AL, Sobrado P. *Enzyme research* 2011, **2011**.

12. Majumder HK. Springer; 2008.

****

2. **Chapter 2:** The first structure of a kinetoplastid phosphoglucomutase at 3.5 Å resolution

The estimated annual occurrence of new cases of leishmaniasis is as high as 0.2 to 0.4 million for visceral leishmaniasis and 0.7 to 1.2 million for cutaneous leishmaniasis [1]. Many cases of cutaneous leishmaniasis, caused by the pathogen *Leishmania major*, goes unreported because the disease mostly occurs in remote areas plus the social stigma associated with the disfiguration and scars discourages patients to seek medical attention [2].

The emergence of strains resistant to the existing first line of defense provided by the pentavalent antimonial compounds and the toxic side effects thereof has appeared as a major health issue in the developing countries. The availability of the new formulations of the old drugs along with combinatorial drug therapy may delay the emergence of resistance, yet are not adequate due to toxicity or prohibitive prices. There is an urgent need for the development of alternative therapies that would provide adequate and affordable drugs with least toxic side effects.

The enzyme phosphoglucomutase (PGM) reversibly catalyses the conversion of glucose-1-phosphate (G1P) to glucose-6-phosphate (G6P) by the transfer of a phosphate between the C6 and C1 hydroxyl groups of glucose. A high degree of conservation of the PGM gene has been reported between the two sister species, *T. cruzi* and *L.major* [3]. Studies indicate that the outer surface of *T. cruzi and T. brucei is* covered with glycoconjugates whose glycans are mainly

made up of galactopyranose (Gal*p)* and galactofuranose(Gal*f*) residues [4]. Glycoconjugates such as mucin-type glycoproteins and glycoinositol phospholipids have been implicated to be virulence factors in *T. cruzi and T. brucei* and play a crucial role in evading hosts immune response [4, 5]. In *T. cruzi*, the biosynthesis of galactose takes place via the sugar nucleotide synthesis pathway involving the conversion of uridine-diphosphate-glucose (UDP-Glc) to UDP-Galactose (UDP-Gal) catalyzed by UDP-Gal-4'-epimerase. Since *T. cruzi* and *T. brucei* hexose transporter does not transport D-Gal, it relies solely on glucose (Glc) as the starting point of the generation of UDP-Gal. Besides being an important metabolite, UDP-Gal is essential for the growth of the parasite and the presumed donor for the synthesis of a rare deoxyneucleotide (β-D-glucosylhydroxymethyluracil) present in the DNA of trypanosomatids [5]. The essentiality of the glycoconjugates for the survival of the parasite inside the host renders enzymes of the sugar nucleotide synthesis pathway as potential therapeutic targets.

Results and discussion

The enzyme phosphoglucomutase was purified and crystals were obtained by hanging drop vapor diffusion method using PEG3350 as precipitant. The X-ray synchrotron data were collected at beamline 21, RRCAT, Indore. The crystal diffracted to 3.5 Å resolution and belonged to space group I121 with unit-cell parameters a =125.915, b = 114.971, c = 198.369; α = 90, β = 106.58, γ = 90. With four molecules in the asymmetric unit, the Mathews coefficient was evaluated to be 2.69 $Å^3$ Da^{-1} [6]. The final LmPGM model was refined to 3.5 Å resolution, with an R factor of 24.37% and R_{free} of 29.9%. For all the four molecules in the asymmetric unit, the initial 24-30 residues could not be modeled due to the absence of definite positive electron density and therefore, the residue contents were: molecule A (residues 25 to 589), molecule B

(residues 31 to 589), molecule C (residues 29 to 589) and molecule D (residues 31 to 589). The final model comprised 2244 amino-acid residues and 4 magnesium ions.

The model was validated using the program *Molprobity* [7]. 83.32% and 9.61% of the residues were within the Ramachandran favored and additionally allowed regions respectively while 7.07% were the outliers. The r.m.s deviation of bond lengths and bond angles from their ideal values were 0.0062 Å and 1.498° respectively.

Each molecule comprises four domains with an overall heart shape similar to other reported phosphoglucomutases. The approximate dimension of the molecule was 73 Å by 66 Å by 49 Å. Domain I, II and III share a common core made up of a mixed α/β structure, with four β strands (three stranded parallel sheet followed by an antiparallel strand) flanked by two α helices. Unlike the rest of the domains, domain IV is a member of the TATA-box-binding-protein like family comprising six stranded anti-parallel β sheet flanked by α-helices.

The active site and specificity pocket

The active site is situated inside a large cleft extending from the top to the centre of the molecule. The site is accessible from the top and through the groove that runs down one of the sides of the molecule extending from top to the bottom. The phosphoryl transfer site of the dephospho-enzyme comprises the catalytic serine (S118), located within loop 'P', playing the dual role of donor and acceptor of phosphate group, followed by H119 and N120. Together they form the catalytic triad as observed in other serine proteases [8]. In *P.aeruginosa*, the mutation of the residue N110 (N120 in LmPGM) to an alanine exhibited profound detrimental effect on the catalytic efficiency, indicating its crucial role towards completion of the catalytic process [9]. Residues D299, D301 and D303 situated on loop 'M' and close to the

phosphoryl transfer site, interact with the divalent cation Mg^{2+}, the physiological activator of the enzyme [10]. The catalytic serine S118 was found to be directed towards the metal binding loop and the S118- O^{γ} and to interact with the divalent cation.

Catalytic properties of LmPGM

The purified enzyme was catalytically active in the absence of the activator glucose-1,6-bisphosphate. The kinetic constants K_m, V_{max} and K_{cat}/K_m were calculated to be 21.35±3.1 µM, 0.55±0.02 $\mu Ms^{-1}\mu M^{-1}$of LmPGM and 2.58X10^4 $M^{-1}s^{-1}$ respectively. Addition of 0.1 µM glucose-1,6-bisphosphate resulted in almost twofold increase in V_{max} (1.02±0.06 μMs^{-1} μM^{-1} of LmPGM) and catalytic efficiency K_{cat}/K_m (4.52 X10^4 $M^{-1}s^{-1}$).

Conformational variability of domain IV

The conformational flexibility of the fourth domain was closely modeled by normal mode analysis. Compared to other domains, domain IV exhibited maximum conformational flexibility via the hinge joint. The least affected were the domain II and III. The domain motion analysis by the program *DynDom* [11] revealed that the switching between the "closed" and "open" enzyme involved ~10° rotation of the fourth domain. The motion of the fourth domain causes the tip of the loop **L** to move towards the central cleft by ~ 10Å and together with flap **F** actuates the closure of the lid over the active site.

Conclusion

The structural analysis lead to the identification of a pocket situated in the close proximity of the hinge region and bounded by residues from the two extended loops containing regions of

insertion. The site features included a negatively charged region at the top and a hydrophobic region at the bottom in addition to binding features such as concavity and surface roughness. The site is unique to the pathogen and provides an opportunity for the strategic design of inhibitors. It may also facilitate the design of bifunctional drugs where two partners bind to two different site and are tethered together by a linker molecule resulting in an increase of total binding affinity. The presented report, thus, provides an insight into this potential drug target.

References:

1. Alvar, J., Vélez, I. D., Bern, C., Herrero, M., Desjeux, P., Cano, J., Jannin, J., Boer, M., D. (2012). *PLoS One*. 7(5): e35671 (doi:10.1371/journal.pone.0035671).
2. Reithinger, R., Mohsen, M., Aadil, K., Sidiqi, M., Erasmus, P., Coleman, P. G. (2003). *Emerging Infectious Diseases*, 9, 727–729.
3. Penha, L. L., Sant'Anna, C. B., Mendonça-Previato, L., Cunha-e-Silva, N. L., Previato, J. O., Lima, A. P.(2009). *Glycobiology*. Vol. 19, no. 12, 1462–1472.
4. Penha, L. L., Mendonça-Previato, L., Previato, J. O., Scharfstein, J., Heise, N., C. de A. Lima, A. P. (2005). *Glycobiology*. Vol. 15, no. 12, 1359–1367.
5. Bandini, G., Mariño, K., Güther, M.L.S., Wernimont, A. K., Kuettel, S., Qiu, W., Afzal, S., Kelner, A., Hui, R., Ferguson, M. A. J., (2012). *Molecular Microbiology* Volume 85, Issue 3, pages 513–534.
6. Matthews, B. W. (1968). *J Mol Biol.* 33(2), 491-497.
7. Chen VB, Arendall III WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. (2010). *Acta Cryst.* D, **66**: 12-21.

8. Müller, S., Diederichs, K., Breed, J., Kissmehl, R., Hauser, K., Plattner, H., Welte, W. (2002). *J Mol Biol.* Jan 11, 315(2), 141-153.

9. Schramm, A. M., Beamer, L. J. (2006). *J Biol Chem.* 281(22), 15564-71.

10. Lin, Z., Konno, M., Abad-Zapatero, C., Wierenga, R., Murthy, M. R., Ray, W. J. Jr., Rossmann, M. G. (1986). *J Biol Chem*. 261(1), 264-74.

11. Hayward, S., Berendsen, H. J. C. (1998). *Proteins: Structure, Function and Genetics*. 30, 144-154.

****

LIST OF FIGURES

Chapter I

Chapter II

Chapter III

Appendix

LIST OF TABLES

Chapter I

Chapter II

Chapter III

Appendix

List of Supplementary materials (Provided separately with the CD-ROM)

CHAPTER II

Supplementary file 1.xlsx: The full list of sorted 86 prioritized proteins.

Supplementary file 2.xlsx: List of hypothetical proteins which scored hits in the pdb.

Supplementary file 3.xlsx: Target proteins and their inhibitors identified from the literature.

Supplementary file 4.xlsx: List of atomic contacts between inhibitors and the target enzymes.

Supplementary file 5.xlsx: Pharmacophore designed from inhibitors.

CHAPTER III

rcsb086012.pdb: Coordinate file of the solved structure

rcsb086012-sf.cif: Structure factor file of the solved structure

Domain motion.avi: Supplementary movie revealing the conformational flexibility of the molecule

CHAPTER I

1. Introduction

The parasites belonging to the order trypanosomatida are responsible for a broad spectrum of diseases collectively referred to as leishmaniasis. The disease generally occurs in three forms: (i) visceral (VL) (ii) muco-cutaneous (MCL) and (iii) cutaneous (CL). Visceral leishmaniasis is caused by *Leishmania donovani* and *Leishmania infantum* and can be fatal if left untreated. Cutaneous leishmaniasis usually appears as self healing skin lesions generally caused by *L. major* which could lead to permanent scars in patients and may occasionally require prolonged treatment for a complete cure. In the muco-cutaneous form of the disease, the lesion spreads to the mucosal tissues leading to the disfiguration of the nasal septum, lips and palate. Post kalaazar dermal leishmaniasis (PKDL) generally occurs subsequent to VL with the appearance of self healing skin rashes. Immuno-compromised patients (e.g. HIV positive) residing at endemic regions are most vulnerable to leishmaniasis with high chances of relapse and treatment failure [1].

The disease is endemic in 98 countries including 72 developing and 13 under developed nations. The annual occurrence of new cases are estimated to be 0.2 to 0.4 million, 0.7 to 1.2 million for VL and CL respectively [2]. Almost 90% of VL patients are found in India (primarily Bihar state), Bangladesh, Sudan, Ethiopia and Brazil. More than 350 million people living in the endemic areas are at risk of contracting the disease [3]. However, many cases go unreported due to its prevalence in remote areas. Apart from that, the social stigma associated with disfiguration by scars discourages patients to seek medical attention [4]. The infected female sandflies [*Phlebotomus spp.* (**Fig. 1*a***) and *Lutzomyia spp*.] are known to transmit more than 20 species of the parasite to human [1]. The progressive clinical manifestation of the

disease, starting from a papule and its subsequent development to an ulcer is shown in **Fig. 1*b-d***. Depending on the sites of parasitic development within the abdomen of the sand fly vector, leishmania can be divided into two subgenera: (i) *Leishmania* (prevalent in both the New and the Old World) and (ii) *Viannia*, which is prevalent primarily in the Old World [5]**.** The lifecycle of the parasite is primarily divided two stages distinguished by their characteristic morphological forms: (i) the vector stage consisting of 15-20 µm flagellated promastigotes inside the mid-gut of sand fly and (ii) the host stage comprising 3-5 µm long non-flagellated amastigotes within cells of monocyte-macrophage origin [5].

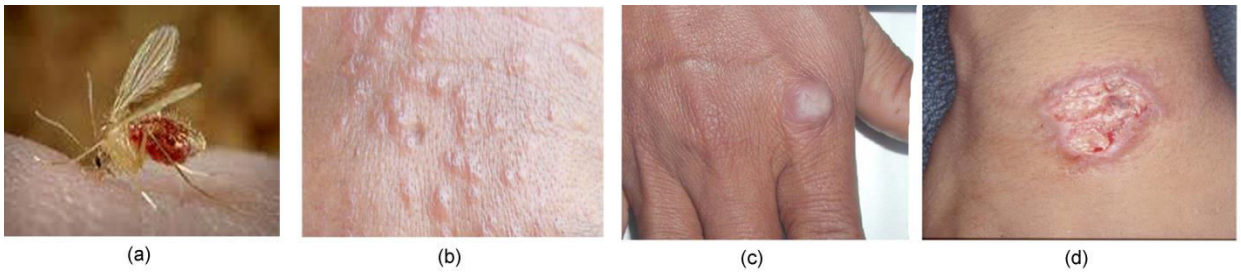


Figure 1: The vector and the clinical manifestation of the disease. (a) Phlebotominae sandfly which is a carrier for the disease (http://en.wikipedia.org/wiki/Phlebotominae) (b) Progressive manifestation of CL from a papule (c) to a nodule (d) finally leading to an ulcer. Figures adapted from [6].

The infected female sand fly injects promastigotes into the skin of the host during the course of a bloodmeal which are readily engulfed by the host macrophages. Once inside the macrophage, the promasigotes transform into amastigotes and reproduce finally leading to the lysis of the macrophage, releasing amastigotes to infect other macrophages.

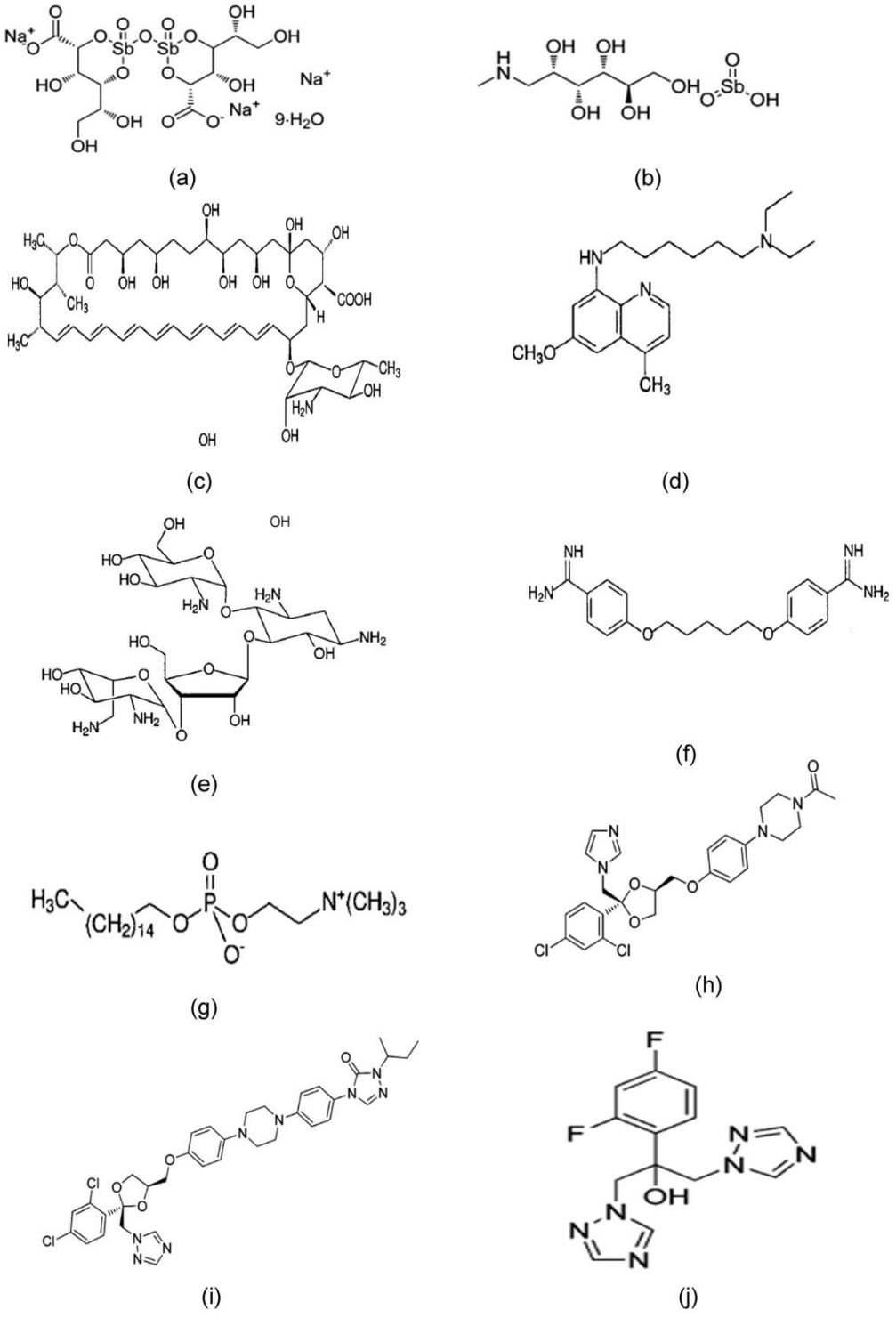


Figure 2: Available antileishmanials. (*a*) Sodium stibogluconate (Pentostam), (*b*) Melgumine antimoniate, (*c*) Amphotericin B, (*d*) Sitamaquine, (*e*) Paromomycin sulphate (Aminosidine), (*f*) Pentamidine, (*g*) Miltefosine, (*h*) Ketoconazole, (*i*) Itraconazole, (*j*) Fluconazole. Figures obtained from [7].

Drugs that are already in use for the treatment of leishmaniasis include pentavalent antimonials, pentamidine, several formulations of amphotericin B, paromomycin and miltefosine (**Table 1, Fig. 2**). For the past 50 years, pentavalent antimonials have been the first line of defense against VL, CL and MCL (**Fig. 2*a*, 2*b***), and still provides the cheapest formulation available for the treatment of leishmaniasis. Antimonials inhibit energy metabolism and fatty acid oxidation [7] in the parasites and in-vitro studies indicate that treatment with Sb(III) leads to DNA fragmentation, apoptosis in promastigotes. Sb(III) also inhibits the reduction of thiols (trypanothione, glutathione) following their oxidation by diamides, thereby, causing intracellular thiol levels to drop significantly in the parasite, leading to a decrease in cellular redox potential .

Lately, increased parasitic resistance to antimonials have been detected in India especially Bihar state, in which case second line drugs include amphotericin B (**Fig. 2*c***), pentamidine (**Fig. 2*f***) and miltefosine (**Fig. 2*g***). However, most of these drugs are expensive, difficult to adminster and induce severe side effects. For example, patients treated with pentamidine have reported insulin dependent diabetes mellitus, hypoglycemia, hypotension, fever, myocarditis and renal toxicity [5].

In contrast, miltefosine exhibits relatively less severe side effects compared to other antileishmanials and can be administered orally. A relatively new and promising antileishmanial is paromomycin which is being extensively used against both VL and CL. It was found to be more effective against *L. major* compared to *L. amazonensis* or *L. panamensis* [7] during experimental and clinical trials. Recent studies showed that topical application of paromomycin (15% paromomycin and 0.5% gentamycin) was effective against CL [8]. Other prospective antileishmanial candidates include the quinolin derivative sitamaquine (**Fig. 2*d***) and antifungal azoles (**Fig. 2*h-j***) against CL, which inhibit the ergosterol biosynthesis pathway (Lanosterol 14

Table 1: List of available drugs for leishmaniasis. Abbreviation used: VL- visceral leishmaniasis, CL – cutaneous leishmaniasis, MCL – muco-cutaneous leishmaniasis, PKDL – post kalaazar dermal leishmaniasis. Table adapted from [5].

Drug	Target Pathway and Mechanism of inhibition	Clinical manifestation and route of drug administration	Side effects	Advantage/ Disadvantage
Pentavalent antimonial compounds: 1. sodium stibogluconate 2.meglumine antimoniate	Targets glycolytic pathway and fatty acid oxidation. Inhibition of ATP and GTP formation depending upon the dose administered.	VL, CL, MCL, PKDL Intralesional: CL Intra-muscular/intra-venous.	Systemic: Pancreatitis, thrombocytopenia, leucopenia, cardiac arrhythmia, deranged liver enzymes	Cheapest among the available drugs. Some parasite strains are resistant to the formulation.
Pentamidine isethionate	Inhibition of polyamine biosynthesis and disruption of mitochondrial membrane potential	CL, MCL Intralesional: CL Intra-muscular	Pain at injection site Hypoglycaemia Hypotension Diabetes Renal dysfunction	Development of resistance is Problematic. Adverse effects limit its use
Amphotericin B and lipid formulations	Inhibition of cell membrane synthesis by binding to ergosterol Pore formation in cell membrane	VL, CL, MCL, PKDL Intra-venous.	Fever, chills, bone pain, Hypokalaemia, nephrotoxicity	Lipid formulations less toxic than amphotericin B. Effective total dose varies with geographical region. More expensive than antimonials
Paromomycin	Possible interference with RNA synthesis and membrane permeability	VL Topical: CL Intra-muscular	Topical: Pain, erythema, blistering Systemic: hepatotoxicity,	Combination with antimonials results in higher cure rates of VL in India, but not in Africa

			reversible VIII nerve damage	
Allopurinol	Interference with protein synthesis (purine salvage cycle)	VL, CL Oral	Rash	Ineffective as monotherapy: used in combination with sodium stibogluconate for VL
Azole derivatives: -fluconazole, ketoconazole, Itraconazole	Inhibition of 14a-lanosterol demethylase required for ergosterol biosynthesis	CL Oral	Hepatotoxicity	Inconsistent success between species
Alkylphosphocoline analogues: Miltefosine (hexadecyl-phosphocholine)	Alteration of glycosylphosphatidylinositol anchor synthesis, ether lipid metabolism, signal transduction and alkyl-specific acyl-coenzyme A acyl-transferase	VL, CL, PKDL Oral	Gastrointestinal disturbances Hepatorenal toxicity Teratogenic: contra-indicated in pregnancy	Lack of compliance results in emergence of resistance especially in anthroponotic transmission. Higher cure rates when used concurrently with paromomycin or sequentially after liposomal amphotericin B for VL in India.

α-demethylase) leading to the accumulation of 14 α-methyl sterols. Thus, currently there is an urgent need to identify antileishmanials which are economically viable, effective with minimal side effects and easily deliverable to relatively remote regions of developing countries wherein the disease is endemic.

Recent years have witnessed an extensive application of approaches such as virtual screening using pharmacophores in drug discovery.

2. Pharmacophore Generation

With the rapid advancements in computational methods and protein crystallography, in silico approaches have become an integral part of drug design and discovery. A proper understanding of the metabolic pathways, selection of potential biological targets and identification of target specific ligands is a pre-requisite for the design of small molecule drug candidates. Screening thousands of ligands, based upon their interaction profiles with the biological target for a particular disease, is possible using both automated experimental screening schemes and computational approaches. Over the year, computational approaches have emerged as an attractive alternative against the expensive experimental effort involved in screening large number of compounds [9]. Virtual screening (VS) techniques may be thought of as a hierarchical filtering method where libraries of chemical compounds are searched to identify molecules exhibiting particular required activities. VS find its application in identifying novel lead structures from commercially available databases of small molecules.

Once the bioactive conformation of the ligand becomes available, either by X-ray, NMR or molecular modeling, novel ligands could be searched and ranked by virtual screening techniques using 3D similarity search or by pharmacophore pattern matching. The

pharmacophore hypothesis is used as a template in pharmacophore based VS approaches to identify and rank molecules (hits) that exhibit chemical features similar to those of the template.

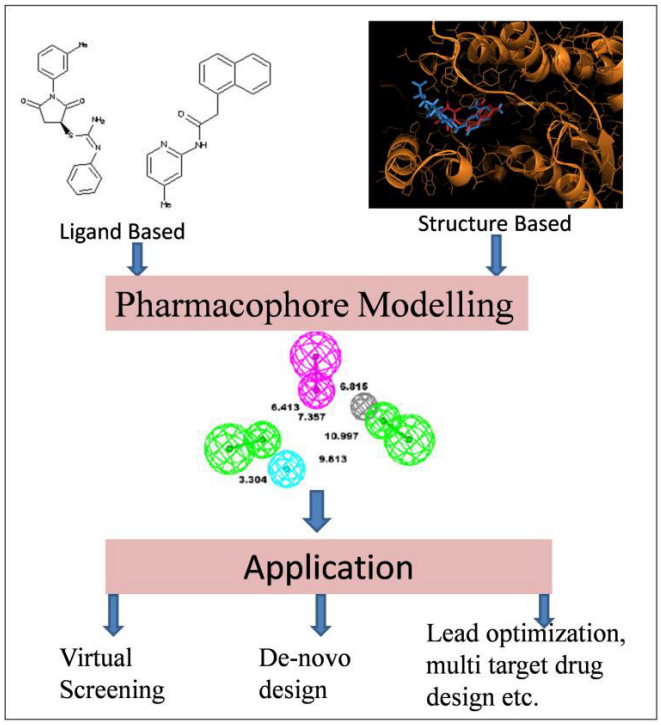


Figure 3: Pharmacophore designing and its application. The pharmacophore once generated, can then be used for virtual screeing, de-novo designing of drugs or other application such as lead optimization and designing drugs for targeting multiple loci.

According to the IUPAC definition 'A pharmacophore is the ensemble of steric and electronic features that is necessary to ensure the optimal supramolecular interactions with a specific biological target structure and to trigger (or to block) its biological response'. A pharmacophore does not represent a real molecule rather it is an abstract concept. A typical pharmacophoric feature indicates regions in a molecule that are hydrophobic, aromatic, electron bond acceptor or donor and the spatial relationship between them. Apart from VS, pharmacophore approaches find application in de novo design, lead optimization and multitarget drug design (**Fig. 3**) [10]. In de-

novo drug design, a set of disconnected pharmacophoric fragments are finally connected by linkers such as atoms, chains and ring moieties.

Cavalli *et al*. had designed pharmacophore from the available thirty one K^+ channel blockers (K^+ Channels encoded by the human ether-a`-go-go related gene, HERG), which were found to exhibit some common features irrespective of their structural differences (**Fig. 4**), e.g. the presence of the basic nitrogen of the piperidine cycle and aromatic rings [11].

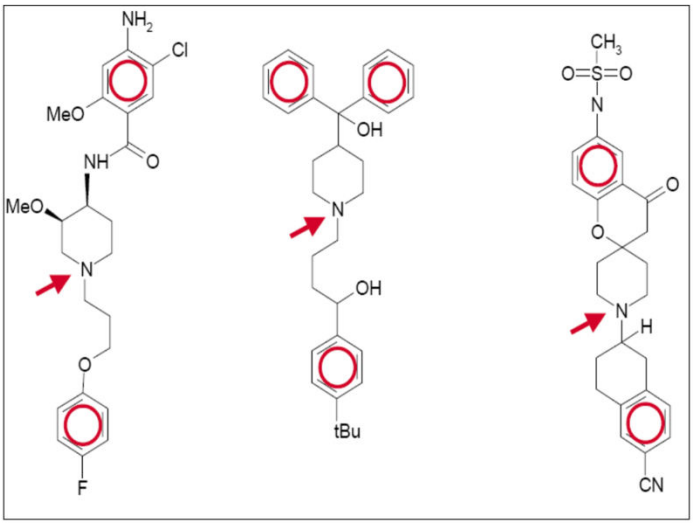


Figure 4: Three of the 31 K^+ channel blockers exhibiting common features (basic nitrogen and aromatic rings are highlighted) which were used to generate the pharmacophore. Figure adapted from [11].

The crystal structure of the most potent K^+ channel blocker was used as the initial reference structure and in the first step, molecules with similar spatial and geometrical characteristics were superposed onto it (**Fig. 5*a***). These superimpositions led to the individuation of other pharmacophoric features, which were then used to superimpose structurally diverse inhibitor molecules and to refine the initial alignment.

The initial reference molecule was assigned three pharmacophoric points, the basic nitrogen of the piperidine cycle (N) and the centre of mass of the two close aromatic moieties (centroids C_0 and C_1). Since a number of molecules from the training set were found to contain an aromatic group connected to the basic nitrogen, a fourth pharmacophoric point was defined as the centroid of the phenyl ring (C_2). Molecules that did not contain all four pharmacophoric points were superimposed either on the basis of available points or on the basis of additional features present on the molecule, e. g. the presence of halogen atom at the para-position on one phenyl ring (C_0) was used to reinforce the fit.

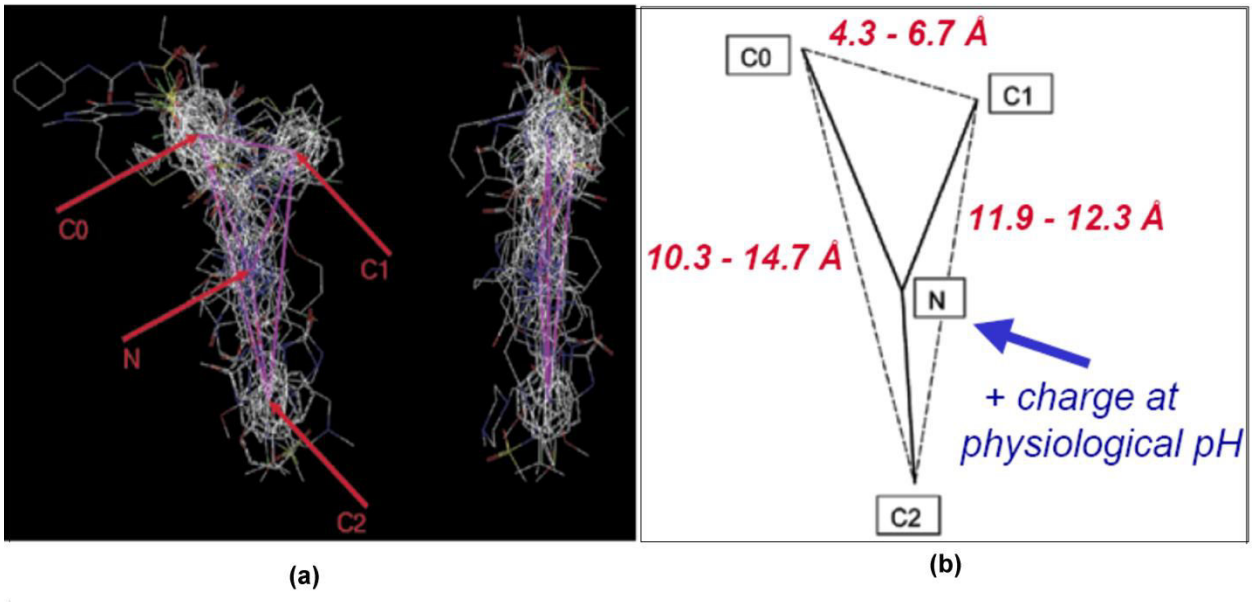


Figure 5: Generation of pharmacophore. (a) The orthogonal views of the superimposition of 31 K^+ channel blockers. The pharmacophoric frame is shown in magenta and the features (basic nitrogen N and three centroids: C_0, C_1 and C_2) are indicated by arrows. (b) The simplified pharmacophoric hypothesis. Figures adapted from [11].

Distances and the angles between the pharmacophoric points were noted carefully and the simplified pharmacophore hypothesis was obtained by joining the average spatial position of all the molecules (**Fig. 5*b***).

Two approaches are widely used for the design of pharmacophores: (i) Ligand based approach and (ii) Structure based approach.

2.1 Ligand based approach

In the absence of a macromolecular target structure, the common chemical features are extracted from a set of known ligands representing the essential interaction features between a ligand and its specific target. The two key steps of the pharmacophore generation are (i) model the conformational flexibility of each ligand of the training set by creating conformational spaces (ii) align large number of ligands to extract the common features and construct the pharmacophore models. To model the conformational flexibility, multiple conformations are either initially generated and saved as a database (pre-enumerating method) or constructed during the pharmacophore modeling process (on-the-fly method).

2.2 Structure based approach

In structure based methods, the complementary chemical features between the receptor active site and its respective ligand(s) along with their mutual spatial relationships are analysed to generate the pharmacophore models. Structure based methods can farther be subdivided into (i) macromolecule-ligand complex based (ii) macromolecule based. The first method requires the availability of the 3D structure of the macromolecule-ligand complex, from which the key

interaction features between the ligand and the macromolecule are utilized for the generation of the pharmacophore models.

When the structure of the protein bound ligand is unknown, a de-novo design approach i.e. LUDI can be used to design potential ligands [12]. LUDI locates potential binding interactions basing on the ability of the ligand and the receptor to participate in hydrogen bonding (donor/acceptor) and hydrophobic interactions. The complements of the functionality at the binding site are represented by the shape and feature types in the interaction map (hydrogen bond donor/acceptor, hydrophobes) (**Fig. 6**). The LUDI interaction map thus generated are then converted to pharmacophoric features.

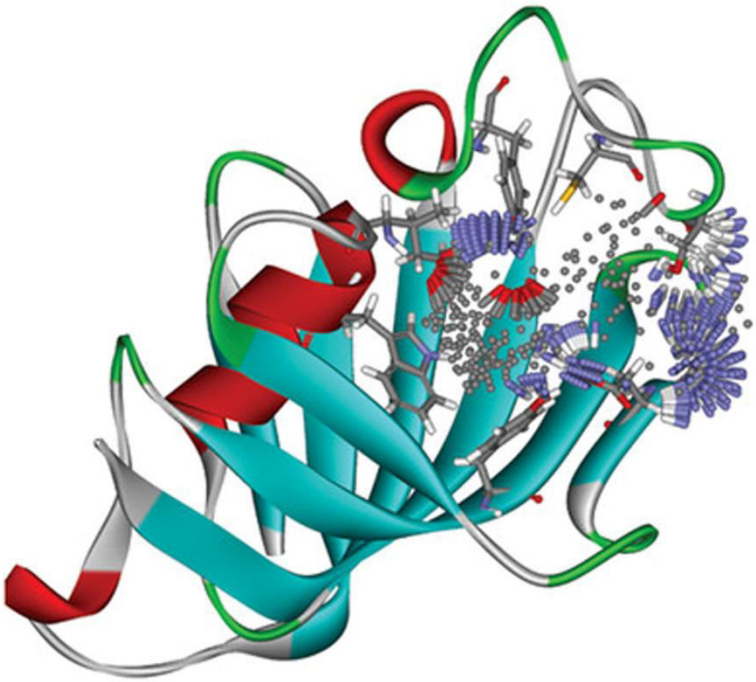


Figure 6: Ludi interaction map. Acceptor: red rods (OH), donor: purple rods (NH), hydrophobes: black spheres. Figure adapted from [13].

The pharmacophores, thus designed may be used as queries to search the 3D-chemical databases. The two major challenges of pharmacophore based VS are: (i) to account for the conformational flexibility of small molecules either by pre-enumerating multiple conformations or comformational sampling on each run and (ii) identification of the pharmacophoric patterns (substructure searching) to check whether a query pharmacophoric feature is present in a particular conformation of a small molecule.

The main aim of drug discovery approaches is to find novel active compounds against preselected biological target. Scaffold hopping or the lead hopping technique provides a measure to discover active compounds with novel backbones with improved drug like properties.

3. Scaffold Hopping

A 3D pharmacophore derived from a series of structurally diverse ligands (sharing a common feature and capable of adopting similar shapes) may contain minimal set of spatially oriented features which are the pre-requisite of an active compound. Those pharmacophores could be used for searching novel scaffolds (scaffold hopping) while maintaining the essential reaction features. In the scaffold hopping technique, the central core structure of a known active compound is modified while retaining its biological activity such that it binds to the same receptor irrespective of the modification [14]. The new compounds generated by scaffold hopping techniques could result in improved binding due to the substitution of a structurally labile by a rigid scaffold; increased solubility by substituting a more polar scaffold in place of a lipophilic one or enhance both the stability and specificity of the ligand to its target thereby reducing its toxic side effects. Discovery of entirely novel structure could also lead to new patents. The four major categories of scaffold hopping techniques are: (i) Heterocycle

replacement (ii) Ring opening or closure (iii) Pseudopeptides or peptidomimetics (iv) Topology/shape based hopping.

3.1 Heterocycle replacement

While designing pharmacophoric features, the core structure of drug molecules generally provide multiple vectors projecting in different directions. In the heterocycle replacement method, the C, N, O, and S atoms in a heterocycle are replaced while maintaining the direction of the vectors. If the heterocycle interacts directly with the target macromolecule, then the modification may result in improved binding affinity. One example of this procedure is the design of the derivatives of the drug rimonabant which exerts inverse agonist effect on the cannabinoid 1 receptor (CB1) thereby causing a reduction in appetite. The methylpyrazole core in rimonabant was replaced by thiazoles, pyrroles, and pyrazines and the newly designed scaffolds resulted in novel classes of CB1 receptor antagonists [15].

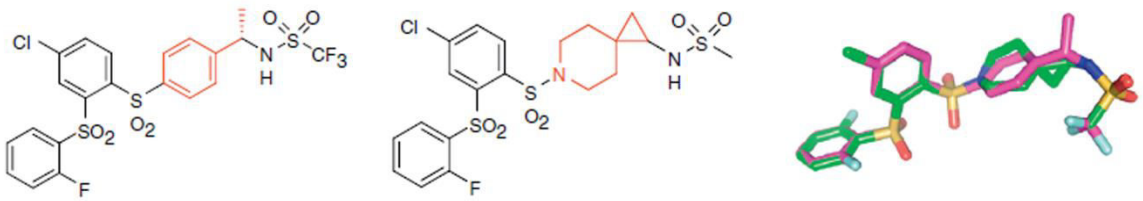


Figure 7: Heterocycle replacement method used in scaffold hopping. (a) Structure of a triaryl bis-sulfone CB2 receptor inhibitor. (b) Biaryl analog of the inhibitor. (c) The superposition of both structures. Figures adapted from [15].

In another similar situation, the central phenyl ring of the triaryl bis-sulfone CB2 receptor inhibitor was replaced with spirocyclopropyl piperidine (**Fig. 7**). Besides being more drug like, the resulting derivative demonstrated same potency against CB2 and selectivity against CB1.

3.2 Ring opening or closure

Generally drug like molecules contain at least one ring system any changes to which is accompanied by changes in the entropy related free energy of binding, membrane penetration and absorption. Therefore, opening or closing of the ring system provides an opportunity to manipulate the flexibility of the molecule by changing the number of free rotatable bonds and thereby improving the drug like properties of a molecule. Conversion of an alkyl chain to cyclohexane, piperizine or piperidine, *o*-hydroxylbenzoyl group to quinazoline and arylamine or arylamide to a fused ring system have been used successfully for the design of a novel scaffold [15].

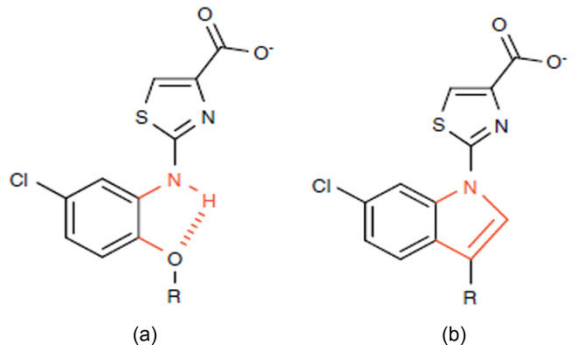


Figure 8: Ring opening or closure method used in scaffold hopping. Prostaglandin EP1 receptor anatagonists (a) biaryl amine series (b) indole series. Figures adapted from [15].

The ring closure design was succefully applied to a series of indole compounds which were known to be Prostaglandin EP1 receptor anatagonists (**Fig. 8**). The intra-molecular hydrogen

bonding between the o-alkoxy group and biaryl NH offered direct hints for the ring closure. One of the resulting indole compounds exhibited higher activities in binding and functional assays.

3.3 Pseudopeptides or peptidomimetics

There are several peptides present in living organisms e.g. peptide hormones, growth factors and neuropeptides. In human, an imbalance of these peptides may lead to diseases such as diabetes,

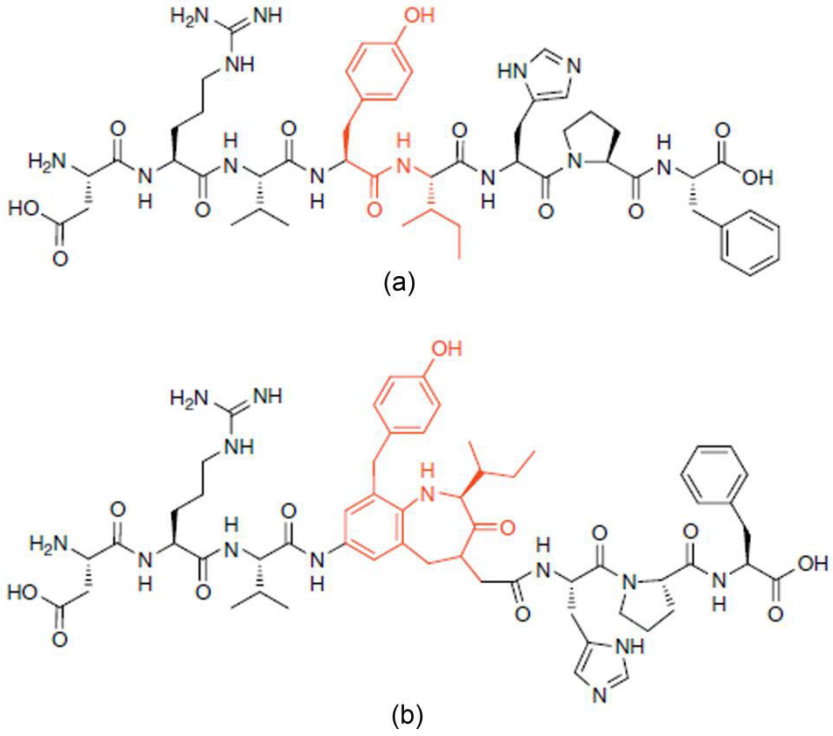


Figure 9: Pseudopeptides or peptidomimetics method used in scaffold hopping. (a) The structures of octapeptide hormone angiotensin II (DRVYIHPF) (b) benzodiazepine-based β-turn mimetic of Ang II. Figures adapted from [15].

osteoporosis, cancer and endometriosis. In the peptide based drug discovery approach, the active peptide conformation is used as templates to design small molecules in order to reduce the peptide character so as to avoid proteolysis while retaining the key interaction features essential

for molecular recognition [15]. For example, while targeting the protein-protein interaction site, the small molecules are designed to mimic the interacting moieties of the protein but are metabolically stable and exhibit increased bioavailability. Generally protein-protein or peptide-protein interfaces involve secondary structural elements such as α-helices, β-sheets or turns. Small molecules have been designed to mimic these secondary structures such that the side chain orientation of the synthetic molecule are consistent with that of the parent structure.

The octapeptide hormone angiotensin II (Ang II) adopts a turn structure centered at Y(4) while activating the angiotensin 1(AT1) receptor. The Y(4) and I(5) were replaced with a benzodiazepine (**Fig. 9**), a well-known β-turn mimetic scaffold, resulting in an increase in its affinity towards its receptor.

3.4 Topology/shape based hopping

Topological pharmacophore models derived from reported inhibitors have been used successfully to generate hits that were found to be significantly different from query compounds. The main advantage of pharmacophore modeling is that it simplifies the complex nature of non covalent target-ligand interactions and thereby renders comparisons of pharmacophoric features by the computer feasible. At the same time, the simplifications are accompanied by a decrease in its efficiency to provide perfect prediction of ligand binding propensities. Yet, the method is practically useful and has been employed successfully in the discovery of new chemical classes exhibiting desired biological activity.

In an attempt to interrupt the bacterial cell wall biosynthesis, the ZipA-FtsZ protein–protein interactions was targeted by choosing a weak high-throughput screening (HTS) hit as the template (**Fig. 10*a***).

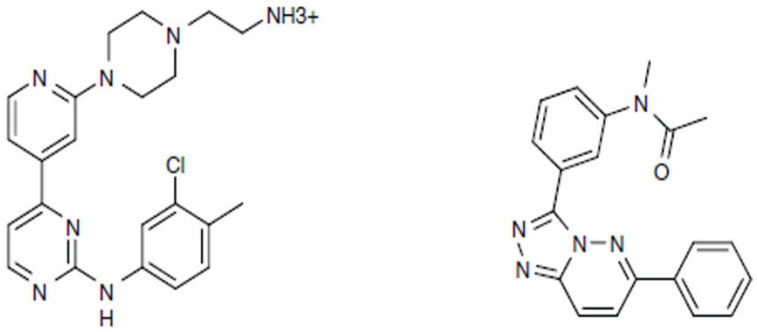


Figure 10: Topology/shape based hopping. Structures of ZipA-FtsZ inhibitors. (a) pyridylpyrimidine template from high-throughput screening (b) ROCS (Rapid Overlay of Chemical Structures)-identified hit. Figures adapted from [15].

The template was found to possess toxicity and intellectual property issues. The shape based ROCS (Rapid Overlay of Chemical Structures) search recorded hits which were less potent but did not show toxicity and comprised novel scaffolds (**Fig. 10*b***).

While using virtual screening algorithms, benchmarking sets are required for the quantitative assessment of the correlation between docking scores and ligand binding affinity. Different tests use separate benchmarking sets for this purpose. Pose prediction test sets are obtained from the crystal structures available in PDB and the docking algorithm is considered reliable if it successfully reproduces the ligand-bound structure. Similarly, enriched test sets are used to assess the ability of the program to enrich (assign better score to a known compound compared to a decoy molecule) active compounds rather than the assumed or known decoys.

The database DUD-E (Directory of Useful Decoys, Enhanced) provides sets of decoys in order to help benchmark molecular docking programs [16]. It contains 22,886 active compounds and their affinities against 102 targets and 50 decoys per each active compound with similar physico-chemical properties but different 2-D topology in mol2 or SDF formats. Besides, the

enhanced decoy generation procedure is fully automated and can be used to generate DUD-E decoys for any input list of ligands.

In order to validate the pharmacophores in retrospect, they are used as query against a database comprising the ligands which were used to generate the pharmacophores, other known active compounds and the decoys. Those ligands which were used to generate the pharmacophores should obtain the highest score. The other validation method employed to test computationally generated pharmacophores is the ROC curve.

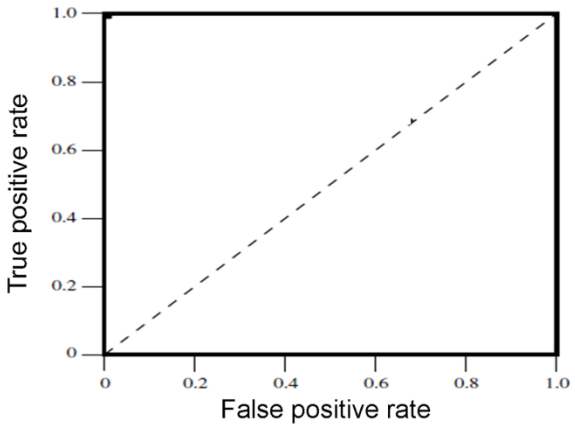


Figure 11: The ROC graph. One point in ROC space is considered better than the other if it is located more to the upper left (True positive rate is higher and false positive rate is lower).

ROC (Relative operating characteristic curve) curves have widely been used to differentiate hits from false alarm or to select the possible optimal models while discarding the suboptimal ones for different threshold values [17]. For a two class prediction problem, where the instances or the outcomes of the two classes are either labeled 'positive' or 'negative' for different threshold values, four outcomes are possible 1. true positive, 2. false negative, 3. true negative and 4. false positive. For a given set of classifier and a test set, a 2×2 matrix can be

constructed to display the possible outcomes. The true positive rate is defined as:

True positive rate = (Positives correctly classified)/ (Total positives)

And the false positive rate is defined as:

False positive rate = (Negatives incorrectly classified)/ (Total negatives)

On a ROC graph, true positive rate is plotted along the Y-axis and the false positive rate along the X-axis (**Fig. 11**). One point in ROC space is considered better than the other if it is located more to the upper left so that the true positive rate is higher and false positive rate is lower. Only those pharmacophores whose area under the ROC curve lies above a certain threshold are generally selected for further analysis, e.g. screening the small molecule databases.

4. Small molecule databases

Modern drug discovery approaches relies heavily on iterative screening of small molecules stored in publicly accessible databases. Besides, the bioassay or chemical reaction data of each entry are regularly used while selecting a list of small molecules against a specific target.

4.1 ZINC database (http://zinc.docking.org/)

The database houses more than 35 million purchasable compounds for virtual screening. Data can be downloaded free of cost in several file formats like SMILES, mol2, 3D SDF, and DOCK flexibase format. The availability of web interface tools enables the database to be searched and browsed and subsets to be created. Several filters can be applied while searching the database,

e.g. net charge, molecular weight, unique serial number assigned to each compound or by uploading a list of SMILES32 strings in a text file [18].

Simplified Molecular Input Line Entry System (SMILES) is a chemical notation language that can be easily processed by computers. Using this line notation method, chemical structure can unambiguously be represented by a linear string of symbols. SMILES denote the two dimensional valence oriented structure of a molecule [19]. The single, double, triple, and aromatic bonds are represented by the symbols -, =, #, and :, respectively although single and aromatic bonds are usually omitted (**Fig. 12**).

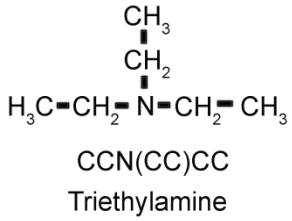


Figure 12: The SMILES notation for triethylamine.

The ZINC server allows users to upload their own compounds and process them for example as positive or negative controls which are not a part of the database. The database is useful not only for VS purposes but also for computational ligand discovery.

4.2 DrugBank (http://www.drugbank.ca/)

DrugBank is a database housing nearly 8000 drug entries including FDA approved small molecule drugs, bio-tech drugs, neutraceuticals (nutritional supplements) and experimental drugs. In addition, more than 40,000 non-redundant protein sequences are directly linked to the drug entries. The web-interface of DrugBank provides an opportunity for searching, viewing and sorting the query results as well as for downloading small molecular structure data, target

sequence and extracting text or images. DrugBank has greatly facilitated *in silico* drug target discovery, drug design, drug docking or screening, drug metabolism prediction, drug interaction prediction and general pharmaceutical education [20].

5. The kinetoplastid genome

The two *Leishmania* species: *L.donovani* and *L.major* are prevalent in the old world and contain 36 chromosomes pairs (0.28 to 2.8 Mb) [21]. *L. Mexicana* group and *L. brazilliensis* group are prevalent in the new world and contain 34, 35 chromosomes pairs respectively. The reduction in chromosome number is due to the fusion of chromosomes 8+29 and 20+36 for the L. *Mexicana* group and 20+34 in the L. *brazilliensis* group (**Table 2**).

Table 2: The chromosome organization of leishmanial and trypanosome species and number of annotated coding sequences

Leishmania spp.	Number of chromosomes	Annotated coding sequences
L.major	36	8316
L. donovani	29[†]	8032
L.mexicana	34	8249
L. braziliensis	35	8056
L. infantum	36	8227
T. brucei	11[*]	9962

[†] Ghosh *et al*. [23]. [*] Excluding the mini and intermediate chromosomes.

The *T.brucei* genome comprises minichromosomes (25-100kb), intermediate chromosomes (100-800kb) and megabase chromosomes (1.1-5.2 Mb). The 25 Mb diploid nuclear genome comprises 11 pairs of chromosomes excluding the mini and intermediate chromosomes [22]. A high degree of conservation of the gene order and sequence has been observed among the ~30 *Leishmania* species. The whole genome sequencing of the three trypanosomes (*L. major*, *T. brucei* and *T. cruzi*; Tri-tryp) revealed that the parasites implements strategies different from their eukaryotic counterpart in DNA repair and initiation of replication and the presence of the unusual mitochondrial DNA [24]. The organization of the protein coding genes into long, strand specific polycistronic structures and the absence of general transcription factors indicate that the regulatory mechanism involving RNA polymerase-II directed transcription to be different from other eukaryotes. Large number of RNA-binding protein encoded within the tri-tryp genome indicates the crucial role of post transcriptional regulation on gene expression [21].

The haploid genome of *Leishmania major* (Friedlin strain) was predicted to contain 911 RNA genes, 39 pseudogenes, and 8272 protein-coding genes, of which 36% were assigned putative functioning [21]. The haploid genome of *Trypanosoma brucei* was found to contain 9068 predicted genes, including ~900 pseudogenes and ~1700 *T. brucei*–specific genes. The large sub-telomeric region was found to contain 806 varieties of surface glycoproteins [22]. It was found that the outer surface the trypanosomes are covered with glycoconjugates. These glycoconjugates protect the parasite from the harsh environments present inside the phagolysosome of the host macrophages. Besides, these parasites are also known to secret glycoconjugates that modulate the macrophage signaling pathways and therefore may help in evading the host's immune response [25]. The diploid genome of *Trypanosoma cruzi* was predicted to contain 22570 protein coding genes, of which 12570 were present as allelic pairs

[24]. A large proportion of the genome comprised repeated sequences of retrotransposons, variant of surface proteins such as trans-sialidases, mucins, gp63s and 91300 copies of mucin-associated surface proteins.

A comparison of the metabolic pathways encoded by the genomes of the tri-tryp indicate that *T. brucei* exhibit least metabolic capability and *L. major* the maximum [22]. The differences observed could be ascribed to the horizontal transfer of genes of bacterial origin. Analysis of the genome of *L. major* revealed that most genes have orthologs in *T. brucei* and *T. cruzi* genome. Only 910 genes of *L. major* have no orthologs in tritryp while 74 were orthologous in *T. brucei* and 482 in *T. cruzi*. Some of these *L. major* specific genes such as some peptidases, transporters, and glycoconjugate biosynthesis components are responsible for the metabolic differences observed in tri-tryps. The two closely related genes encoding proteins that contain a microphage migration inhibition factor (MIF) domain was found to be evolutionarily related to the bacterial clade [21]. Their human counterparts have been reported to activate the macrophages against the invading parasite via a T helper cell 1- type pathway. Studies show that *Brugia malai* (filarial worm) secrets the homologs of human cytokine MIF which directs the macrophages to a T helper cell 2- type pathway. It is, therefore, possible that *leishmania spp.* may recruit similar strategies to evade the host's immune response and promote parasite survival.

Organizational differences of the RNA genes have been observed among the three species. The chromosome of *L.major* contains a single, large tandem array of the 28S, 18S, and 5.8S rRNA genes while they are scattered over different chromosomes in *T. brucei* and *T. cruzi* [21]. On the other hand the 5S rRNA is dispersed at 11 separate loci on multiple chromosomes in *L.major* but as a single tandem array in *T. brucei* and *T. cruzi.*

6. Drug Targets in the Leishmanial genome

Studies on *L. mexicana* promastigotes have shown that the MAP kinase (mitogen activated protein kinase) null mutants are viable and phenotypically similar to the wild-type cells but are non virulent. The essentiality of this gene, supposedly involved in cell signaling, for survival inside the mammalian host makes it an ideal drug target [26]. The enzyme ornithine decarboxylase (ODC) has been found to be involved in the early stages of polyamine biosynthesis in *T. gambiense*. The null mutants are viable in presence of high concentration of exogenous polyamines and its scarcity has been found to be lethal, a condition generally observed within the mammalian host [26]. ODC is thus essential for the survival of the parasite inside the host and therefore is an important drug target. Similarly, dihydrofolate reductase–thymidylate synthase (DHFR–TS; Bifunctional enzyme) null mutants are viable only in the presence of high concentrations of exogenous thymidine in *L. major*, indicating the essentality of this enzyme for the survival of the parasite.

Arginine and ornithine were shown to be essential for parasite growth and proliferation and the synthesis of trypanothione which defends parasitic cells from chemical and oxidative stress. Glutathione (GSH), polyamines and trypanothione were found to be essential for the survival of *T. brucei*. Removal of one of the two alleles of trypanothione reductase gene present in diploid parasite lead to reduced infectivity of activated macrophages while double knockouts could not be generated, suggesting that the enzyme is essential for the parasite [26]. In trypanosomes, the toxic metabolic byproduct methylglyoxal is converted to D-lactate via the trypathione dependent glyoxalase pathway. Unlike bacteria, none of the parasites contained methylglyoxal synthase but *T. cruzi* contains homologs of glyoxalase I and II while *T. brucei* contains glyoxalase II only. Trypanosomatids are dependent on trypanothione dependent

peroxidases for the removal of peroxides due to the lack of catalase and selenium dependent peroxidases. All Tritryp peroxidases are coupled to trypanothione and trypanothione reductase qualifying it as a potential therapeutic target.

Studies indicate that trypanosomatids are unable to synthesize purine, de-novo, due to the absence of the 9 of the 10 genes required for the synthesis of inosine monophoshate (IMP) from phophoribosyl pyrophosphate. The parasites are capable of scavenging purines from the media or host milieu and can synthesize phosphoribosylpyrophosphate (PRPP), the substrate for several enzymes of the purine salvage pathway [27]. They are capable of acquisition of any naturally occurring purine nucleobase or nucleotide to enrich its nucleotide pools. All the genes of this pathway have been identified either by molecular genetics or by genome sequencing approaches. The enzymes of this pathway, therefore, could be targeted for the design of inhibitors.

Trypanosomatids synthesize ergosterol and inhibitions of this pathway by antifungal agents have been found to be lethal except for bloodstream *T. brucei* which obtains cholesterol from the host. Most of the enzymes involved in ergosterol biosynthesis were present in all three parasites and could serve as drug targets.

In trypanosomes, one of the largest families of permeases is constituted by amino acid transporters [22]. *L. major* is predicted to have 29 members while *T. brucei* and *T. cruzi* have 38 and 42 members respectively. Unlike their mammalian host, they lack the functional amino-acid biosynthesis machinery and require exogenous proline as an energy source, glutamine for several biosynthetic pathways, cystine as an additional source of sulfur and tyrosine for the synthesis of protein. Except for the presence of putative biopterin dependent phenylanaline-4-hydroxylase, other ezymes of the aromatic amino acid oxidation pathways are absent in *L. major*. But

Leishmania spp. are auxotrophic for tyrosine indicating that the only enzyme belonging to this pathway may have functions other than those predicted.

The whole genome sequencing of representative organisms from all the domains of life has generated large scale genomic data which are curated, maintained and made available via different public databases such as the NCBI genome database.

7. The sugar nucleotide synthesis pathway of the trypanosomes

The cell surface of the *Leishmania* parasite are lined with a variety of novel glycoconjugates like lipophosphoglycan (LPG), glycoinositolphospholipids (GIPLs), membrane proteophosphoglycan (PPG) and glycosylated GPI-anchored proteins (e.g., GP63, PSA-2/GP46) which are essential for the survival, infectivity and virulence of the pathogen [21]. Generally in eukaryotes, sugar nucleotides are synthesized in the cytosol followed by their transportation to either the Golgi apparatus or to the endoplasmic reticulum to be used by glycosyl transferases [28]. In *T.brucei*, some of the enzymes of the sugar nucleotide synthesis pathway are reported to be present inside the glycosome.

The hexose transporters present in *T. brucei* and *T. cruzi* are unable to transport D-galactose and therefore UDP-glucose serves as the only source for the generation of the important metabolite [28]. In eukaryotes, UDP-glucose is required for the synthesis of glycoconjugates and secondary metabolites. In *T. brucei* UDP-glucose is required for the synthesis of a rare deoxynucleotide (β-D-glucosylhydroxymethyluracil) present in the DNA of the trypanosomatids.

The steps involved in the synthesis of UDP-glucose are: (i) glucose is phosphorylated by hexokinase to form glucose-6-phosphate. (ii) The enzyme phosphoglucomutase (PGM)

reversibly catalyses the formation of glucose-6-phosphate from glucose-1-phosphate and vice-versa, by the transfer of a phosphate between the C6 and C1 hydroxyl groups of glucose. (iii) glucose-1-phosphate is coupled to UTP by UDP glucose pyrophosphorylase leading to the formation of UDP-glucose (**Fig. 13**).

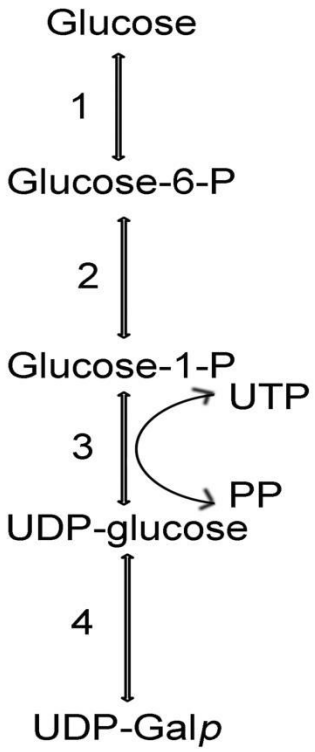


Figure 13: The sugar nucleotide synthesis pathway. Enzymes of this pathway are (1) Hexokinase (2) Phosphoglucomutase (3) UDP glucose pyrophosphorylase (4) UDP-galactose-4' epimerase.

A high degree of conservation of the phosphoglucomutase gene has been reported between the two sister species, *T. cruzi* and *L.major* [29]. Three isoenzymes of the PGM genes: PGM_1, PGM_2 and PGM_3 are present in human [30]. The NCBI GenBank database record indicated the presence of a single copy of the PGM gene on chromosome-21, encoding 589

amino acids in *L.major*, *L.infantum*, *L.mexicana, L.brazilliensis* and *L.donovani* while *T.cruzi* (strain CL Brener) contained two copies of the gene encoding 600 residues.

8. The enzyme: Phosphoglucomutase

The enzyme phosphoglucomutase (PGM), a member of the α-D-phosphohexomutase superfamily, reversibly catalyses the conversion of glucose-1-phosphate (G1P) to glucose-6-phosphate (G6P) by the transfer of a phosphate between the C6 and C1 hydroxyl groups of glucose. The structure comprises four domains which are arranged in an overall heart shape with the presence of a large central cleft harbouring the catalytic residues. The first three domains are similar and share a common core made up of α/β structures whereas domain four is different from the rest and is member of the TATA-box binding protein like fold.

The conformational flexibility of the fourth domain is a common feature observed among the family of phosphoglucomutase with the whole domain moving via a hinge type motion [31].

Available crystal structures of phosphoglucomutase from organisms belonging to the three domains of life are listed in **Table 3**.

Mutations in PGM1 gene in human (53% sequence identity with *L. major* PGM) cause a multisystem disorder with a marked decrease in enzyme activity [37]. The disease is characterized by reduced glycosylated serum glycoproteins. The disease is responsible for defects in the development of nervous system, psychomotor retardation, dysmorphic features, hypotonia, coagulation disorder and immunodeficiency.

In plants (*Arabidopsis thaliana*), phosphoglucomutase exists as plastidial (pPGM) and cytosolic (cPGM) isoforms [38]. pPGM is responsible for the transitory starch synthesis in chloroplast and cPGM is essential for glucose phosphate partitioning, synthesis of sucrose and

cell wall components. The two isoforms of cPGM (PGM 2 and PGM 3) are redundant in fuction and single mutant does not exhibit any alteration in primary metabolism compared to the wild type.

Table 3: Structurally characterized phosphoglucomutases from diverse organisms.

Domains of life	organism	PDB code	Resolution (Å)	References
Bacteria	*S. typhimurium*	2FUV	2.0	To be published
	S. enterica	3NA5	1.7	[32]
	B. melitensis	4HJH	2.1	To Be Published
	T. thermophilus	2Z0F	2.52	To Be Published
	B. thailandensis	3UW2	1.95	[33]
	P. aeruginosa	4MRQ	1.9	[34]
Archaea	*P. horikoshii*	1WQA	2.0	To be Published
	S. tokodaii	2F7L	2.8	To be Published
Eukaryota	*P. tetrauralia*	1KFI	2.4	[35]
	O. cuniculus	3PMG	2.4	[36]

The generation of amiRNA transgenic plant lacking cPGM activity exhibited reduced

growth, shorter root, and reduced production of seed; whereas those lacking both pPGM and cPGM exhibited dwarfness, premature death and inability to develop a functional inflorescence. In S. *cerevisiae* the PGM activity is attributed to different genes encoding the two isoforms PGM1 and PGM2 [39]. The S. *cerevisiae* null mutant lacking both the isoforms (ΔPGM1/ΔPGM2) is not viable in a media containing Gal as the only source of carbon. ΔPGM2 strain exhibits retarded growth coupled with large Ca^{2+} uptake and accumulation. However, expression of *T. cruzi* PGM in a ΔPGM1/ ΔPGM2 strain of S. *cerevisiae* was able to rescue the cell in a media containing Gal as the only source of carbon. Studies indicate that PGM activity is important to maintain intracellular Ca^{2+} homeostasis in S. *cerevisiae* since the level of Glc-6-P and of Glc-1-P play a decisive role in Ca^{2+} uptake and accumulation.

9. Conclusion

Trypanosomiasis and leishmaniasis fall under the most neglected disease category and poses a global threat to public health [40]. The first line of treatment still depends on toxic heavy metal based preparations. Therefore, it has become crucial to explore other newer options. The second chapter elaborates the comparison of whole sets of annotated coding sequences of one representative parasite species with that of human, resulting in the identification of proteins unique to the parasite. These enzymes were then utilized to identify their specific high affinity inhibitors. Crystal structures of the protein (target) - inhibitor complexes were then used to generate structure based pharmacophores. Finally, the ZINC database and Drug Bank has been searched utilizing this set of pharmacophores to generate a set of compounds which could serve as a library in the search for prospective antileishmanial drugs.

The third chapter deals with the solution of the structure of one of the parasitic-enzyme belonging to the indispensable sugar-nucleotide synthesis pathway. The presence of a pocket close to the hinge region may provide an opportunity for the successful design of inhibitors. The appendix section elaborates the initial purification as well as crystallization screening trials of several enzymes from *Leishmania sp*. Combination drug therapy often exhibits greater therapeutic efficacy compared to monotherapy (e.g. Cancer, AIDS and tuberculosis) [41]. In order to explore such possibilities, the two natural products: curcumin and quercetin were tested in combination with the existing antileishmanial drug paromomycin, against *L. major* promastigotes.

References

1. Cecílio P, Pérez-Cabezas B, Starém N, Maciel J, Rodrigues V, Cordeiro da Silva A: **Deception and manipulation: the arms of Leishmania, a successful parasite**. *Frontiers in immunology*. Review article. 2014, **5**(480): 1-16. doi: 10.3389/fimmu.2014.00480 .

2. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, Boer MD: **Leishmaniasis Worldwide and Global Estimates of Its Incidence.** *PLoS One* 2012, **7**(5): e35671. doi:10.1371/journal.pone.0035671.

3. Freitas-Junior LH, Chatelain E, Kim HA, Siqueira-Neto JL: **Visceral leishmaniasis treatment: What do we have, what do we need and how to deliver it?** *International Journal for Parasitology : Drugs and Drug Resistance* 2012, **2**: 11–19.

4. Reithinger R, Mohsen M, Aadil K, Sidiqi M, Erasmus P, Coleman PG: **Anthroponotic Cutaneous Leishmaniasis, Kabul, Afghanistan**. *Emerging Infectious Diseases* 2003, **9**: 727-729.

5. Pace D: **Leishmaniasis**. *Journal of Infection* 2014, **69:** S10eS18.

6. Olliaro P, Vaillant M, Arana B, Grogl M, Modabber F, Magill A, Lapujade O, Buffet P, Alvar J: **Methodology of Clinical Trials Aimed at Assessing Interventions for Cutaneous Leishmaniasis**. *PLOS Neglected Tropical Diseases* 2013, **7**(3): e2130. DOI: 10.1371/journal.pntd.0002130.

7. Croft SL, Sundar S, Fairlamb AH: **Drug Resistance in Leishmaniasis**. *Clinical Microbiology Reviews*. 2006, **19**(1): 111–126. doi:10.1128/CMR.19.1.111–126.2006

8. Nagle AS, Khare S, Kumar AB, Supek F, Buchynskyy A, Mathison CJN, Chennamaneni NK, Pendem N, Buckner FS, Gelb MH, Molteni V: **Recent Developments in Drug Discovery for Leishmaniasis and Human African Trypanosomiasis.** *Chem. Rev.* Special Issue: Drug Discovery and Development for Neglected Diseases 2014. dx.doi.org/10.1021/cr500365f.

9. Reddy AS, Pati SP, Kumar PP, Pradeep HN, Sastry GN: **Virtual Screening in Drug Discovery – A Computational Perspective**. *Current Protein and Peptide Science* 2007, **8**: 329-351.

10. Yang SY: **Pharmacophore modeling and applications in drug discovery: challenges and recent advances.** *Drug Discovery Today* 2010, **15**: 444-450.

11. Cavalli A, Poluzzi E, Ponti FD, Recanatini M: **Toward a Pharmacophore for Drugs Inducing the Long QT Syndrome: Insights from a CoMFA Study of HERG K+ Channel Blockers.** *J. Med. Chem.* 2002, **45:** 3844-3853.

12. Chasman D (Editor): **Protein Structure: Determination, Analysis, and Applications for Drug Discovery**. *CRC Press* 2003, 457-458.

13. Harikishore A, Niang M, Rajan S, Preiser PR, Yoon HS: **Small molecule *Plasmodium* FKBP35 inhibitor as a potential antimalaria agent.** *Scientific Reports* 2013, **3**(2501). doi:10.1038/srep02501.

14. Böhm HJ, Flohr A, Stahl M: **Scaffold hopping.** *Drug Discovery Today: Technologies* 2004, **1**(3): 217-224. doi: 10.1016/j.ddtec.2004.10.009.

15. Sun H, Tawa G, Wallqvist A: **Classification of scaffold hopping approaches.** *Drug Discov Today* 2012, **17**(7-8): 310–324. doi:10.1016/j.drudis.2011.10.024.

16. Mysinger MM, Carchia M, Irwin JJ, Shoichet BK: **Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking.** *J. Med. Chem.* 2012, **55**: 6582−6594.

17. Fawcett T: **An introduction to ROC analysis.** *Pattern Recognition Letters* 2006, **27**: 861–874. doi:10.1016/j.patrec.2005.10.010.

18. Irwin JJ, Shoichet BK: **ZINC – A Free Database of Commercially Available Compounds for Virtual Screening.** *J Chem Inf Model* 2005, **45**(1): 177–182.

19. Weininger D, Chem J: **SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules.** *J. Chem. Inf. Comput. Sci.* 1988, **28** (1): 31–36. DOI: 10.1021/ci00057a005.

20. Wishart DS, Knox C, Guo AC, Cheng D, Shrivastava S, Tzur D, Gautam B, Hassanali M: **DrugBank: a knowledgebase for drugs, drug actions and drug targets**. *Nucleic Acids Research* 2008, **36**: D901–D906. doi:10.1093/nar/gkm958.

21. Ivens AC *et al*: **The Genome of the Kinetoplastid Parasite, *Leishmania major*.** *Science* 2005, **309**(436): 436-442. DOI: 10.1126/science.1112680.

22. Berriman M *et al.*: **The Genome of the African Trypanosome *Trypanosoma brucei*.** *Science* 2005, **309**(416): 416-422. DOI: 10.1126/science.1112642.

23. Ghosh SS, Mukerjee S, Adhya S: **Chromosome profile of Leishmania donovani: Interstrain and interspecific variations**. *J. Biosci* 1998, **23**(3): 247-254.

24. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G *et al*: **The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease**. *Science* 2005, **309**(5733):409-415.

25. Descoteaux A, Turco SJ: **Glycoconjugates in Leishmania infectivity**. *Biochimica et Biophysica Acta* 1999, **1455**: 341-352.

26. Barrett MP, Mottram JC, Coombs GH: **Recent advances in identifying and validating drug targets in trypanosomes and leishmanias.** Trends in microbiology 1999, **7**(2): 82-88.

27. Boitz JM, Ullman B, Jardim A, Carter NS: **Purine salvage in Leishmania: complex or simple by design?** *Trends in Parasitology* 2012, **28**(8): 345-352.

28. Bandini G, Mariño K, Güther MLS, Wernimont AK, Kuettel S, Qiu W, Afzal S, Kelner A, Hui R, Ferguson MAJ: **Phosphoglucomutase is absent in Trypanosoma brucei and redundantly substituted by phosphomannomutase and phospho-N-acetylglucosamine mutase**. *Molecular Microbiology* 2012, **85**(3): 513-534.

29. Penha LL, Sant'Anna CB, Mendonça-Previato L, Cunha-e-Silva NL, Previato JO, Lima AP: **Sorting of phosphoglucomutase to glycosomes in Trypanosoma cruzi is mediated by an internal domain**. *Glycobiology* 2009, **19**(12): 1462-1472.

30. Quick CB, Fisher RA, Harris H: A kinetic study of the isozymes determined by the three human phosphoglucomutase loci PGM1, PGM2, and PGM3. *Eur. J. Biochem* 1974, **42**: 511-517.

31. Schramm AM, Mehra-Chaudhary R, Furdui CM, Beamer LJ: **Backbone Flexibility, Conformational Change, and Catalysis in a Phosphohexomutase from *Pseudomonas***

aeruginosa. *Biochemistry* 2008**, 47**: 9154–9162.

32. Mehra-Chaudhary R, Mick J, Tanner JJ, Henzl MT, Beamer LJ: **Crystal structure of a bacterial phosphoglucomutase, an enzyme involved in the virulence of multiple human pathogens.** *Proteins* 2011. **79**(4): 1215-29. doi: 10.1002/prot.22957.

33. Baugh L, Gallagher LA, Patrapuvich R, Clifton MC, Gardberg AS, Edwards TE, Armour B, Begley DW, Dieterich SH, Dranow DM, Abendroth J, Fairman JW, Fox D 3rd, Staker BL, Phan I, Gillespie A, Choi R, Nakazawa-Hewitt S, Nguyen MT, Napuli A, Barrett L, Buchko GW, Stacy R, Myler PJ, Stewart LJ, Manoil C, Van Voorhis WC: **Combining functional and structural genomics to sample the essential Burkholderia structome.** *PLoS One* 2013, **8**(1): e53851. doi: 10.1371/journal.pone.0053851.

34. Lee Y, Villar MT, Artigues A, Beamer LJ: **Promotion of enzyme flexibility by dephosphorylation and coupling to the catalytic mechanism of a phosphohexomutase.** J Biol Chem. 2014, **289**(8): 4674-82. doi: 10.1074/jbc.M113.532226.

35. Müller S, Diederichs K, Breed J, Kissmehl R, Hauser K, Plattner H, Welte W: **Crystal structure analysis of the exocytosis-sensitive phosphoprotein, pp63/parafusin (phosphoglucomutase), from Paramecium reveals significant conformational variability.** J Mol Biol. 2002, **315**(2): 141-53.

36. Liu Y, Ray WJ Jr, Baranidharan S: **Structure of rabbit muscle phosphoglucomutase refined at 2.4 A resolution.** *Acta Crystallogr D Biol Crystallogr*. 1997, **53**(4): 392-405.

37. Tegtmeyer LC, Rust S, Scherpenzeel MV, Ng BG, Losfeld ME, Timal S, Raymond K, He P, Ichikawa M, Veltman J, Huijben K, Shin YS, Sharma V, Adamowicz M,

Lammens M *et al*.: **Multiple phenotypes in phosphoglucomutase 1 deficiency**. New Engl. J. Med. 2014, 370(6): 533-542.

38. Malinova I, Kunz HH, Alseekh S, Herbst K, Fernie AR, Gierth M, Fettke J: **Reduction of the Cytosolic Phosphoglucomutase in Arabidopsis Reveals Impact on Plant Growth, Seed and Root Development, and Carbohydrate Partitioning**. PLOS ONE 2014, 9 (**11**): e112468.

39. Penha LL., Mendonça-Previato L, Previato JO, Scharfstein J, Heise NC, De A, Lima AP: **Cloning and characterization of the phosphoglucomutase of Trypanosoma cruzi and functional complementation of a Saccharomyces cerevisiae PGM null mutant.** *Glycobiology* 2005, **15**(12): 1359-1367.

40. Changtam C, de Koning HP, Ibrahim H, Sajid MS, Gould MK, Suksamrarn A: **Curcuminoid analogs with potent activity against Trypanosoma and Leishmania species.** *Eur J Med Chem.* 2010, **45**(3): 941-56. doi: 10.1016/j.ejmech.2009.11.035.

41. Fivelman QL, Adagu IS, Warhurst DC: **Modified fixed-ratio isobologram method for studying in vitro interactions between atovaquone and proguanil or dihydroartemisinin against drug-resistant strains of Plasmodium falciparum.** *Antimicrob Agents Chemother*. 2004, **48**(11): 4097-102.

CHAPTER II

1. Introduction

Leishmaniasis, a broad spectrum of diseases, is caused by more than 20 sister species of protozoa belonging to the family Trypanosomatidae and genus *Leishmania*. These diseases are generally classified into three forms: visceral (VL), cutaneous (CL) and mucocutaneous (MCL), of which VL is lethal if left untreated, whereas CL, MCL generally self cure, though with the possibility of leaving permanent scars on the patient. The Indian subcontinent along with Sudan and Brazil account for the overwhelming majority of cases in VL, while the incidence of CL predominantly occurs in Afghanistan, Algeria, Brazil, Iran, Peru, Saudi Arabia and Syria [1] . Overall, about 10 million people are affected worldwide. The vector for this disease is the phlebotominae sand fly, which injects the parasites into the host in the course of a blood meal. The parasites thus exist in two forms: as mobile, flagellated promastigotes in the gut of the sandfly and non-motile, non-flagellated amastigotes which multiply within the phagolysosomal compartment of the macrophage in the mammalian host [2].

The first line of defense against the parasites traditionally continues to be generic pentavalent antimonials (sodium stibogluconate), especially in those regions where resistant strains have yet to appear. With the emergence of strains resistant to antimonials (especially in Bihar state of India) [3-5] second line drugs such as amphotericin-B (along with its liposomal formulations) and miltefosine are being extensively used [6]. However, both these drugs are more expensive than antimonials, toxic with reportedly severe side effects. Pentamidine and paromomycin are other drugs currently in use though their ready availability in endemic areas appears to be limited [7]. It is thus clear that there is an urgent need to search for and identify therapeutic alternatives to combat the disease.

With the availability of full genome sequences, search for drug targets in pathogenic organisms have been greatly facilitated. Comparative genomics allows the identification of genes unique to an organism, determination of parasitic genes absent in human and the evolutionary conservation of genes, probably reflecting upon their essentiality [8]. Gene conservation across pathogenic species also gives the added advantage that a single broad based antiparasitic targeting a conserved protein, could be used as a drug for several ailments. The genomes of five leishmanial species *L. major*, *L. infantum*, *L. donovani* , *L. mexicana* and *L. braziliensis* have been sequenced; with the first three consisting 36 chromosomes each, while *L. braziliensis* contains only 35. Notably, *L. braziliensis* has been assigned to a different subgenus *Leishmania* (*Viannia*) *sp.* and is thus somewhat distantly related to the others [12] ,which belong to the subgenus *Leishmania* (*Leishmania*) *sp.*. This reduction in chromosome number in *L. braziliensis* is due to a fusion event joining chromosome 20 and 34 (as numbered in *L. major*). Likewise, *L. mexicana* is two chromosomes less due to two fusions between four chromosomes (chromosome 8 and 29; chromosome 20 and 36). These genomes have approximately 8300 protein coding regions of which only about 40% can be ascribed a putative function [9-11]. In addition, the genomes of *T. brucei* (11 chromosomes) and *T. cruzi* are also available. Generally, the genomes of kinetoplastidae exhibit a high degree of synteny (conservation of gene order) in the organization of their genomes. Comparison between the genomes of *T. brucei*, *T. cruzi* and *L. major* revealed a conserved core of approximately 6200 trypanosomatid genes and about 1000 ORFs [11] were notable for their presence in the genome of *L. major* alone. Further, upon comparing the genomes of leishmanial species, 5, 26 and 47 genes were identified to be exclusively and specifically present in the genomes of *L. major* , *L. infantum* and *L. braziliensis* respectively [12].

Leishmanial genomes consist of several novel metabolic pathways whose enzymes could serve as potential drug targets. Some of the distinctive features of these genomes include the presence of atypical protein kinases lacking the SH2, SH3, FN-III and immunoglobulin like domains which occur most frequently in humans [14, 15]. The cellular surface of leishmania consists of several unique glycoproteins which are essential for immune evasion and host – parasite interaction. The most abundant of these glycoproteins are attached to the surface of the plasma membrane via GPI (glucosylphosphatidyl inositol) anchors, which are essential for parasitic survival. Other novel pathways involve trypanothione metabolism, essential for cell growth and differentiation, which is replaced by glutathione in humans. The first enzyme in trypanothione synthesis is the enzyme ornithine decarboxylase targeted by the drug diflouromethyl ornithine, prescribed for human sleeping sickness. Enzymes of the glycolytic pathway, ergosterol synthesis in sterol metabolism and the purine salvage pathway also offer potential drug targets for therapeutic intervention [14]. Some of these pathways will be discussed in greater detail in the later sections of this paper.

Due to the exponential increase in genomic information, researchers are now confronted with a rapidly expanded list of gene products from which to select prospective targets. Several scoring schemes have been proposed which surveys the genome of a pathogenic organism and ranks genes according to their potential as drug targets [8, 16, 17]. Most schemes give a high weightage to the essentiality of the protein in the life cycle of the parasite, conservation of the target among different sister species and its corresponding absence in humans. Experimentally, either the lethality of gene deletion or insertion of transposons into the selected gene has been used to determine its essentiality. Non-essential genes could also be selected as targets, provided they play a vital role in the infective virulence of the pathogens. Other considerations includes

the assayability of the protein, expression level during the life cycle of the organism possibly determined by microarray data and computational flux based analysis to gauge the effect of protein inhibition on the integrity of biochemical networks. In this connection, the TDR-target database is one of the most well cited amongst such databases [16]. This database consists of an exhaustive list of drug targets from the genomes of *L. major*, *T. brucei*, *M. leprae* and a host of other pathogens responsible for neglected tropical diseases. A useful feature of the database is its ability to prioritize a set of drug targets, where each criterion is assigned a weight and there is flexibility to change the weights associated with different factors (pathogenicity, essentiality, etc.) in the scoring scheme to extract a 'custom-made' list of targets relevant to the research interest of the user.

Several crystal structures of trypanosomal proteins, either individually or complexed with inhibitors are currently available in the Protein Data Bank, such as trypanothione reductase (*T. brucei),* trypanothione synthetase (*L. major*), pteridine reductase 1 (*L. major*), nucleoside hydrolase (*T. vivax*) and ATP dependent phosphofructokinase (*T. brucei*), which provide a detailed three dimensional structure of their active sites facilitating the design of specific inhibitors. In order to generate a library of prospective ligands which could have high affinities towards the active sites of targeted proteins, drug databases could be searched with structure based pharmacophores derived from protein ligand complexes. 'Scaffold hopping' or 'chemotype switching' [18-20], which involves identifying molecules with dissimilar backbone structures yet exhibiting very similar pharmacological properties, is one of the widely used techniques to generate compound libraries for eventual screening [21-25]. Lately, considerable success has been achieved in the application of structure based pharmacophores in the identification of lead compounds [21-25]. In the current work the human and the *L. major*

genome have been compared to identify a set of proteins unique to the parasite. Crystal structures of these proteins or those of closely related homologues have been extracted from the Protein Data Bank (PDB) and the literature has been extensively surveyed to identify their specific high affinity inhibitors. Crystal structures of the protein (target) - inhibitor complexes have then been utilized to generate structure based pharmacophores. In case the crystal structures of the ligand bound targets were not available in the PDB, the inhibitors were computationally docked into the active sites of their receptors. Finally, the ZINC database and Drug Bank has been searched utilizing this set of pharmacophores to generate a set of compounds which could serve as a library in the search for prospective antileishmanial drugs.

2. Material and methods

The annotated coding sequences (CDS) from the genomes of *L. major* (8316 CDS), *L. donovani* (8032 CDS), *L. mexicana* (8249 CDS), *L. braziliensis* (8056 CDS), *L. infantum* (8227 CDS), *T. brucei* (9962 CDS) and human (~41961 CDS) were downloaded from the NCBI genome database (http://www.ncbi.nlm.nih.gov/genome/; updated as on September 1, 2013). The standalone BLAST [26] was obtained from the NCBI ftp server (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/). Here BLAST refers to the BLASTp program of the NCBI standalone BLAST package which aligns protein amino acid sequences. To compare the annotated protein sequences between human and *L. major*(CDS's) , the CDS's from *L. major* were first fed in as a query (in FASTA format) whereas the human CDS's were processed in the makeblastdb tool to form a BLAST database. This was followed by a second run of BLAST (with identical parameters) in which the human sequences were considered to be 'query' and the *L. major* proteins the database. Proteins which simultaneously passed identical filters in both the runs of BLAST were considered for the second step in the

pipeline. A sole exception was made in the case of ATP dependent phosphofructokinase (PFK) in the first filter as the possibility of PFK being a drug target for trypanosomatids has been mentioned in the literature [27-29]. To compare *L. major* CDS's with those of closest – related species, *L. major* proteins were fed in as a query and sequences from the other genome were made the database. All alignments with an E-value less than one were output from the program and default options were used for all other parameters. The software to analyze the BLAST outputs was developed locally in C/C++, Perl and the final results were displayed on Microsoft Excel sheets for further analysis.

In order to identify the unique metabolome in *L. major* with respect to its human host the 'Comparative Analysis and Statistics' option in the BioCyc Database (http://www.biocyc.org/) was used [30]. The metabolic pathways and enzymes associated with this unique set of metabolites from *L.major* were manually culled from the LeishCyc database (http://biocyc.org/LEISH/organism-summary?object=LEISH).

Template based homology modeling was performed using the MODELLER software both in the standalone mode and also implemented in Accelrys Discovery Studio 2.5. In addition, the GUI version of Modeller 9.11v (Easy Modeller 4.0) [31] was also used.

GOLD 5.2 (http://www.ccdc.cam.ac.uk/Solutions/GoldSuite/Pages/GOLD.aspx) was used to dock specific ligands onto the active site of their corresponding enzymes with the following parameters: population size 100, selection pressure 1.1, number of operations 100000(min) - 125000(max), islands 5, niche size 2, crossover frequency 95, mutation frequency 95, migration frequency 10 and search efficiency 100%. The program was run at least 10 times in order to confirm the best docking solution which was identified based on three criteria a) the

CHEMPLP score b) rmsd between the docked solution and the initial placement of the ligand and c) visual survey and examination of the contacts between the docked ligand and protein. In case sufficient prior information was already available with regard to the position of the ligand in the active site of the protein (such as in Group B : See Section on **The Protein – Ligand Complexes**), the docking solution with the minimum rmsd (generally less than 1.0 Å) and whose CHEMPLP score was amongst the top three (comparable to the score derived from the original protein – ligand complex available in the PDB), was accepted as the most reasonable solution, subsequent to the visual inspection of the ligand bound active site. In cases where such information was either ambiguous or limited (Group C : See Section on **The Protein – Ligand Complexes)** the threshold on rmsd was relaxed to about 2.0 - 2.75Å, the interactions of the ligand with active site were visually examined and the solutions grouped into sets with similar geometry with respect to the binding site. Amongst these sets, the pose with the maximum number of physically meaningful interactions and the best CHEMPLP score was accepted as the most favoured solution. All solutions which exhibited significant displacement of the ligand from the putative active site of the protein (>3.75Å) were not considered. The decisions regarding both ligand flexibility and the flexibility of residues constituting the active site were decided on a case by case basis which will be discussed below (in the **Results and Discussion** section). Generally, in case the protein-ligand complex was available in the Protein Data Bank (PDB) and the leishmanial protein modeled utilizing the molecule from such a complex as a template, the ligand was transferred onto the modeled protein by utilizing the rotation matrix, translation vector derived from the superposition of the modeled protein to the template (Dali Server; http://ekhidna.biocenter.helsinki.fi/dali_lite/start) [32]. Most often in such instances both the ligand and the active site were held rigid whilst docking with GOLD. For ligands, whose

complexes with the targeted protein (from *L. major*) or its close homologue were not available in the PDB, the ligand coordinates were derived from its closest related structure (as found as a complex in the PDB) by fourth atom fixing techniques and energy minimized by the semi-empirical quantum mechanical method in the program GAMESS [33, 34]. It goes without saying that the protein in such a complex would either be the target (*L. major*) or a close homologue from a sister trypanosomatid species. With the initial placement of ligand into the protein active site using the same techniques mentioned above, the newly added chemical groups to the parent compound (obtained from the PDB) were rendered flexible in the subsequent docking by GOLD, in addition to selected active site residues which could provide steric hindrance in the docking process. These residues were identified by visual examination of the initial docked position and examination of the list of contacts. In GAMESS the self-consistent field wave function with the semi-empirical basis set (AM1 model Hamiltonian) was used and the optimum tolerance of the energy minimization cycle was set to 1.0e-5. Where no information was available with regard to the association of the ligand with the target protein, blind docking was attempted subsequent to the placement of the ligand at the centroid of the putative active site of the enzyme.

LigandScout version 3.12 (build 20130912) was used to generate the structure based pharmacophores from the crystallographic and modeled/docked complexes, with manual monitoring of the entire process at every stage. To validate the pharmacophores, the ligand along with other active compounds (with relatively less IC_{50} values though with similar structures) were made the kernel of a database which also included decoys generated from the D.U.D.E. decoy generator (http://www.dude.docking.org/generate). The specific ligand, along with other actives and decoys were next submitted to OMEGA 2.5.1.4 [35] to generate two conformers per decoy and each of the active molecules from the docked/crystal structures. Thus,

the database for each ligand (inclusive of the other actives and decoys) consisted of about 1800 molecular conformers in all. The library generating tool of LigandScout was then utilized to convert the database into a library (*.ldb format), prior to searching the library with the corresponding pharmacophore. Invariably, the specific ligand used to derive the structure based pharmacophore would be detected at the topmost rank. Every validation database was split into two, one consisting of actives and the other of decoys. Ligand screening option in LigandScout was then invoked to search both the databases with the specific pharmacophore as the query (with the advanced options, scoring function: 'Pharmacophore - Fit'; Screening Mode: 'Match all query features'; Retrieval Mode: 'Get best matching conformation').For every case, the maximum number of omitted features were varied to get optimal results and the ROC curve. Those pharmacophore queries, which gave screening results with area under the ROC curve less than 0.75 were not utilized in subsequent calculations. Finally, the ZINC database was searched by all the structure based pharmacophores derived from the protein ligand complexes.

3. Results & Discussion

3.1 Comparative Genomics – Human versus Leishmania

The whole set of annotated protein-coding genes from *L. major* genome (8316 CDS) was compared against the ones from human genome (41961 CDS) and those parasitic proteins (4991 among 8316 sequences) which could not align with any human gene, (**first filter**) with pident (percentage sequence identity) >35% and a simultaneous query coverage >50% , in two way reciprocal BLAST runs with *L. major* as query, human as database and vice versa (**Materials & Methods**) , were selected for the next filter. Hypothetical sequences (4407 CDS) were removed from this list (**second filter**) and the Protein Databank (PDB) searched for each of the remaining

sequences, which amounted to a total of 584 putative sequences. The PDB database was downloaded from the RCSB-PDB (http://www.rcsb.org/pdb/home/home.do) website and incorporated into BLAST using the methods given above (**Materials and Methods**). Only those genes were selected for subsequent analysis (a total 90 sequences) which recorded hits in the PDB with >40 % sequence identity and a simultaneous query coverage >75% (**third filter**). Each gene from this set (consisting of *L. major* proteins alone) were then checked for BLAST hits (pident >40% & query coverage >75%) in the genomes of *L. donovani*, *L. mexicana*, *L. braziliensis*, *L. infantum* and also in the clusters of orthologous proteins in the Tritryp database (based on the OrthoMCL annotation [36]).

Upon merging both sets of data, only those (*L. major*) proteins with homologues/BLAST hits in all the five genomes were retained (**fourth filter**). The final set of genes consisted of a total of 86 polypeptide chains (Supplementary **File 1,** Supplementary files are provided separately with the CD-ROM). The schematic representation of the successive filters to arrive final list of drug targets is described in **Figure 1**.The separated list of hypothetical sequences (5135 CDS) were independently searched in the PDB and eighteen sequences scored hits satisfying the above criteria (given in **Supplementary File 2**). Upon the application of more stringent criteria (pident > 25 and query coverage > 33) in the first filter, 47 out of the original 86 genes satisfied the new threshold values, of which 13 genes were retained from the first thirty proteins of the original set (**Supplementary File 1**).

3.2 Target Prioritization

This list of parasitic proteins were then sorted according to a "*weighted union*" scoring scheme based on the following factors: i) **essentiality** as determined from experimental studies,

ii) **virulence factor** iii) **expression profile** and iv) whether the natural substrate to the protein is a ligand unique to *Leishmania spp.* with respect to human.

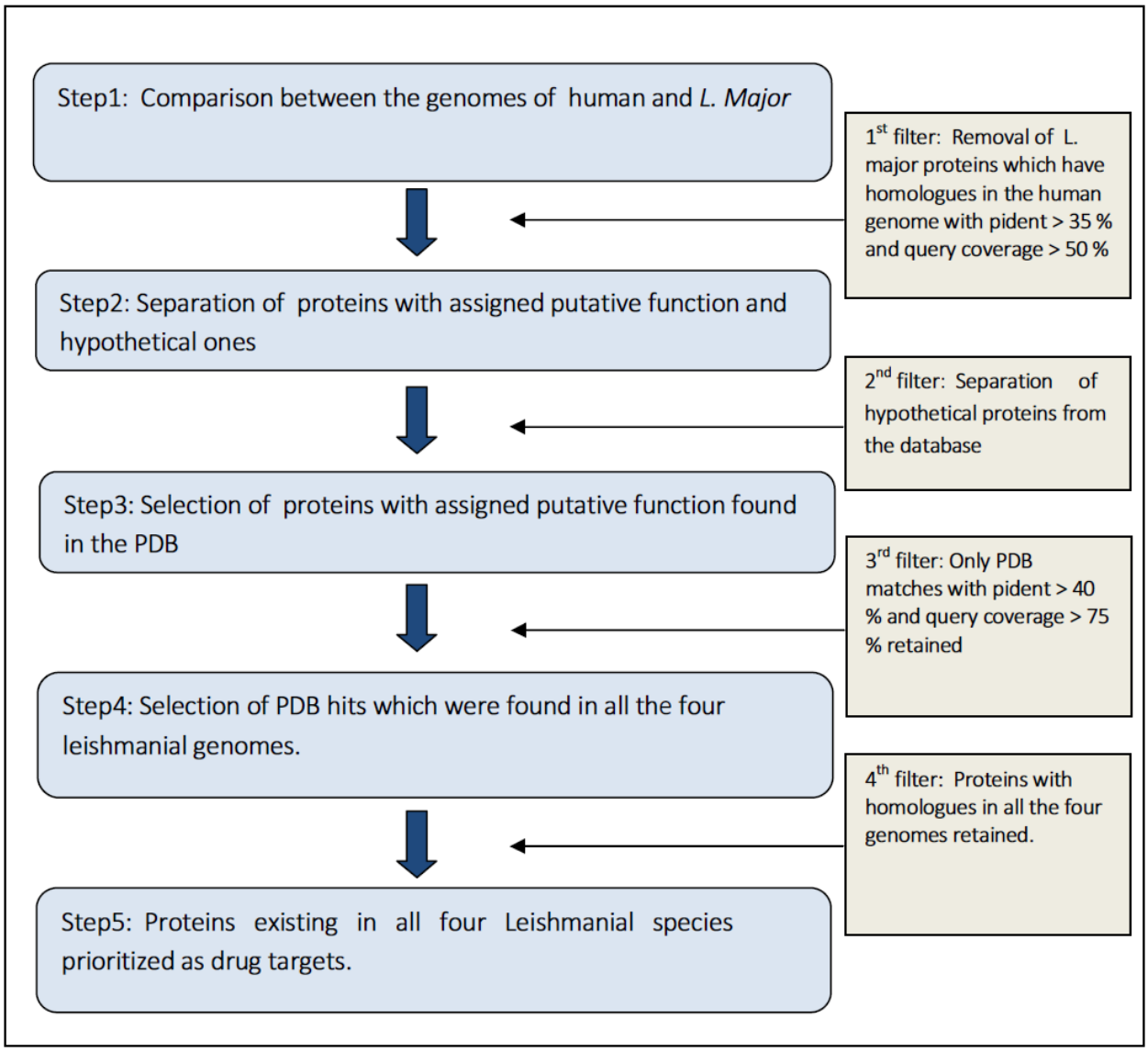


Figure 1: Schematic representation of steps involved in the selection of drug targets from the genome of *L. major* (**Supplementary File 1,** Supplementary files are provided separately with the CD-ROM). The separated set of hypothetical proteins were independently searched against the PDB (**Supplementary File 2**).

Information regarding essentiality and virulence were obtained by an extensive literature survey, in addition to consulting the TDR database**.** A list of metabolites unique to *Leishmania spp*. were

identified by searching the leishmanial and human metabolome databases [37]. The metabolome of *L. major* was obtained from the Biocyc Database (http://biocyc.org/LEISH/class-tree?object=Compounds) (see **Materials and Method**) and the corresponding metabolome from human was available in the Human Metabolome database (HumanCyc; http://biocyc.org/HUMAN/organism-summary?object=HUMAN). Utilizing bioinformatics tools available in the *Biocyc,* the two metabolomes were compared and the ligands unique to *Leishmania spp.* were filtered out. From this comparison 129 metabolites were identified to be unique to the parasite. The full set of

enzymes corresponding to these substrates were assembled into a blast database and run against the initial list of 86 genes. Only eight polypeptide chains from this list of 86 proteins were found to be associated with unique ligands. As has been mentioned previously, leishmania has two stages in its life cycle (amastigote and promastigote) and the information whether the gene was 'constitutively expressed' in both stages or in only one of them was obtained primarily from GeneDB [38], which provides a convenient platform to cull information with regard to leishmanial gene expression from the work of Leifso *et al.* [39]. A score of 100 was awarded on the full satisfaction of any one of the criteria given above and thus the highest possible score obtained by any protein could be 400. Targets lacking experimental data with regard to essentiality were still given 50 in case strong arguments existed in favour of their being essential genes (e.g. phosphofructokinase in the glycolytic pathway) and 20 if the protein was found to be indispensable in a sister species. A score of 100 was assigned to the gene which was constitutively expressed in both stages and 50 if expressed in only one of them. The final list of prioritized proteins are given in **(Supplementary File 1,** Supplementary files are provided separately with the CD-ROM**)** and the first thirty proteins from this list (**Table 1**) are described

in some detail in the context of unique metabolic pathways of leishmania parasites. The only available information with regard to the 18 hypothetical proteins (**Supplementary File 2,** provided separately with the CD-ROM)) was that they are constitutively expressed in both stages of the leishmanial life cycle [39] , and thus upon prioritization, these proteins (all with an identical score of 100) did not find their place amongst the first thirty, in the list of annotated proteins (given in **Table 1**).

As expected, prominent amongst the list of prioritized proteins (**Table 1**) are three enzymes associated with trypanothione i)**trypanothione reductase** (**TR**: 1), ii)putative **trypanothione synthetase** (**TS**: 3) and iii)**trypanothione - dependent glyoxalase I** (**GLO1**: 10). The trypanothione system in leishmania which replaces the ubiquitous glutathione system present in humans, enables the parasite to survive the high oxidative stress found in the host immune system and the presence of toxic heavy metals [40]. Both **trypanothione synthetase** (which synthesizes trypanothione from glutathione and spermidine) and **trypanothione reductase** (which keeps it in its reduced form in the presence of NADPH) are attractive drug targets [41], as this system is the only pathway involved in the crucial regulation of oxidative stress in the parasites. Reduced trypanothione in turn causes the reduction of tryparedoxin which then transfers electrons to the recycling enzyme tryparedoxin peroxidase [40]. Although **TR** and human glutathione reductase(GR) exhibits 35% sequence identity and shares many physicochemical properties , yet their corresponding active sites are different due to their diverse substrate specificities [42]; **TR** binding only to the oxidized forms of positively charged glutathionyl – polyamine conjugates whereas human GR associates only with negatively charged glutathione [43].

Table 1: Drug Targets from the Genome of *L. major*. Prioritized list of drug targets consisting of the first thirty proteins. Information regarding their metabolic pathway, genebank ID and association with a ligand unique to leishmania w.r.t human are also included in the table. Proteins given in bold were subsequently used for pharmacophore calculations.

NO.	TARGET PROTEINS	METABOLIC PATHWAY	GENE NAME (GENEDB ID)	UNIQUE LIGAND	PRIORITIZATION SCORE
1	**Trypanothione reductase**	**trypanothione metabolism**	**LmjF05.0350**	**Trypanothione**	**400**
2	**Pteridine reductase 1**	**Reductase**	**LmjF23.0270**	**tetrahydropteroyltri-L-glutamate, 5-methyltetrahydropteroyltri-L-glutamate**	**400**
3	**Putative trypanothione synthetase**	**trypanothione metabolism**	**LmjF27.1870**	**Trypanothione**	**350**
4	**UDP-galactopyranose mutase**	**sugar metabolism**	**LmjF18.0200**	**β-D-ribopyranose**	**350**
5	**Putative udp-glc 4'-epimerase**	**sugar Metabolism**	**LmjF33.2300**	**Absent**	**300**
6	Inhibitor of cysteine peptidase	peptidase/protease	LmjF24.1770	Absent	220
7	Peptidase m20/m25/m40 family-like protein	peptidase/protease	LmjF31.1890	Absent	220
8	Putative endoribonuclease L-PSP (pb5)	RNase	LmjF23.0200	Absent	220
9	Putative peptidase M20/M25/M40	peptidase/protease	Lmjf33.1610	Absent	220
10	**Trypanothione-dependent glyoxalase I**	**trypanothione metabolism**	**GLO1**	**S-lactoyl trypanothione**	**220**
11	**Putative deoxyuridine triphosphatase**	**purine/pyrimidine metabolism**	**DUT**	**Absent**	**200**
12	Putative metacaspase protein	Peptidase/protease	LmjF35.1580	Absent	200

13a	Macrophage migration inhibitory factor-like protein	host-parasite interaction, macrophage migration inhibitory factor	MIF1	Absent	200
13b	Macrophage migration inhibitory factor-like protein	host-parasite interaction, macrophage migration inhibitory factor	MIF2	Absent	200
14	Putative calpain-like cysteine peptidase	peptidase/protease	LmjF20.1185	Absent	200
15	Putative proteasome activator protein pa26	peptidase/protease	LmjF35.0750	Absent	200
16a	Putative serine peptidase	peptidase/protease	LmjF27.2630	Absent	200
16b	Putative serine peptidase	peptidase/protease	LmjF29.1270	Absent	200
17	Putative NADP-dependent alcohol dehydrogenase	glycolytic pathway/ gluconeogenesis/ glycerolipid metabolism	LmjF23.0360	1-alkyl-2acyl-phosphatidyl-inositol (an alcohol)	200
18	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase	amino acid metabolism	LmjF31.0010	5-methyltetrahydropteroyltri-L-glutamate	200
19	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	glycolytic pathway	PGAM	Absent	150
20	ADF/Cofilin	cellular motility	LmjF29.0510	Absent	150
21	**ATP-dependent phosphofructokinase**	**glycolytic pathway**	**LmjF29.2510**	**Absent**	**150**
22	Glucokinase	glycolytic pathway	LmjF36.2320	Absent	150
23	Glycerol-3-phosphate dehydrogenase	glycerophospholipid metabolism	GPD	absent	150

	[NAD+],glycosomal				
24	Nonspecific nucleoside hydrolase	purine/pyrimidine metabolism	NH	Absent	150
25	Putative dipeptylcarboxypeptidase	peptidase/protease	DCP	Absent	150
26	UDP-sugar pyrophosphorylase	sugar metabolism	USP	Absent	150
27	putative glutamate dehydrogenase	Amino acid metabolism	LmjF28.2910	Absent	120
28	putative isocitrate dehydrogenase	TCA cycle	LmjF33.2550	Absent	120
29	putative OMPDCase-OPRTase	Pyrimidine metabolism	LmjF16.0550	Absent	120
30	3-mercaptopyruvate sulfurtransferase	Amino acid metabolism	LmjF05.0970	Absent	120

The difference in specificity is primarily due to the presence of five amino acid residues in the **TR** active site, which confers enhanced hydrophobicity, negative charge and wider access to its binding pocket relative to human GR [43]. Several inhibitors specifically designed for TR have yet to be entirely successful as drugs, probably due to the wide active site of the enzyme which poses obstacles for structure-based drug design, coupled to the pharmacokinetic properties of the inhibitors [43]. In addition, trypanothione is also implicated in the Glyoxalase I, II systems in the parasite (again replacing glutathione in humans) which is responsible for the removal of toxic and mutagenic methylglyoxal formed as a byproduct of glycolysis. The crystal structure of leishmanial **glyoxalase I** (**GLO I**) reveals differences with respect to the corresponding human enzyme in its active site architecture [44], which includes increased negative charge and hydrophobicity along with the truncation of a loop which could be involved in the catalytic activity of the human enzyme.

The surface glycocalyx of *Leishmania spp.* consists of several unique sugars and glycoconjugates which mediate host – parasite interaction and virulence. A significant fraction of these glycoconjugates consists of lipophosphoglycans (LPG) implicated in the adhesion of leishmania to the host cell and glycoinositolphospholipids (GIPLs) involved in pathogenesis [45]. Both LPGs and GIPLs, have β-galactofuranose (β-Gal*f*) as one of their main constituents, an unusual sugar not found in vertebrates. Three proteins in **(Table 1)**, **UDP-galactopyranose mutase** (**UGM**: 4), **putative UDP-Glc 4'-epimerase** (**galE**: 5), **UDP-sugar pyrophosphorylase** (**USP**: 26) are constituents of biochemical pathways, either directly or indirectly responsible for the synthesis of β–Gal*f*. β – Gal*f* is synthesized from its precursor UDP-galactose (UDP–Gal) by the enzyme **UDP-galactopyranose mutase (UGM)**, inhibition of which is known to regulate parasitic virulence and hence is an attractive target [46]. The cellular pool of UDP-Gal is

contributed by the Isselbacher and Leloir pathways [45, 46]. In the Leloir pathway, UDP–Gal is synthesized from galactose **-** 1 **-** phosphate by **UDP - sugar pyrophosphorylase** (**USP**), whereas in the Isselbacher pathway galactose–1–phosphate is converted to UDP-Gal and glucose–1-phosphate by galactose–1–phosphate uridyltransferase. Within this pathway, the reversible and bidirectional enzyme **UDP-Glc 4'-epimerase (GalE)** can convert UDP – Gal to UDP – Glucose and vice versa. Despite low sequence identity of about 33 % between human and parasitic GalE , high resolution crystal structures of both proteins reveal a common overall topology and similar protein-ligand interactions at the active site [47]. GalE holds great promise as a drug target in *T. brucei*.

Next, a set of five proteins implicated in purine/pyrimidine metabolism occupied fairly prominent positions in **Table 1**: **putative deoxyuridine triphosphatase nucleotidohydrolase** (**dUTPase**: 11), **nonspecific nucleoside hydrolase** (**NNH**: 24), and **putative OMPDCase – OPRTase** (**OMPDC-OPRT**: 29). Unlike their human and other mammalian hosts leishmania lack the molecular machinery to synthesize purine nucleotides *de novo* and is thus dependent on a purine salvage pathway [48]. Extensive genetic studies on the purine salvage pathway show it to be highly complex with several redundant links. For example, mutant strains individually lacking one of the key enzymes of the pathway: adenine phosphoribosyl transferase (APRT), hypoxanthine – guanine phosphoribosyltransferase (HGPRT), adenosine kinase (AK) and xanthine phosphoribosyl transferase (XPRT) were all found to be viable [48]. However, the phenotypic characterization of the double Δ*hgprt/*Δ*xprt* mutant indicated that purine salvage from extracellular sources is primarily funneled through XPRT, HGPRT with AK and APRT being by and large superfluous [49]. Thus, the central role played by these two enzymes (HGPRT & XPRT) confers functional importance to downstream molecules which distributes

their products into adenylate and guanylate nucleotides. Adenylosuccinate synthetase (ADSS) and adenylosuccinate lyase (ASL) are two such enzymes which sequentially convert IMP to AMP, the former catalyzing the GTP dependent formation of adenylosuccinate from IMP and aspartic acid while the latter removes a fumarate from adenylosuccinate formed by ADSS. Knock out mutants of ASL shows highly attenuated infectivity of the parasites [49].

Null mutants of purine transporters LdNT1, LdNT2 (Δ*ldnt1*, Δ*ldnt2* and Δ*ldnt1* /Δ*ldnt2*) do not appear to interfere with parasitic growth based on natural purine sources (with the exception of xanthosine) [48]. Subsequently, the enzyme **nonspecific nucleoside hydrolase** (**NNH**) was identified to perform the non-specific conversion of purine nucleosides to nucleobases, which can then be transported by other transporters LdNT3 and LdNT4 [48]. In *Leishmania spp.* **NNH** hydrolyzes the N-glycosidic bond of both purine and pyrimidine nucleosides to yield ribose and other bases. Upon intake and conversion, adenine bases are irreversibly deaminated to hypoxanthine by the enzyme adenine aminohydrolase (AAH). However, despite its unique presence in the parasite (w.r.t mammals and humans), the enzyme was found to be non – essential as demonstrated by the viability of Δ*aah* knockouts [50].

Both humans and leishmania are capable of the de novo synthesis of pyrimidines though there exists considerable discrepancy in the organization of their corresponding enzymes into multifunctional polypeptides, cellular localization and allosteric regulators [51]. This is especially true of the last two enzymes in the pyrimidine synthesis pathway, orotate phosphoribosyltransferase (OPRT) and orotidine monophosphate decarboxylase (OMPDC), which are fused into one bifunctional protein, in both human and parasite. However the order of the polypeptide chains are reversed in both cases [51]. As **putative OMPDCase - OPRTase (OMPDC - OPRT)** is active in the final step of pyrimidine biosynthesis , its inhibition is

expected to be lethal for the parasite. Finally, **putative deoxyuridine triphosphatase nucleotidohydrolase (dUTPase)** hydrolyzes dUTP to dUMP and pyrophosphate leading to the maintenance of the dTTP:dUTP ratio in the cell ensuring precision in DNA replication [52].

Another unique feature of protozoa belonging generally to kinetoplastids is the compartmentalisation of the first seven glycolytic enzymes (and therefore glycolysis) into organelles called glycosomes, in contrast to other organisms where glycolysis generally occurs in the cytosol. This feature is essential for the regulation of glycolysis in the parasite and also to effectively switch over to anaerobic forms of respiration [53]. Three such glycolytic enzymes **2,3 - bisphosphoglycerate – independent phosphoglycerate mutase** (**PGM**: 19), **ATP – dependent phosphofructokinase** (**PFK**: 21) and **glycerol - 3 - phosphate dehydrogenase** (23), including two other enzymes either upstream or downstream of the glycolytic pathway, **putative NADP-dependent alcohol dehydrogenase** (17) and **glucokinase** (22) appeared in **Table 1**. Since glycolysis is the only known source for ATP in leishmania these enzymes offer attractive targets, especially **phosphoglycerate mutase (i-PGM)** which is distinct in terms of structure, catalytic mechanism and whose reduced expression was also found to be lethal for cultured *T. brucei*. However, despite intense effort on some of these validated targets effective pharmaceutical interventions have yet to emerge.

Traditionally, drugs inhibiting folate metabolism, specifically dihydrofolate reductase (DHFR) and thymidylate synthase (TS) have been successful as antibacterials. DHFR maintains the THF (N^5, N^{10}-methylene tetrahydrofolate) pool in the cell by the NADPH dependent reduction of dihydrofolate (DHF), which in turn is utilized by TS to catalyze the conversion of dUMP to dTMP. Lack of dTMP curtails DNA replication leading to cell death [54]. In leishmania both these enzymes are conjoined into a bifunctional enzyme DHFR-TS which is the

primary source of reduced folate and also the lone source of thymidylate in the parasite [55]. However, inhibition of this enzyme is ineffective in promoting lethality due to the presence of another short chain non – specific dehydrogenase/reductase **pteridine reductase 1** (**PTR 1**: 2), which acts both as a modulator and bypass for inhibitors targeting DHFR – TS. **PTR 1** is responsible for the essential salvage of unconjugated pterins (such as biopterins) as it catalyzes the NADPH dependent two step reduction of oxidized pterins to their active tetrahydro forms. Deletion mutants of **PTR-1** alone were non – viable and hypersensitive to the drug methotrexate (MTX) and had to be simultaneously inhibited [54], in case DHFR-TS was being targeted.

Among the first 30 prioritized targets a group of peptidases: **peptidase m20/m25/m40** (7, 9), **putative calpain – like cysteine peptidase** (14), **putative serine peptidase** (16a, 16b), **putative proteosome activator protein pa26** (15), **putative dipeptidylcarboxypeptidase** (25) and a **putative metacaspase protein** (12) were found in **Table 1**. A wide range of proteases spanning most of the major classes have been identified in the leishmanial genomes, with *L. braziliensis* alone having at least 44 cysteine, 23 serine and 97 metalloproteases [56]. Of these, cysteine proteases (CP) have been confirmed as virulence factors playing a major role in mediating host – parasite interactions; with parasites (*L.tropica*) treated with CP inhibitors exhibiting reduced viability, growth and pathogenicity. Metalloproteinases have also been known to be expressed on the surface of *Leishmania spp*., protecting the pathogen from the defensive action of host enzymes and the phagolysozome of macrophages [56]. In addition, a CP inhibitor; **inhibitor of cysteine peptidase** (6) also appeared at a prominent position in **Table 1** by virtue of its being a virulence factor and a probably protecting the parasite from the hydrolytic environment of the sandfly gut or the internal environment of host macrophages.

The rest consisted of a miscellaneous collection of enzymes such as **putative endoribonuclease L-PSP** (8), involved in mRNA salvage and protein synthesis [57]; **macrophage migration inhibitory factor like protein** (13a,13b), implicated in the evasion of innate host immunity by arresting the apoptosis of infected macrophages [57]; **5-methyl-tetrahydropteroyltriglutamate-homocysteine S-methyltransferase** (18), which plays important role in the synthesis of cysteine/methionine and also their interconversion [58], **putative glutamate dehydrogenase** (27), **3-mercaptopyruvate sulfurtransferase** (30) involved in amino acid metabolism and **ADF – cofilin**(20)for cellular growth and motility [57].

3.3 The Protein – Ligand Complexes

An exhaustive literature survey was conducted to identify inhibitors for the first thirty proteins from **Table 1**. Inhibitors with experimentally determined IC_{50} or K_i values were found in the literature for only 8 out of the 30 proteins. A total of 27 inhibitors (**Table 2**) were shortlisted for the above mentioned eight target proteins by selecting those ligands with the lowest IC_{50} values from a given family of compounds (that is a class of compounds with a conserved backbone/kernel and diverse peripheral substituents).

Thus a total of 32 protein-ligand complexes (27 from *L. major* and 5 from *T. brucei*, iTb4-8) were divided into three sets (**Table 3; Group A, B & C**):

(I) In the first set (Group **A**), the crystal structures of the ligand bound protein complexes were readily available in the protein data bank and were utilized directly for computing the structure based pharmacophores (**Table 3; inhibitors i1,i2 and i3**). Henceforth the inhibitors will be referred to by the number enumerated in **Table 2**.

(II) In Group **B**, the crystal structures of the ligand-protein complexes were available, with the protein either being the actual targeted molecule (from *L. major*) or a closely related homologue, with sequence identity exceeding 60 % with respect to the corresponding protein from *L. major*. In the latter cases the homologous protein present in the PDB was used as a template to model the parasitic protein. Likewise, the specific ligand used to form the complex could either be the original small molecule found in the PDB file; or the ligand coordinates from the crystal structure were used as a template to add peripheral chemical groups. Specific ligands were docked onto the corresponding parasitic proteins using the GOLD software. In addition, the original protein-ligand complexes present in the PDB (from other trypanosomes) were also included in the subsequent calculations (**Table 3; inhibitors i4 – i9, iTb4 – iTb8 & i10 – i24**). The ability of the docking protocol as implemented in the GOLD software to independently locate the ligand position as found in the crystal structure was verified for all the complexes used in this study.

(III) In Group **C** no information regarding both the proteins and their specific ligands were available in the PDB, though crystal structures exceeding a sequence identity of 50% with respect to the target were present in the database, which were used as templates to obtain the three dimensional models of the leishmanial proteins. The ligand coordinates were either obtained from the PubChem database (NCBI) or constructed *ab initio* (**Material & Methods**) by fourth atom fixing techniques. Blind docking (by GOLD) was used to position the inhibitor onto the putative active site (as obtained from the literature) of the enzyme (**Table 3**; **inhibitors i25, i26 & i27**).

Table 2: Drug Target Inhibitors. Inhibitors identified for eight of the proteins from Table 1 along with their respective IC_{50}'s or K_i values with respect to their corresponding target proteins. The inhibitors will be referred to by the numbers assigned in the first column.

No.	PROTEIN	Inhibitor name	IC_{50} /K_i
i1	Pteridine Reductase 1	methotrexate (nature structural biology, volume 8, number 6, june 2001)	1.1 µM
i2	Pteridine Reductase 1	Trimethoprim (Experimental Parasitology 87, 157-169 (1997))	12 µM
i3	Pteridine Reductase 1	10 – propargyl-5,8-dideazafolic acid (J. Mol. Biol. (2005) 352, 105–116)	> 10 µM
i4	Trypanothione Reductase	methyl [(4S) - 6 - bromo - 2 - methyl - 4 - phenylquinazolin - 3(4H)-YL] acetate (J. Med. Chem. 2011, 54, 6514–6530)	6.8 µM
i5	Trypanothione Reductase	(4s) - 3 - benzyl - 6 - chloro - 2 - methyl - 4 - phenyl - 3,4 – dihydroquinazoline (J. Med. Chem. 2011, 54, 6514–6530)	0.93 µM
i6	Trypanothione Reductase	n - {2 - [(4s) - 6 - chloro - 2 - methyl - 4 - phenylquinazolin - 3(4h) - yl] ethyl}furan - 2 – carboxamide (J. Med. Chem. 2011, 54, 6514–6530)	0.86 µM
i7	Trypanothione Reductase	(4s) - 6 - chloro - 3 - {2 - [4 - (furan - 2 - ylcarbonyl)piperazin - 1 - yl]ethyl} -2 - methyl - 4 - phenyl - 3,4 –dihydroquinazoline (J. Med. Chem. 2011, 54, 6514–6530)	0.42 µM
i8	Trypanothione Reductase	3 - [(4s) - 6 - chloro - 2 - methyl - 4 - (4 - methylphenyl) quinazolin - 3(4h) -yl] - n,n - dimethylpropan - 1 – amine (J. Med. Chem. 2011, 54, 6514–6530)	0.23 µM
i9	Trypanothione Reductase	a C6 - substituted and C8 - substituted 3,4 - dihydroquinazoline analogue (J. Med. Chem. 2011, 54, 6514–6530)	0.35 µM
i10	Pteridine Reductase 1	a quinazoline derivative (Experimental Parasitology 1997, 87, 157-169)	0.4 µM

i11	Pteridine Reductase 1	a 2,4-Diaminopyrimidine derivative (Experimental Parasitology 1997 ,87, 157-169)	0.4 µM
iTb4 to iTb8	Trypanothione Reductase crystal complex from T. brucei	Above inhibitor i1 to i5. (J. Med. Chem. 2011, 54, 6514–6530)	Same as above
i12	ATP dependent Phospho Fructo Kinase	a N, N0 - substituted - 1-amino - 2, 5 - anhydro - 1 - deoxy - 1 - D - mannonamide derivatives (Bioorg. Med. Chem. 2008, 16 , 5050–5061)	49 µM
i13	ATP dependent Phospho Fructo Kinase	2, 5 - anhydro - 1 - deoxy - 1 - (3, 4 - dichlorobenzylamino) - D – mannitol (Bioorg. Med. Chem. 2008, 16 , 5050–5061)	0.4 µM
i14	ATP dependent Phospho Fructo Kinase	2, 5 - Anhydro - 1 - deoxy - 1 - (3, 4 - dichlorobenzylamino) - D - 3, 4 - dichlorobenzylmannonamide (Bioorg. Med. Chem. 2008, 16 , 5050–5061)	23 µM
i15	Deoxyuridine Triphosphate Nucleotidohydrolase	a 5' - tert - butyldiphenylsilyloxy derivative (J. Med. Chem. 2005, 48, 5942-5954)	3.0 µM
i16	Deoxyuridine Triphosphate Nucleotidohydrolase	a 5' - Ph3CNH derivative (J. Med. Chem. 2005, 48, 5942-5954)	NA; substrate analog
i17	Deoxyuridine Triphosphate Nucleotidohydrolase	5' - tritylamino - 3' - fluoro - 2', 3', 5' – trideoxyuridine (J. Med. Chem. 2005, 48, 5942-5954)	3.6 µM
i18	Deoxyuridine Triphosphate Nucleotidohydrolase	5' - O - triphenylsilyl - 2', 3' - didehydro - 2', 3' – dideoxyuridine (J. Med. Chem. 2005, 48, 5942-5954)	12 µM
i19	Nonspecific Nucleoside Hydrolase	immucillin A (J. Biol. Chem. 1999,274(30, 21114-21120)	15 nM
i20	Nonspecific Nucleoside Hydrolase	immucilin ACAP (J. Biol. Chem.1999, 274(30), 21114-21120)	6.5 nM
i21	Nonspecific Nucleoside Hydrolase	N - (9 - deaza - adenin - 9 - yl) methyl - 1, 4 - dideoxy - 1, 4 - imino - D - ribitol (Antimicrobial Agents and Chemotherapy, 2010, 1900-1908)	0.49 µM

i22	Nonspecific Nucleoside Hydrolase	immucillin-H (1, 4 - dideoxy - 4 - aza - 1 - (s) - (9 - deazahypoxanthin - 9 - yl) - d - ribitol) (Biochem Biophys. Acta.2009, 1794, 953-960)	K_i = 6.2 nM
i23	Nonspecific Nucleoside Hydrolase	7 - (((2R, 3R, 4S) - 3, 4 - dihydroxy - 2 - (hydroxymethyl) pyrrolidin - 1 - yl) methyl) - 3H - pyrrolo [3, 2-d] pyrimidin - 4 (5H) - one (Biochem Biophys. Acta. 2009, 1794 , 953-960)	Ki = 4.4 nM
i24	Nonspecific Nucleoside Hydrolase	(2R, 3R, 4S) - 2 - (hydroxymethyl) - 1 - (quinolin - 8 - ylmethyl) pyrrolidine - 3, 4 – diol (Biochem Biophys. Acta. 2009, 1794, 953-960)	K_i = 10.8 µM
i25	Trypanothione Synthetase	1-[3-(3-fluorophenyl)indazol-1-yl]-3,3-dimethylbutan-2-one (J.Biol.Chem., 2009, 284(52), pp. 36137–36145)	0.095 µM
i26	Putative-UDP-glc-4'-epimerase	ebselen (Bioorg. & Med. Chem. Lett. 2006, 16 , 5744–5747)	0.62 µM
i27	GLO1	S-4-bromobenzylglutathionylspermidine (Molecular Microbiology,2006, 59, 1239–1248)	Ki = 0.54 µM

Table 3: Protein – inhibitor Groups. Rationale behind grouping the protein – inhibitor complexes into classes A,B,C.

Groups	Status considered	Proteins
A	Protein-inhibitor complex structure available Modelling not required for protein Inhibitor's position and coordinates known Docking not necessary	PTR I
B	Protein-inhibitor complexes not available Modelling required for protein from close homolog or structure available Inhibitor pose & coordinates taken from known homologous complex Docking performed with positioned ligand	TR (from *L. major* & *T. brucei*) ATP dependant PFK dUTPase NNH
C	Protein-inhibitor complex not available Modelling required for protein from close homolog Inhibitors are drawn ab initio and energy minimized Blind docking performed based on the putative active site found in the literature	TS GalE GLO1

Thus **32 s**tructure based pharmacophores were computed from their corresponding protein - ligand complexes, of which 3 complexes belonged to Group A, 21to Group B and 3 to Group C. In addition, 5 complexes (for inhibitor no. iTb4 to iTb8) whose proteins belonged to other trypanosomatids were also included in the calculation, by virtue of their being templates for the leishmanial proteins in Group B.

Complexes in Group A (**Table 3**) consists of **pteridine reductase 1** (**PTR 1)** bound to inhibitors methotrexate (1E7W; inhibitor no. i1), trimethoprim (2BFM; inhibitor no. i2) and 10 – propargyl-5, 8-dideazafolic acid (2BFA; inhibitor no. i3). Crystal structures of these three complexes include the cofactor NADPH. **PTR 1** is a homotetramer, with individual subunits displaying the double Rossmann Fold composed of a central 7 - stranded parallel β-sheet with three α-helices on either side (**Figure 2**) [54]. All three structures exhibit high structural conservation, with the active site being an elongated L-shaped cleft constituted by the C terminal

section of the strands β_1 - β_6, parts of the helices α_1, α_4 and a loop interconnecting a strand (β_6) and a helix (α_6) [54]. Two more complexes involving **PTR1** with a quinazoline derivative (inhibitor no. i10) and a 2, 4-diaminopyrimidine derivative (inhibitor no. i11) were included in Group B. The methotrexate structure from the PDB file 1E7W was used as a template to build the quinazoline derivative (inhibitor no. i10) and the best solution with a CHEMPLP score of 49.44 was finally selected out of several GOLD runs (See **Materials & Methods**).

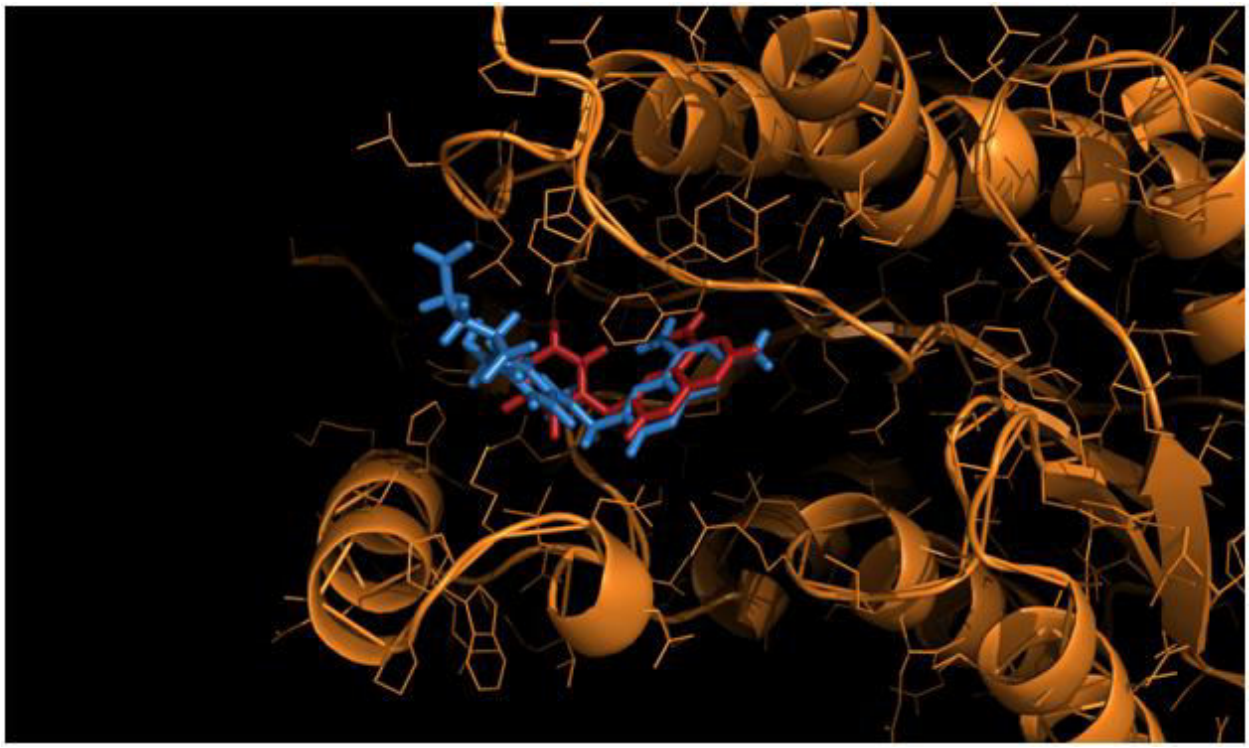


Figure 2: **Active site of Pteridine Reductase 1 (PTR1) complexed with methotrexate (i1 - blue) as found in the crystal structure IE7W and the docked quinazoline derivative (i10 – red).**

The rmsd between the atoms common to methotrexate (pteridine ring or its derivative) as located in the **PTR1** active site and the docked inhibitor-i10 was 0.77Å. Protein ligand contacts involving residues Ser 111, Phe 113, Asp 181, Leu 188, Tyr 194, Leu 226 , Leu 229, Asp 232

and Met 233 were common both for methotrexate and Inhibitor i10 (**Supplementary file 4,** Supplementary files are provided separately with the CD-ROM). Notably, contacts between the pteridine ring and Phe 113, Ser 111 were prominent in both cases. Additional contacts were observed in methotrexate with respect to the inhibitor due to the more elongated character of the molecule, extending from the pteridine ring (**Supplementary file 4; Figure 2**). A similar procedure adopted for inhibitor-i11 based on trimethoprim (PDB code 2BFM) as a template gave a corresponding CHEMPLP score of 79.80. Although the orientations for both ligands were fairly similar, a translational shift in the position of inhibitor i11 was due to the substitution of -CH_2Ph in place of -H in the pyrimidine ring of trimethoprim. The common residual contacts for both ligands were Phe 113, Asp 181, Leu 188, Tyr 194, Leu 226, Leu 229, and His 241 (**Supplementary file 4**). All the four ligands also maintained atomic contacts with NADPH.

Other protein targets in Group B apart from **PTR – 1** were **TR**, **PFK, dUTPase** and **NNH**. Five complexes of **TR** from *T. brucei* were included (inhibitor numbers iTb4 – iTb8; See **Table 2**) directly from the crystal structures with PDB codes 2WP5, 2WP6, 2WPE, 2WPC and 2WPF [59]. **TR** from *T. brucei* has a sequence identity of 66.5 % with respect to the corresponding protein from *L. major* and was used as a template to model the parasitic protein. **TR** (*T. brucei*) is a homodimer with each subunit consisting of three domains, the inhibitor binding cleft being formed by a congregation of α helices in domain I [60]. The binding site exhibits conformational flexibity indicating an induced fit of the ligand to the binding pocket. Thus for each ligand (i4 – i8) the leishmanial protein was repeatedly modeled from its original complex (2WP5, 2WP6 etc as the case maybe). The ligands were initially placed in the active site of the modeled proteins based on the rotation matrices and translation vectors obtained upon superposing the Cα coordinates (Dali server; http://ekhidna.biocenter.helsinki.fi/dali_lite/start) of

the template onto its associated model. Subsequent GOLD runs gave high CHEMPLP scores greater than 60.0 for all the five complexes and rmsd's ranging from 0.2 – 0.8Å between the final docked structure and the initial position of the ligand in the modeled protein (**Figure 3**).

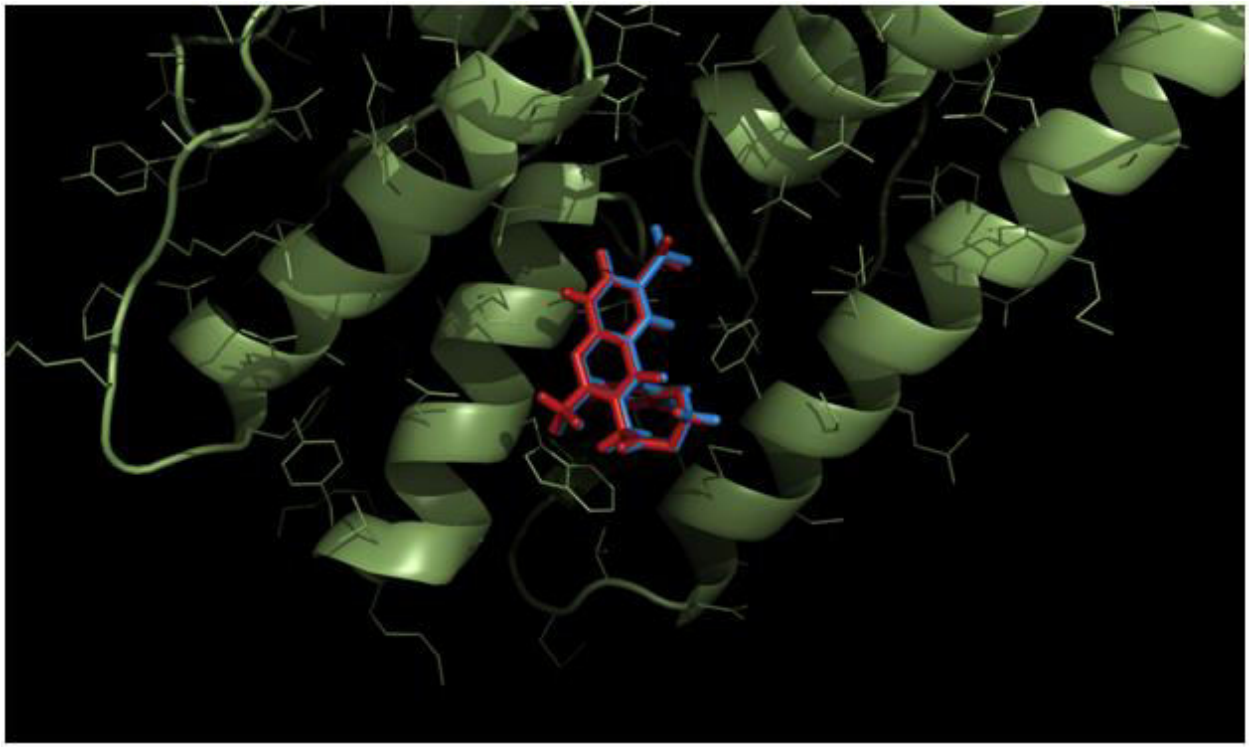


Figure 3: Active site of the modeled Trypanothione Reductase (TR) from *L. major* complexed with ligand i4 (red) along with iTb4 (blue) placed utilizing the superposition matrices and vectors obtained from superposing the protein model onto its template (2WP5) .

The active sites for all the five ligands (from 2WP5 – 2WPF) were completely conserved in both *L . major* and *T. brucei* , with residues Leu 17, Glu 18, Trp 21, Tyr 110, Met 113, Phe 114 making atomic contacts with all the ligands (with the exception of inhibitor i9 in *L. major*), in both the enzymes (**Supplementary file 4,** Supplementary files are provided separately with the CD-ROM). In addition, Ser 14, Leu 17, Gly 49, Leu 120 and Ile 339 were found in the vicinity

of the ligands, occasionally in some of the complexes. i9 (**Table 2**) was obtained from (4s)-3-benzyl-6-chloro-2-methyl-4-phenyl-3,4-dihydroquinazoline (inhibitor i8:2WP6) and the final docked position (allowing for flexibility in residues Glu 18, Trp 21 and Met 113 in the active site) had a rmsd of 0.66 Å (with respect to common atoms of 3,4-dihydro-quinazoline analogues) and a score of 34.69. The same set of core residues (Leu 17, Glu 18 etc.) including Ser 14 and Leu 120 also formed the active site for inhibitor i9 (**Supplementary file 4,** Supplementary files are provided separately with the CD-ROM).

The active site for fructose-6-phosphate (F6P) of the trypanosomatid ATP dependent **PFK** from *T. brucei* exhibits significant structural differences compared to its human counterpart and is located at the interface of two subunits in the homotetramer. Each subunit is composed of three domains (A, B, C) with the ATP binding site (between domains B and C) lying adjacent to the F6P site [61]. The complete tetramer of **PFK** from *L. major* was modeled based on the homologous protein from *T. brucei* (3F5M) with respect to which it shares a 71 % sequence identity and ATP (along with the Mg^{2+}) were docked/placed in the model, prior to placement of the substrate. The crystal structure of PFK from *T. brucei* in 3F5M is in complex with ATP and does not contain the substrate F6P, whose coordinates were extracted from 1MTO which consists of PFK from *B. stearothermophilus* in complex with F6P [61] and docked onto the corresponding active site in leishmanial PFK, subsequent to initial positioning of the molecule, following similar methods given above. Beginning with the coordinates of the furan ring of F6P, three other inhibitors were built by making appropriate substitutions : a N,N0-substituted-1-amino-2, 5-anhydro-1-deoxy-1- D-mannonamide derivative (inhibitor no. i12); 2,5-anhydro-1-deoxy-1-(3,4-dichlorobenzylamino) -D-mannitol (inhibitor no. i13) and 2,5-Anhydro-1-deoxy-1-(3,4-dichlorobenzylamino)-D-3,4-dichlorobenzylmannonamide (i14). Subsequent docking with

GOLD on a rigid active site gave CHEMPLP scores 70.35, 67.80 and 46.81 for the inhibitors i12, i13 and i14 respectively. Repeated attempts to dock the ligands on a flexible site did not yield physically meaningful results. The only common feature between the inhibitors – i12, i13 and i14 was the furan ring (from F6P) and considerable variability in the remaining features tended to shift the ligands from their initial position depending on the length and chemical character of the substituents on either side of the furan ring. Consequently with a few exceptions, the constellation of residues constituting their binding pockets were significantly different (**Supplementary file 4,** Supplementary files are provided separately with the CD-ROM).

The crystal structure of **dUTPase** from *L. major*, is a dimeric all α protein (in contrast to its trimeric all β human homologue) was obtained from the PDB (2CJE) [62]. The active site is located in the vicinity of the interface between the rigid and mobile domains which constitute each subunit [62]. In addition, the site on one subunit is also constituted by a loop contributed by the other monomer. Crystal structures of the closed and open forms of the enzyme from *T. cruzi* revealed a significant movement of the mobile domain and rearrangement of the secondary structural elements [62]. The closed ligand bound form of **dUTPase** from 2CJE was used to model complexes with three other inhibitors. Since the enzyme sits in a special position in the crystal structure, the entire dimer of **dUTPase** was generated prior to the placement of the other ligands. The bound substrate analogue DUPNHP (2'-deoxyuridine 5'-alpha,beta-imido-diphosphate from 2CJE) was used to design the three inhibitors : a 5'-tert-butyldiphenylsilyloxy derivative (i15), a 2'-deoxyuridine 5'-alpha,beta-imido-diphosphate (i16), 5'-tritylamino-3'-fluoro-2',3',5'-trideoxyuridine (i17) and 5'-O-triphenylsilyl-2',3'-didehydro-2',3'-dideoxyuridine (i18). CHEMPLP scores for all the four inhibitors ranged from 90-100 and the rmsd's between

the starting and final docked position ranged from 0.5 – 1.2 Å. Based on the visual examination of the initial ligand position in the active site of the enzyme and survey of the ligand – protein atomic contacts, selected residues were rendered flexible in the docking process which could provide steric hindrance to the optimal orientation of the ligand or adopt alternative conformations in the binding pocket (inhibitor i15 - flexible residues Asn25, Glu48, Glu51, Glu76 , Tyr191 ; inhibitor- i17 : Asn25, Glu51 , Tyr 191; inhibitor- i18 : Glu48, Glu51, Glu76 and Tyr191).

For **NNH** (Nonspecific Nucleoside Hydrolase) six inhibitors (inhibitor no. i19 to i24) were chosen for docking. Complexes of three of these inhibitors (i22, i23, i24) with a homologous protein from *T. vivax* were available in the PDB (2FF2, 3EPW and 3EPX respectively), whereas the uncomplexed individual structure of **NNH** from *L. major* was found in 1EZR. The α|β enzyme from *L. major* is a homotetramer with an indispensable calcium ion in its active site [63]. Coordinates for the inhibitors i19, i20 were built starting from the pyrimidine group in the structure of immucilin – H present in 2FF2 and inhibitor i21from the ligand (i23) present in 3EPW. Docking of these inhibitors in the active site of the protein (including the Ca^{2+} ion), exhibited CHEMPLP scores and rmsd's (with respect to their original placement) ranging from 50 – 60 and 1.2 – 1.6, respectively. Flexibility was allowed for residues Phe167 and His240 in the enzyme active site during the docking process for all ligands associated with this protein. Interactions with residues Asp 15, Asp 14 (with the exception Inhibitor – i22) , Thr 126, Met 152 (except Inhibitor – i20) Asn 160, Glu 166, Phe 167, Asn 168, His 240 , Asp 241 and the calcium ion were common to all the ligands. Contacts with Leu 191 were found only for Inhibitors – i19 and i22 (**Supplementary file 4,** Supplementary files are provided separately with the CD-ROM).

Group C

Due to the lack of available prototypes or templates in terms of actual crystal structures depicting the position of the ligands in their binding sites, the confidence level associated with the docked complexes in Group C is necessarily low and thus the discussion of these complexes will be fairly brief. The crystal structure trypanothione sythetase (**TS**) from *L. major* (2VOB) has three putative binding sites for ATP, spermidine and glutathione (GSH). Inhibitor-7 (1-[3-(3-fluorophenyl) indazol-1-yl]-3, 3-dimethylbutan-2-one) binds with uncompetitive inhibition for both the putative ATP and GSH binding sites whereas exhibits competitive inhibition for the site associated with spermidine. Thus, the inhibitor (i25) was placed at the centroid of this site constituted by residues Arg 613, Arg 328, Ser 351, Glu 355, Phe 249 and Glu 407. As mentioned previously, inhibitor – i25 was constructed by *ab initio* fourth atom fixing techniques coupled to energy minimization. Several iterations with flexible ligand and rigid side chains of the active site led to a final CHEMPLP score of 47.32. Introduction of side chain flexibility did not appear to improve final docking poses. Likewise, inhibitor – i26 (ebselen) was positioned in the putative UDP binding pocket of UDP-glc-4'-epimerase (**GalE**) of *L. major* based on the centroid of residues R335, R268, N202 and H221. **GalE** from *L. major* was modeled based on the homologous enzyme from *T. brucei (*~ 58% sequence identity : 1GY8). Coordinates of ebselen were built by the same methods mentioned above and the final docked position had a CHEMPLP score of 36.96. The crystal structure of glyoxalase-I (**GLO I**) from *L. major* was obtained from 2C21 [44] and S-4-bromobenzylglutathionylspermidine (inhibitor-i27) was docked into the putative active site of the enzyme constituted by residues : A chain - His8, Arg12, Arg33, Trp35, Val37, Glu52, Glu59, Asn63 and B chain - His77, Asp100, Tyr101, Phe107, Met108, Tyr118, Glu120, Met127 and Lys130. The CHEMPLP docking score of 79.43

was obtained for this docking. For all docking runs in case of i27 both the energy minimized ligand and the residues composing the protein active site were held rigid, as additional trials with either flexible ligands and/or side chains led to significant shifts in their position away from the putative binding sites. As mentioned previously the confidence level is relatively low for these complexes.

4. Pharmacophore Design and Screening of Zinc Database

34 structure based pharmacophores were derived from their corresponding ligand bound three dimensional structures using LigandScout version 3.12 (build 20130912). Those pharmacophores whose area under the ROC curve (See **Materials & Methods**), were less than 0.75 whilst validation, were filtered out (i8, i11, i17 and i20). In addition, pharmacophores with either too few (i2, i20, i26: 3 features) or too many features (i14:13 features, i16:16, i27:12) were removed, leaving a total of 23 pharmacophores for subsequent calculations (**Table 4**). These pharmacophores were used to search the ZINC database using ZINCPharmer (html search engine) with parameters: 'Max Hits per Conf' =1, 'Max Hits per Mol' = 1, 'Max Total Hits' =20 and 'Max RMSD' = 0.5, 0.75, 1. The Max RMSD was gradually increased only if no hits were recorded in the initial cutoffs. The topmost hit of every pharmacophore with the least RMSD, along with hits which were similar to approved drugs (generally greater than 90%) are shown in **Table 4** and all the hits are given in **Supplementary file 5 (**Supplementary files are provided separately with the CD-ROM). A total number of 344 hits were recorded from the ZINC database which were then used to search the Drug Bank (http://www.drugbank.ca) with a cutoff in similarity score set to 70 %, so as to identify similar molecules actually in use as pharmaceutical products. From the 344 compounds distributed over 23 pharmacophores, 9 exhibited similarities to drugs under investigation, 319 showed similarities to experimental drugs

(*known* to bind to specific proteins in mammals, bacteria, viruses, fungi, or parasites) and **16** were similar to approved drugs (in at least one country). Of these 344 hits, 40 were from complexes in Group A, 304 from Group B and none from Group C.

Table 4: Results from searching the ZINC Database using structure based pharmacophores. The topmost hits from the ZINC database utilizing the pharmacophores from the protein – inhibitor complexes. For each pharmacophore the area under the ROC curve and the number of omitted pharmacophoric features while validation have been given in columns 3,4. Also included are hits (given in bold) which have features very similar to approved drugs. The ZINC Id of the hit is given in the last column.

Inhibitor no.	No. of pharmaco -phoric features	No. of Omitted features	Area Under the ROC Curve (AUC)	SMILES of topmost hit	ZINCID
i1	9	2	0.97	C/C2=N/C=1/N=C(/N)N=C(N)C=1/N=C2/C=3C=CC([Cl])=C([Cl])C=3	ZINC34515729
				C#CC[C@@](CC=2C=NC1=NC(N)=NC(N)=C1N=2)(C=3C=CC(=CC=3)C([O-])=O)C([O-])=O	**ZINC22012802**
				C#CC[C@](CC=2C=NC1=NC(N)=NC(N)=C1N=2)(C=3C=CC(=CC=3)C([O-])=O)C([O-])=O	**ZINC22012807**
				CC=1C=CC=C(C=1)C2=NC=3C(/N=C2/N)=NC(N)=NC=3N	**ZINC01566881**
				C#CC[C@]([H])(CC=2C=NC1=NC(N)=NC(N)=C1N=2)C=3C=CC(=CC=3)C([O-])=O	**ZINC22012811**
				C#CC[C@@]([H])(CC=2C=NC1=NC(N)=NC(N)=C1N=2)C=3C=CC(=CC=3)C([O-])=O	**ZINC22012815**
i3	8	3	0.88	CC=1C=CC=CC=1NC(=O)C=3[S]C=2/N=C(/N)C(C#N)=CC=2C=3N	ZINC18240380
				C#CC[C@@]([H])(CC=2C=NC1=NC(N)=NC(N)=C1N=2)C=3C=CC(=CC=3)C([O-])=O	**ZINC22012815**
i4	6	1	0.97	CN(C)C=[N+]C(=[S])NC=1C=CC([F])=CC=1	ZINC03028809
i5	7	2	0.97	CC=1/C=C(/C)N=C(N=1)N/C(=N\C(=[S])NC=2/C=C(/OC)C([Cl])=CC=2OC)N3CCC[C@@](C)([H])C3	ZINC14160212
i6	6	1	0.97	CN(C)C=1[N+]=CC=CC=1CNC(N[C@@]2([H])CC[C@@]([H])(C2)[S]C)=[N+]C	ZINC72776752
i7	5	1	0.9	COC=1C=CC(=CC=1NC(=[S])NC=2C=CC([F])=CC=2)[N+]([O-])=O	ZINC00493353

				C/C1=C/C=NC=2C1=CC(=CC=2N[C@@](C)([H])CCC[N+]CCC)OC	**ZINC01600860**
i9	6	1	0.97	CN(C)C=[N+]C(=[S])NC=1C=CC([F])=CC=1	ZINC03028809
iTb4	6	1	0.97	CN(C)C=[N+]C(=[S])NC=1C=CC([F])=CC=1	ZINC03028809
iTb5	7	2	0.99	CC=1/C=C(/C)N=C(N=1)N/C(=N\C(=[S])NC=2/C=C(/OC)C([Cl])=CC=2OC)N3CCC[C@@](C)([H])C3	ZINC14160212
iTb6	7	2	0.98	CC=1/C=C(/C)N=C(N=1)N/C(=N/C(=[S])NC=2C=CC([F])=CC=2)NCC=3C=CC([F])=CC=3	ZINC14156881
iTb7	5	1	0.9	C/C1=C(/C(=NN1C[C@@]2([H])CCC[N+]2C)C([F])([F])[F])C(=O)NCCC[N+](C)C	ZINC49362845
				C/C1=C/C=NC=2C1=CC(=CC=2N[C@@](C)([H])CCC[N+]CCC)OC	**ZINC01600860**
iTb8	8	3	0.99	CC=1/C=C(/C)N=C(N=1)N/C(=N\C(=[S])NC=2/C=C(/OC)C([Cl])=CC=2OC)N3CCC[C@](C)([H])C3	ZINC14160209
i10	7	2	0.91	C/C2=C/C1=N/N=C(/[S]CCCCCO)N1C(C)=N2	ZINC72058109
i12	7	3	0.97	CC(=O)NC=4C=1/C=C(/[F])C=CC=1N3C[C@](C)(C(=O)N[C@@]2([H])CCCC2)N(CC[N+](CC)CC)C(=O)C3=4	ZINC21866480
i13	4	1	0.99	CC=1C=CC(=CC=1)C[N+]2CCC([H])(CC2)CC(N)=O	ZINC40540751
i15	5	2	0.78	C[C@@]1([H])CN(C[C@@](C)([H])O1)[C@]([H])(C(=O)NC=2C=CC(=CC=2)NC(=O)COC)C(C)(C)[H]	ZINC58203407
i18	4	1	0.76	C[C@@]1([H])C[C@]([H])(CC)N(C1)C(=O)CC[C@@]2([H])NC(=O)NC2=O	ZINC73336547
				CC=1C=C/C=C(/C)C=1NC(=O)CNC=2C=CC=C(C=2)NC(C)=O	**ZINC29396021**
i19	6	2	0.93	C[C@@]34CC[C@@]1([H])[C@@]([H])(CC[C@@]2(O)C[C@](O)([H])CC[C@]12/C=[N+]/[C@@]([H])(CC)CO)[C@@]3(O)CC[C@]4([H])C5=CC(=O)OC5	ZINC09167567
				C[C@]3([H])O[C@]([H])(O[C@@]1([H])[C@@](O)([H])[C@@](O)([H])[C@]([H])(O[C@]1([H])CO)O[C@@]2([H])[C@@](O)([H])[C@@](O)([H])[C@@](O)([H])O[C@]2([H])CO)[C@](O)([H])[C@](O)([H])[C@]3([H])[N+](C)[C@]4([H])C=C(CO)[C@@](O)([H])[C@@](O)([H])[C@@]4(O)[H]	**ZINC77302460**
i21	6	1	0.88	O[C@@]1([H])CO[C@@]([H])([C@]1(O)[H])[C@@](O)([H])CO	ZINC05157080
				C[C@](O)([H])[C@](O)([H])[C@@](O)([H])[C@@](O)([H])CO	**ZINC03872643**
				CC(=O)C(=O)[C@@](O)([H])[C@](O)([H])[C@@](O)([H])[C@@](O)([H])CO	**ZINC64219378**
				CN/C2=N/C1=C(/N=C(/N)NC1=O)N2[C@]3([H])O[C@]([H])(CO)[C@](O)([H])[C@]3(O)[H]	**ZINC13361972**
i22	6	2	0.75	OCC(CO)(CO)[N+]C[C@@](O)([H])CN2C1=C/C=C(/[Br])C=C1C=3/C=C(/[Br])C=CC2=3	ZINC10384387

				[N+][C@]2([H])C[C@@]([N+])([H])[C@]([H])(O[C@@]1([H])O[C@]([H])(C[N+])[C@](O)([H])[C@@](O)([H])[C@]1(O)[H])[C@@](O)([H])[C@@]2([H])O[C@@]3([H])O[C@](H])(CO)[C@](O)([H])[C@@]([N+])([H])[C@]3(O)[H]	**ZINC08214767**
				C[C@]3(O)CO[C@@]([H])(O[C@]2([H])[C@@]([N+])([H])C[C@]([N+])([H])[C@]([H])(O[C@@]1([H])CC(=CC[C@]1([N+])[H])C[N+])[C@]2(O)[H])[C@@](O)([H])[C@@]3([H])[N+]C	**ZINC70672630**
				[N+][C@@]2([H])[C@](O)([H])[C@]([H])(O[C@]1([H])O[C@@]([H])(CO)[C@](O)([H])[C@](O)([H])[C@]1([N+])[H])[C@@]([H])(CO)O[C@]2(O)[H]	**ZINC43758958**
i23	7	2	0.87	OC=1C=C/C(=C(/O)C=1)C2=NN/C(=C2/C4=N/C=3C=CC=CC=3[S]4)C([F])([F])[F]	ZINC04126006
				[O-]C(=O)C[C@](O)([H])C[C@](O)([H])C=CC=2C(=C1C=CC=CC1=NC=2[C@]3([H])CC3)C=4C=CC([F])=CC=4	**ZINC11616582**
i24	7	2	0.87	CC=1C(O)=CC=C(C=1O)C2=NN=C[C@@]2([H])C4=NC=3C=CC=CC=3[S]4	ZINC18188334
				CN2C=1/N=C(/NCCO)N(CCO)C=1C(=O)NC2=O	**ZINC01876281**
i25	6	2	0.8	**CC=1C=CC(=CC=1)CN2C(=O)C(=CNC2=O)CC([O-])=O**	**ZINC20156415**

The structure based pharmacophore derived from methotrexate (i1) bound to pteridine reductase returned 20 small molecule compounds (**Supplementary file 5,** Supplementary files are provided separately with the CD-ROM) from the ZINC database, with the pteridine ring being the principal pharmacophoric feature. Most of these compounds from ZINC were variable chemical substitutions around the pteridine ring. Inhibitor i2, i10 and i11 (all complexed with pteridine reductase) failed to give any hit whereas the pharmacophore corresponding to i3 – pteridine reductase, again returned 20 compounds. Two approved drugs pralatrexate and triamterene were found to be similar to hits from pharmacophores involving i1 and i3 (**Table 5**). Pharmacophores from inhibitors i4,i5,i6,i7,i8,i9 complexed with trypanothione reductase found 20,4,20,20,20 compounds from the ZINC database respectively. For most of these compounds the phenyl ring and the terminal carboxyl (for example in i4 - methyl [(4S) - 6 - bromo - 2 - methyl - 4 - phenylquinazolin - 3(4H)-YL] acetate) appeared to be crucial

pharmacophoric features. The approved drug primaquine was found to be similar to the compound ZINC01600860 corresponding to i7. Inhibitors i12 (2 hits), i13 (20 hits) and i14(0 hit) complexed with parasitic phosphofructokinase failed to find any approved drug from Drugbank, whereas lidocaine and tocainide were found to be similar to ZINC29396021 (i18).

Table 5 : Approved drugs similar to hits in the ZINC database. Approved drugs with similarity score greater than 0.70 to specific ZINC hits. Information of the protein – inhibitor complex corresponding to the pharmacophore is also given.

Protein	Inhibitor	ZINC Code	Approved Drug (Drug Bank Code)	Similarity Score
Pteridine Reductase 1	i1	ZINC34515729	Triamterene (DB00384)	0.78
	i1	ZINC01566881	"	0.96
	i1	ZINC22012802	Pralatrexate (DB06813)	0.73
	i1	ZINC22012807	"	0.73
	i1	ZINC22012811	"	0.75
	i1,i3	ZINC22012815	"	0.75
Trypanothione Reductase	i7	ZINC01600860	Primaquine (DB01087)	0.84
Deoxyuridine Triphosphatase	i18	ZINC29396021	Lidocaine (DB00281)	0.80
		ZINC29396021	Tocainide (DB01056)	0.72

Nucleotido hydrolase				
Nonspecific	i19	ZINC77302460	Acarbose (DB00284)	0.98
Nucleoside	i21	ZINC03872643	Mannitol (DB00742)	1.0
Hydrolase	i21	ZINC64219378	CalciumGluceptate (DB00326)	0.70
	i21	ZINC13361972	Nelarabine (DB01280)	0.76
	i21	ZINC13361972	Didanosine (DB00900)	0.76
	i21	ZINC13361972	Vidarabine (DB00194)	0.76
	i22	ZINC08214767	Kanamycin (DB01172)	1.0
	i22	ZINC43758958	“	0.85
	i22	ZINC08214767	Tobramycin (DB00684)	0.96
	i22	ZINC43758958	“	0.88
	i22	ZINC08214767	Neomycin (Db00094)	0.94
	i22	ZINC43758958	“	0.88
	i22	ZINC08214767	Framycetin (DB00452)	0.94
	i22	ZINC43758958	“	0.88
	i22	ZINC08214767	Paromomycin (DB01421)	0.94
	i22	ZINC43758958	“	0.88
	i22	ZINC08214767	Gentamicin (DB00798)	0.76
	i22	ZINC70672630	“	0.77
	i22	ZINC08214767	Glucosamine(DN01296)	0.74
	i22	ZINC43758958	“	0.88
	i22	ZINC70672630	Netilmicin (DB00955)	0.71

	i23 i24	ZINC11616582 ZINC01876281	Pitavastatin (DB08860) Dyphylline (DB00651)	1.0 0.71

For inhibitors i15 (6 hits), i16 (0),i17(0) and i18(20) complexed with **deoxy uridine triphosphatase nucleotide hydrolase,** the principal pharamacophoric feature(s) responsible for the hits appeared to be the uridyl moiety coupled to the pentose sugar ring. Especially, fruitful were pharmacophores due to complexes with **nonspecific nucleoside hydrolase** as they yielded acarbose (i19 – 5 hits); mannitol, calcium gluceptate, nelarbine, didanosine, vidarbine (i21 – 20 hits); kanamycin, tobramycin, neomycin, framycetin, paromomycin, gentamicin, glucosamine, netilimicin (i22 – 20 hits); pitavastatin (i23 – 20 hits) and diphylline (i24 – 20 hits). In this case the essential pharmacophoric recognition features were the pyrimidine ring coupled to a pentose sugar. Of the remaining pharmacophores from i25 (20 hits), i26(0),i27(0) complexed with **trypanothione synthetase** no drug could be recovered from DrugBank. The information with regard to the list of approved drugs has been summarized in **Table 5**.

Interestingly, the search for approved drugs similar to ZINC compounds led to paromomycin (with a similarity score of 0.944 with respect to immucilin H in complex with **nonspecific nucleoside hydrolase**, i22)spontaneously appearing in the list, a drug having passed all clinical trials and is now currently being prescribed for visceral leishmaniasis. Paromomycin has also been successfully used in topical creams for the treatment ulcerative cutaneous leishmaniasis [64]. The inclusion of paromomycin provides some confidence that some of the

listed drugs (in **Table 5**) could possibly exhibit some measure of antileishmanial activity. Likewise framycetin, neomycin, gentamicin, netilimicin and tobramycin all belong to the same class of aminoglycoside antibiotics generally known to inhibit protein synthesis. Framycetin and neomycin have found extensive use in topical ointments and creams. Didanosine and vidarbine are antiviral drugs the former being a nucleoside analogue of guanosine with hypoxanthine attached to the sugar ring and the latter an analogue of adenosine, in this case D – arabinose replacing D-ribose. Nelarabine on the other hand is a purine nucleoside analogue currently being applied in the chemotherapy of T-cell acute lymphoblastic leukemia. Other drugs include lidocaine (and its analog tocainide) an amino amide type local anesthetic , primaquine a member of the 8 – aminoquinoline group of drugs used in the treatment of malaria/ pneumocystis pneumonia, pralatrexate an anti-folate for anti-cancer therapy and triamterene a diuretic drug for hypertension. Notably, **pteridine reductase**, **trypanothione reductase**, **deoxyuridine triphosphatase** have been found to be essential for survival and **nonspecific nucleoside hydrolase** plays a central role in the purine salvage pathway. Currently, our aim is to experimentally test the anti – leishmanial character of these compounds/approved drugs.

5. Conclusions

The work reported in this paper demonstrates the series of computational steps beginning with the comparison of genomes, prioritization of prospective drug targets, culling or assembly of inhibitor – target complexes through template based model building and docking, generation of pharmacophores and their subsequent use for searching small molecule databases (such ZINC / Drug Bank), to rationally assemble a set of lead compounds for experimentally testing as potential antileishmanials. The natural appearance of paromomycin, a drug currently being employed against visceral leishmaniais, in the list of lead compounds lends some confidence to

the adoption of such scaffold – hopping techniques in order to generate a library of prospective antileishmanials.

References:

1. Alvar J, Velez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, den Boer M: **Leishmaniasis worldwide and global estimates of its incidence**. *PLoS One* 2012, **7**(5):e35671.
2. Killick-Kendrick R: **The life-cycle of Leishmania in the sandfly with special reference to the form infective to the vertebrate host**. *Ann Parasitol Hum Comp* 1990, **65 Suppl 1**:37-42.
3. Ashutosh, Sundar S, Goyal N: **Molecular mechanisms of antimony resistance in Leishmania**. *J Med Microbiol* 2007, **56**(Pt 2):143-153.
4. Jeddi F, Piarroux R, Mary C: **Antimony resistance in leishmania, focusing on experimental research**. *J Trop Med* 2011, **2011**:695382.
5. Rojas R, Valderrama L, Valderrama M, Varona MX, Ouellette M, Saravia NG: **Resistance to antimony and treatment failure in human Leishmania (Viannia) infection**. *J Infect Dis* 2006, **193**(10):1375-1383.
6. Ouellette M, Drummelsmith J, Papadopoulou B: **Leishmaniasis: drugs in the clinic, resistance and new developments**. *Drug Resist Updat* 2004, **7**(4-5):257-266.
7. Guerin PJ, Olliaro P, Sundar S, Boelaert M, Croft SL, Desjeux P, Wasunna MK, Bryceson AD: **Visceral leishmaniasis: current status of control, diagnosis, and treatment, and a proposed research and development agenda**. *Lancet Infect Dis* 2002, **2**(8):494-501.
8. Hopkins AL, Groom CR: **The druggable genome**. *Nat Rev Drug Discov* 2002, **1**(9):727-730.

9. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B *et al*: **The genome of the African trypanosome Trypanosoma brucei**. *Science* 2005, **309**(5733):416-422.

10. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G *et al*: **The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease**. *Science* 2005, **309**(5733):409-415.

11. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C *et al*: **Comparative genomics of trypanosomatid parasitic protozoa**. *Science* 2005, **309**(5733):404-409.

12. Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, Peters N, Adlem E, Tivey A, Aslett M *et al*: **Comparative genomic analysis of three Leishmania species that cause diverse human disease**. *Nat Genet* 2007, **39**(7):839-847.

13. Ghedin E, Bringaud F, Peterson J, Myler P, Berriman M, Ivens A, Andersson B, Bontempi E, Eisen J, Angiuoli S *et al*: **Gene synteny and evolution of genome architecture in trypanosomatids**. *Mol Biochem Parasitol* 2004, **134**(2):183-191.

14. Fairlamb AH: **Novel biochemical pathways in parasitic protozoa**. *Parasitology* 1989, **99 Suppl**:S93-112.

15. Croft SL, Coombs GH: **Leishmaniasis--current chemotherapy and recent advances in the search for novel drugs**. *Trends Parasitol* 2003, **19**(11):502-508.

16. Aguero F, Al-Lazikani B, Aslett M, Berriman M, Buckner FS, Campbell RK, Carmona S, Carruthers IM, Chan AW, Chen F *et al*: **Genomic-scale prioritization of drug targets: the TDR Targets database**. *Nat Rev Drug Discov* 2008, **7**(11):900-907.

17. Hu W, Sillaots S, Lemieux S, Davison J, Kauffman S, Breton A, Linteau A, Xin C, Bowman J, Becker J *et al*: **Essential gene identification and drug target prioritization in Aspergillus fumigatus**. *PLoS Pathog* 2007, **3**(3):e24.

18. Schneider P, Tanrikulu Y, Schneider G: **Self-organizing maps in drug discovery: compound library design, scaffold-hopping, repurposing**. *Curr Med Chem* 2009, **16**(3):258-266.

19. Sekhon BS, Bimal N: **Scaffold hopping in drug discovery**. *RGUHS Journal of Pharmaceutical Sciences* 2012, **2**(4):10.

20. Tsunoyama K, Amini A, Sternberg MJ, Muggleton SH: **Scaffold hopping in drug discovery using inductive logic programming**. *J Chem Inf Model* 2008, **48**(5):949-957.

21. Böhm H-J, Flohr A, Stahl M: **Scaffold hopping**. *Drug Discovery Today: Technologies* 2004, **1**(3):217-224.

22. Kaminski JJ, Rane D, Snow ME, Weber L, Rothofsky ML, Anderson SD, Lin SL: **Identification of novel farnesyl protein transferase inhibitors using three-dimensional database searching methods**. *Journal of medicinal chemistry* 1997, **40**(25):4103-4112.

23. De Lucca GV, Lam PY: **De novo design, discovery and development of cyclic urea HIV protease inhibitors**. *Drugs Future* 1998, **23**(9):987-994.

24. De Esch IJ, Mills JE, Perkins TD, Romeo G, Hoffmann M, Wieland K, Leurs R, Menge WM, Nederkoorn PH, Dean PM: **Development of a pharmacophore model for histamine H3 receptor antagonists, using the newly developed molecular modeling program SLATE**. *Journal of medicinal chemistry* 2001, **44**(11):1666-1674.

25. Barreca ML, Gitto R, Quartarone S, De Luca L, De Sarro G, Chimirri A: **Pharmacophore modeling as an efficient tool in the discovery of novel noncompetitive AMPA receptor antagonists**. *Journal of chemical information and computer sciences* 2003, **43**(2):651-655.

26. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL: **BLAST+: architecture and applications**. *BMC bioinformatics* 2009, **10**(1):421.

27. Lopez C, Chevalier N, Hannaert V, Rigden DJ, Michels PAM, Ramirez JL : **Leishmania donovani phosphofructokinase gene characterization, biochemical properties and structure modeling studies.** *Eur.J. Biochem.* 2002, **269** : 3978-3989.

28. Nowicki MW, Tullock LB, Woralll L, McNae IW, Hannaert V, Michels PAM, Fothergill-Gilmore LA, Walkinshaw MD, Turner NJ : **Design, synthesis and trypanocidal activity of lead compounds based on inhibitors of parasite glycolysis.** *Bioorganic & Medicinal Chemistry.* 2008,**16** : 5050–5061.

29. Verlinde CLMJ, Hannaert V, Blonski C, Willson M, Perie JJ, Fothergill-Gilmore LA, Opperdoes Fr, Gelb MH, Hol WGJ, Michels PAM : **Glycolysis as a target for the design of new anti – trypanosome drugs.** *Drug Resistance Updates.* 2001,**4**, 50–65.

30. Karp PD, Ouzounis CA, Moore-Kochlacs C, Goldovsky L, Kaipa P, Ahrén D, Tsoka S, Darzentas N, Kunin V, López-Bigas N: **Expansion of the BioCyc collection of pathway/genome databases to 160 genomes**. *Nucleic acids research* 2005, **33**(19):6083-6089.

31. Kuntal BK, Aparoy P, Reddanna P: **EasyModeller: A graphical interface to MODELLER**. *BMC research notes* 2010, **3**(1):226.

32. Holm L, Rosenström P: **Dali server: conservation mapping in 3D**. *Nucleic acids research* 2010, **38**(suppl 2):W545-W549.

33. Gordon MS, Schmidt MW: **Advances in electronic structure theory: GAMESS a decade later**. *Theory and Applications of Computational Chemistry: the first forty years* 2005:1167-1189.

34. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S: **General atomic and molecular electronic structure system**. *Journal of Computational Chemistry* 1993, **14**(11):1347-1363.

35. Hawkins PC, Skillman AG, Warren GL, Ellingson BA, Stahl MT: **Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database**. *Journal of chemical information and modeling* 2010, **50**(4):572-584.

36. Li L, Stoeckert,Jr. CJ, Roos, DS : **OrthoMCL : identification of ortholog groups for eukaryotic genomes.** *Genome Res. 2003,* **13**(9):2178 – 2189.

37. Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, Paley S, Rhee SY, Shearer AG, Tissier C: **The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases**. *Nucleic acids research* 2008, **36**(suppl 1):D623-D631.

38. Logan-Klumpler *et al.,* : **GeneDB - an annotation database for pathogens**. *Nucleic acids research* 2012, **40(D1)** : D98 – D108.

39. Leifso, K., Cohen-Freue, G., Dogra, N., Murray, A., McMaster, W.R. : **Genomic and proteomic expression analysis of leishmania promastigote and amastigote life stages**

: the leishmania genome is constitutively expressed. *Mol. Biochem. Parasitol.* 2007, **152** (1) : 35 – 46.

40. Krauth-Siegel RL, Meiering SK, Schmidt H: **The parasite-specific trypanothione metabolism of Trypanosoma and Leishmania**. *Biological chemistry* 2003, **384**(4):539-549.

41. Olin-Sandoval V, Moreno-Sanchez R, Saavedra E: **Targeting trypanothione metabolism in trypanosomatid human parasites**. *Current drug targets* 2010, **11**(12):1614-1630.

42. Irigoín F, Cibils L, Comini MA, Wilkinson SR, Flohé L, Radi R: **Insights into the redox biology of *Trypanosoma cruzi*: Trypanothione metabolism and oxidant detoxification**. *Free Radical Biology and Medicine* 2008, **45**(6):733-742.

43. Iribarne F, Paulino M, Aguilera S, Murphy M, Tapia O: **Docking and molecular dynamics studies at trypanothione reductase and glutathione reductase active sites**. *Molecular modeling annual* 2002, **8**(5):173-183.

44. Ariza A, Vickers TJ, Greig N, Armour KA, Dixon MJ, Eggleston IM, Fairlamb AH, Bond CS: **Specificity of the trypanothione-dependent Leishmania major glyoxalase I: structure and biochemical comparison with the human enzyme**. *Molecular microbiology* 2006, **59**(4):1239-1248.

45. Oppenheimer M, Valenciano AL, Sobrado P: **Biosynthesis of galactofuranose in kinetoplastids: novel therapeutic targets for treating leishmaniasis and Chagas' disease**. *Enzyme research* 2011, **2011**.

46. Lamerz A-C, Damerow S, Kleczka B, Wiese M, Van Zandbergen G, Lamerz J, Wenzel A, Hsu F-F, Turk J, Beverley SM: **Deletion of UDP-glucose pyrophosphorylase reveals**

a UDP-glucose independent UDP-galactose salvage pathway in Leishmania major. *Glycobiology* 2010, **20**(7):872-882.

47. Urbaniak MD, Tabudravu JN, Msaki A, Matera KM, Brenk R, Jaspars M, Ferguson MA: **Identification of novel inhibitors of UDP-Glc 4'-epimerase, a validated drug target for African sleeping sickness**. *Bioorganic & medicinal chemistry letters* 2006, **16**(22):5744-5747.

48. Majumder HK: **Drug targets in kinetoplastid parasites**: Springer; 2008.

49. Boitz JM, Strasser R, Yates PA, Jardim A, Ullman B: **Adenylosuccinate Synthetase and Adenylosuccinate Lyase Deficiencies Trigger Growth and Infectivity Deficits in Leishmania donovani**. *Journal of Biological Chemistry* 2013, **288**(13):8977-8990.

50. Boitz JM, Strasser R, Hartman CU, Jardim A, Ullman B: **Adenine Aminohydrolase from Leishmania donovani unique enzyme in parasite purine metabolism**. *Journal of Biological Chemistry* 2012, **287**(10):7626-7639.

51. French JB, Yates PA, Soysa DR, Boitz JM, Carter NS, Chang B, Ullman B, Ealick SE: **The Leishmania donovani UMP synthase is essential for promastigote viability and has an unusual tetrameric structure that exhibits substrate-controlled oligomerization**. *Journal of Biological Chemistry* 2011, **286**(23):20930-20941.

52. Hemsworth GR, Moroz OV, Fogg MJ, Scott B, Bosch-Navarrete C, González-Pacanowska D, Wilson KS: **The crystal structure of the Leishmania major deoxyuridine triphosphate nucleotidohydrolase in complex with nucleotide analogues, dUMP, and deoxyuridine**. *Journal of Biological Chemistry* 2011, **286**(18):16470-16481.

53. Michels PA, Bringaud F, Herman M, Hannaert V: **Metabolic functions of glycosomes in trypanosomatids**. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* 2006, **1763**(12):1463-1477.

54. Schüttelkopf AW, Hardy LW, Beverley SM, Hunter WN: **Structures of *Leishmania major* Pteridine Reductase Complexes Reveal the Active Site Features Important for Ligand Binding and to Guide Inhibitor Design**. *Journal of molecular biology* 2005, **352**(1):105-116.

55. Hardy L, Matthews W, Nare B, Beverley S: **Biochemical and Genetic Tests for Inhibitors of *Leishmania* Pteridine Pathways**. *Experimental parasitology* 1997, **87**(3):158-170.

56. Silva-Almeida M, Pereira BAS, Ribeiro-Guimarães ML, Alves CR: **Proteinases as virulence factors in Leishmania spp. infection in mammals**. *Parasites & vectors* 2012, **5**(1):1-10.

57. Silverman JM, Chan SK, Robinson DP, Dwyer DM, Nandan D, Foster LJ, Reiner NE: **Proteomic analysis of the secretome of Leishmania donovani**. *Genome Biol* 2008, **9**(2):R35.

58. Alves JM, Klein CC, da Silva FM, Costa-Martins AG, Serrano MG, Buck GA, Vasconcelos ATR, Sagot M-F, Teixeira MM, Motta MCM: **Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers**. *BMC evolutionary biology* 2013, **13**(1):190.

59. Patterson S, Alphey MS, Jones DC, Shanks EJ, Street IP, Frearson JA, Wyatt PG, Gilbert IH, Fairlamb AH: **Dihydroquinazolines as a novel class of Trypanosoma brucei**

trypanothione reductase inhibitors: discovery, synthesis, and characterization of their binding mode by protein crystallography. *Journal of medicinal chemistry* 2011, **54**(19):6514-6530.

60. Zhang Y, Bond CS, Bailey S, Cunningham ML, Fairlamb AH, Hunter WN: **The crystal structure of trypanothione reductase from the human pathogen Trypanosoma cruzi at 2.3 Å resolution**. *Protein science* 1996, **5**(1):52-61.

61. Riley-Lovingshimer MR, Ronning DR, Sacchettini JC, Reinhart GD: **Reversible ligand-induced dissociation of a tryptophan-shift mutant of phosphofructokinase from Bacillus stearothermophilus**. *Biochemistry* 2002, **41**(43):12967-12974.

62. Harkiolaki M, Dodson EJ, Bernier-Villamor V, Turkenburg JP, González-Pacanowska D, Wilson KS: **The crystal structure of *Trypanosoma cruzi* dUTPase reveals a novel dUTP/dUDP binding fold**. *Structure* 2004, **12**(1):41-53.

63. Shi W, Schramm VL, Almo SC: **Nucleoside hydrolase from Leishmania major Cloning, expression, catalytic properties, transition state inhibitors, and the 2.5-Å crystal structure**. *Journal of Biological Chemistry* 1999, **274**(30):21114-21120.

64. Ariza A, Vickers TJ, Grieg N, Armour KA, Dixon MJ, Eggleston IM, Fairlam AH, Bond CS: **Specificity of the trypanothione-dependent Leishmania major glyoxalase I : structure and biochemical comparison with the human enzyme.** *Mol. Microbiol.* 2006, **59**:1239 – 1248.

CHAPTER III

1. Introduction

The parasitic protozoa belonging to the order trypanosomatida are responsible for a wide range of diseases collectively termed as leishmaniases, affecting humans as well as livestock. The estimated annual occurrence of new cases of the disease are 0.2 to 0.4 million for visceral leishmaniasis and 0.7 to 1.2 million for cutaneous leishmaniasis [1]. Cutaneous leishmaniasis is generally caused by *L. tropica* and *L. major* in the old world and by the pathogens from the two subgenera *Leishmania leishmania* and *Leishmania viannia* in Latin America [2]. People with compromised immune systems are especially susceptible to the disease [2].

The emergence of strains resistant to the existing first line of defense provided by the pentavalent antimonial compounds and its associated toxic side effects has appeared as a major health issue in the developing countries. The availability of the new formulations of the old drugs along with combinatorial drug therapy may delay the emergence of resistance, yet are not adequate due to toxicity or prohibitive prices. There is an urgent need for the development of alternative therapies that would provide adequate and affordable drugs with the least toxic side effects.

The enzyme phosphoglucomutase (PGM) reversibly catalyses the conversion of glucose-1-phosphate (G1P) to glucose-6-phosphate (G6P) by the transfer of a phosphate between the C6 and C1 hydroxyl groups of glucose. A high degree of conservation of the PGM gene has been reported between the two sister species, *T. cruzi* and *L.major* [3]. Studies indicate that the outer surfaces of trypanosomatids are covered with glycoconjugates whose glycans are primarily composed of galactopyranose (Gal*p)* and galactofuranose (Gal*f*) residues [4]. Glycoconjugates such as mucin-type glycoproteins and glycoinositol phospholipids have been implicated as

virulence factors and play a crucial role in evading the host's immune response [4, 5]. In *T. cruzi* as well as other leishmanial species , the biosynthesis of galactose takes place via the sugar nucleotide synthesis pathway involving the conversion of uridine-diphosphate-glucose (UDP-Glc) to UDP-galactose (UDP-Gal) catalyzed by UDP-Gal-4'-epimerase. Since the generally available trypanosomatid hexose transporters do not transport D-Gal, *leishmania spp.* relies solely on glucose as the starting point for the generation of UDP-Gal [5]. Besides being an important metabolite, UDP-Gal is essential for the growth of the parasite and the presumed donor for the synthesis of a rare deoxyneucleotide (β-D-glucosylhydroxymethyluracil) present in the DNA of trypanosomatids [5]. The essentiality of the glycoconjugates for the survival of the parasite inside the host renders enzymes of the sugar nucleotide synthesis pathway as potential therapeutic target.

The two structurally characterized eukaryotic PGMs from rabbit (OcPGM) and *P. tetrauralia* (PtPGM) share ~50% sequence homology with the phosphoglucomutase from *L. major* (LmPGM). Out of the three PGM genes present in human (PGM_1, PGM_2 and PGM_3) [6], the rabbit muscle PGM was found to share 97% sequence homology with PGM_1. The NCBI GenBank indicated the presence of a single copy of the PGM gene on chromosome-21, encoding 589 amino acids in *L.major*, *L.infantum*, *L.mexicana, L.brazilliensis* and *L.donovani* while *T.cruzi* (strain CL Brener) contained two copies of the gene encoding 600 residues.

Apart from the glycolytic enzymes, phosphoglucomutase was found to be located inside glycosomes, the peroxisome like organelle, in *T.cruzi* [3]. A novel internal non-peroxisomal targeting motif (PTS; corresponding residues 260-380 in LmPGM) was sufficient for its transportation into the glycosome in the absence of the classical terminal PTS motifs. The

compartmentalization of the enzymes involved in the biosynthesis of sugar nucleotide, may have regulatory effect on the synthesis of metabolites [5].

Assays developed using the PGM gene sequence as a marker finds extensive application in distinguishing different strains of trypanosomatids based upon their isoenzyme patterns [4, 7]. In vitro studies indicate that the invasion of mammalian cells by the metacyclic trypanomastigotes (*T. cruzi*) involves the activation of distinct signal transduction pathways triggered by the interaction of parasitic surface molecules with target cells [4]. This leads to the intracellular Ca^{2+} mobilization and subsequent increase in cytosolic Ca^{2+} concentration both in parasite and the host cell, a pre-requisite for invasion. In yeast, the accumulation, storage and maintenance of intracellular Ca^{2+} homeostasis is coupled to the relative concentration of G1P and G6P, which in turn, is regulated by phosphoglucomutase. It may, therefore, be possible that LmPGM plays a similar role in the maintenance of the intracellular Ca^{2+} homeostasis during normal cell functioning and increase in Ca^{2+} level during invasion via the regulation of cellular G1P-G6P concentrations.

This chapter consists of the crystal structure of the third eukaryotic phosphoglucomutase from *Leishmania major*, with an estimated molecular weight of ~64 kDa. The overall fold of the enzyme was found to be similar to other reported eukaryotic phosphoglucomutases, with the presence of four domains arranged in an overall heart-shape and a large central cleft harbouring the catalytic residues. There were seven insertion regions compared to OcPGM, present on the periphery of the molecule, in the form of extended loops, away from the active site. Normal mode analyses revealed the conformational variability of the fourth domain, via a hinge region, as was previously observed in other PGMs of α-D-phosphohexomutase superfamily. The

catalytic properties of the enzyme and the prospect for inhibitor design have also been discussed in this chapter.

2. Materials and methods

2.1 Materials

Lysozyme, glucose-1-phosphate, glucose-6-phosphate dehydrogenase from baker's yeast (*S.cerevisiae*) were obtained from Sigma Aldrich and EDTA free protease inhibitor cocktail tablets, thrombin protease were procured from Roche Diagnostics and Novagen respectively. Sephadex G-100 for gel filtration was from Amersham Biosciences and bovine albumin fraction V was bought from HIMEDIA. All other chemicals if not mentioned otherwise, were from Sisco Research Laboratories Pvt. Ltd.

2.2 Protein expression, purification and crystallization

The plasmid construct containing the phosphoglucomutase gene from *Leishmania major* (GeneDB: LmjF21.0640; UniProtID: Q4QCF1, EC: 5.4.2.2) was purchased from GenScript (http://www.genscript.com/). In this construct, the PGM gene was cloned into the Nde1 and BamH1 restriction sites of the pET-28a (+) vector with an N-terminal His_6 tag followed by a thrombin restriction site. This construct was used to transform *E.coli* Rosetta competent cells followed by overnight incubation on LB-agar/kanamycin-chloramphenicol plates at 37° C. A single colony was used to inoculate the LB media containing kanamycin, chloramphenicol and incubated at 37°C until OD_{600} attained a value of ~0.6. IPTG (isopropyl-β-D-thiogalactopyranoside) was then added to reach a final concentration of 1mM for the induction of

LmPGM, followed by overnight incubation at 17°C. The cells were harvested by centrifugation at 7000g for 10 minutes and washed with buffer containing 10mM $Na_2HPO_4.2H_2O$, pH 7.4, 2mM KH_2PO_4, 137mM NaCl and 2.7mM KCl. The cell pellets were stored at -80°C until further use.

The cell pellet was thawed at 4°C and all the subsequent steps were performed at 4°C. The pellet was suspended in lysis buffer (50mM Tris-HCl, pH 8.0, 300mM NaCl, 10mM Imidazole, 5mM β-mercapto-ethanol, 0.01% Triton X-100, 2 mM phenyl methane sulphonyl fluoride) and lysozyme, protease inhibitor cocktail were added to reach a final concentration of 1 mg/ml and 1X respectively. Subsequent to sonication, the homogenized lysate was centrifuged at 14000g for 45 minutes and the supernatant was loaded onto a pre-equilibrated Ni-NTA column (QIAGEN Ni-NTA super flow).The column was washed with 50 ml of wash buffer (50mM Tris-HCl, pH 8.0, 300mM NaCl, 20mM Imidazole, 5mM β-mercapto-ethanol) and finally eluted with elution buffer (50mM Tris-HCl, pH 8.0, 300mM NaCl, 250mM Imidazole, 5mM β-mercapto-ethanol). To remove the N-terminal leader sequence (His_6 tag), 0.1U/mg of restriction grade thrombin protease was added to the eluted protein and was dialyzed overnight against the dialysis buffer (25mM Tris-HCl, pH 7.5, 100mM NaCl, 3mM DL-dithiothreitol, 1% glycerol, 0.02% sodium azide). The dialyzed protein mixture was loaded onto a Sephadex G-100 column, pre-equilibrated with the dialysis buffer containing 25mM Tris-HCl, pH 7.5, 100mM NaCl, 3mM DL-dithiothreitol, 1% glycerol, 0.02% NaN_3. In order to estimate the molecular weight of LmPGM, Dextran blue (MW: 2000000) was used to estimate the void volume and bovine serum albumin (66kDa), lysozyme (14.4kDa) were used as standards. The enzyme was concentrated using VIVASPIN 15 (Sartorius stedim, 10 kDa molecular weight cutoff) by centrifugation at 2500g. The purity of the protein was confirmed by SDS-PAGE

electrophoresis which exhibited a single band of expected molecular weight. The concentration of the protein was determined by the Bradford method using bovine serum albumin as the standard [8]. The purified protein was divided into 50 µl aliquots, flash cooled in liquid nitrogen and stored at -80^{o}C.

2.3 Activity assay for the conversion of G1P to G6P

The activity of the enzyme was assayed by coupling the formation of G6P to NADPH using the enzyme glucose-6-phosphate dehydrogenase (G6PDH). LmPGM, at a final concentration of 1µM, was added to the 400 µL reaction mixture containing 50mM MOPS, pH 7.4, 3mM $MgCl_2$, 0.2mM $NADP^+$ and 1.25 U/ml G6PDH. For the kinetic analysis, the concentration of G1P was varied between 10- 200 µM. The formation of NADPH was monitored by measuring the absorbance at 340nm in the absence and presence of the activator glucose-1,6-bisphosphate. Kinetic data were fitted to the appropriate equations ($Y=B_0 + B_1X + B_2X^2 + B_3X^3$) by nonlinear regression methods using *GraphPad Prism 5*. Initial velocities versus substrate concentration were fitted to the Michaelis-Menten equation and the K_m, V_{max} and K_{cat}/K_m were calculated using the same software. All the data were measured in duplicate and the results are shown as the mean±SD.

2.4 Crystallization

Initial crystallization screens were set up at 20°C using the hanging drop vapour diffusion method using PEG3350 as the precipitant. Drops containing 2µl of protein solution at a concentration of 10mg/ml and 2 µl of crystallization buffer were sealed over a 1 ml reservoir. The protein crystallized in crystallization buffer containing 0.1M Tris-HCl, pH 8.0, 10% (w/v)

PEG 3350, 10mM $MgCl_2$ and 0.02% NaN_3. For cryo-protection, the crystals were transferred to a solution containing 0.1M Tris-HCl, pH 8.0, 15% (w/v) PEG 3350, 10mM $MgCl_2$, 30% (v/v) glycerol and 0.02% NaN_3. The crystals were mounted in nylon loops, flash frozen in liquid nitrogen and stored until required for data collection.

2.5 Data collection and processing

Diffraction data for the LmPGM crystals were collected on Beamline-21 of the synchrotron at the Raja Ramanna Centre for Advanced Technology (RRCAT, Indore, India) utilizing a Rayonix MX-225 CCD detector. Several crystals were tried until one crystal diffracted to a resolution of 3.5 Å at 100° K. The data were integrated using the software *XDS* [9] and scaled with *Aimless* from *CCP4* suite of programs [10]. The data collection statistics are given in **Table 1.**

2.6 Structure determination and refinement

LmPGM was found to share 53% sequence identity with rabbit muscle phosphoglucomutase (resolution 2.4 Å; PDB accession code: 3PMG) [11]. After the removal of all water molecules and ions, a polyALA model from 3PMG was generated by *Chainsaw* from the CCP4 package and used as the search model to generate the initial solution by molecular replacement using *Phaser* [12, 13]. The solution was found to contain four molecules in the asymmetric unit, corresponding to a solvent content of 54.34%. The calculation was repeated with 3 and 5 molecules in the asymmetric unit which did not yield significant results. The solvent content for 3,4,5 molecules in the asymmetric unit was 65.8, 54.34 and 43.0 %, respectively. 5% of the dataset was set aside for cross validation prior to refinement.

Table 1: The data collection and structure-refinement statistics.

Statistics for the highest-resolution shell are shown in parentheses.

Data collection statistics	
X-ray source	BL-21, RRCAT, INDIA
Space group	I121
Unit cell dimensions	a = 125.92, b =114.97, c = 198.37 α = 90 , β = 106.58 , γ = 90
λ (Å)	0.979470
Resolution (Å)	39.65-3.50 (3.71-3.50)
R_{merge} (all I+ and I-)	0.148 (0.512)
R_{merge} in top intensity bin	0.033
Total number of observations	91999 (14825)
Total number of unique reflection	29659 (4882)
Mean ((I)/σ(I))	8.3 (2.3)
Completeness (%)	86.6 (88.5)
Multiplicity	3.1 (3.0)
Refinement Statistics	
[a]R_{work}	0.2437 (0.3168)
[b]R_{free}	0.2990 (0.3687)
Number of non-hydrogen atoms	17136
Number of macromolecules	17132
Ions (Mg^{2+})	4
Protein residues	2244
[c]RMS(bonds Å)	0.006
RMS(angles °)	1.50
Ramachandran favored (%)	83.32
Additionally allowed (%)	9.61
Ramachandran outliers (%)	7.07
Clashscore	24.21
Average B-factor (Å^2)	
macromolecules	71.90
Ions (Mg^{2+})	11.80

[a]$R_{work} = \sum|F_o-F_c|/\sum|F_c|$, where F_o and F_c are the observed and the calculated structure factors respectively. [b]R_{free} is the R factor calculated from 5% of the reflections not included in refinement. [c]RMS, root mean square deviation. Refinement statistics were generated using *Phenix.refine* [15].

Electron density was extremely poor for the first 30 residues in all the four chains and these stretches of polypeptide chains were removed from subsequent calculations (LmPGM comprises a total of 589 residues). The Non-crystallographic symmetry (NCS) operators were calculated using *Phenix_Find_NCS* [14], and restraints on NCS (type: torsion-angle) were maintained throughout the refinement process. Iterative cycles of refinement using the software *Phenix.refine* [15] and extensive manual model building using the software *COOT* [16] were performed. With the progress in refinement missing sections of the polypeptide chain gradually emerged. In the final cycles of refinement, the model was refined using TLS groups generated by *Phenix.refine*. Since the crystallization buffer contained magnesium, the presence of significant electron density at the putative metal binding led to the placement of the divalent cation, Mg^{2+}.

The final model, validated using *Molprobity* [17], had a R_{work} and R_{free} of 24.4 and 29.9 % respectively. 7.07 % of the residues were found in the disallowed region of the Ramachandran Plot. The r.m.s deviation of bond lengths and bond angles from their ideal values were 0.0062 Å and 1.498° respectively. For all the four molecules in the asymmetric unit, the first 24-30 from the amino terminal could not be modeled due to the absence of definite positive electron density and thus the final model contained residues 25 to 589, 31 - 589, 29 - 589 and 31 - 589 for the four polypeptide chains (A,B,C,D respectively.) The final model comprised 2244 amino-acid residues and 4 magnesium ions.

The structural figures were prepared using *Pymol* [18]. The coordinates were deposited in the Protein Databank with the assigned code 4QG5.

2.7 Normal mode analysis

The low frequency normal modes were used to model the conformational flexibility of the fourth domain of LmPGM using the web-server: elNémo [19]. The crystal structure of LmPGM was used as the initial model. The protein-inhibitor complex of rabbit exhibited maximum conformational variability compared to LmPGM (C_α rmsd 0.988; PDB accession code: 1C4G, Baranidharan *et al*., unpublished work) than the phosphorylated (C_α rmsd 0.872; PDB accession code: 3PMG) and dephospho-enzyme-intermediate complex of OcPGM (C_α rmsd 0.818; PDB accession code: 1C47). Molecule A of '1C4G' was modified as per the requirements of the program and was used as the second model. A series of low frequency normal modes were calculated and perturbations were applied to the lowest vibrational mode (mode 7) with amplitude varying from -100 to 100 in steps of 20. A total of 11 models were generated by the program.

2.8 Docking the substrate and the intermediate to the active site

Molecular docking was performed using the program *AutoDock4* [20]. While docking, the key residues of the receptor and the substrate were treated as flexible and the divalent cation (Mg^{2+}) was retained. The receptor molecule was either the crystal structure of LmPGM or one of the models generated by normal mode analysis. In order to mimic the phospho form of the enzyme, the coordinates corresponding to the phosphate group were manually incorporated into LmPGM from the phospho-enzyme-substrate complex of *P. aeruginosa* after superposition (PDB code: 1P5D) [21]. The newly constructed phosphoserine SEP118 was maintained flexible throughout the docking.

The genetic algorithm was selected as the search algorithm, number of individuals in population (pop_size) was set to 150, rate of gene mutation (mutation_rate) was set to 0.02 and rate of crossover (crossover_rate) was set to 0.8. The ligands as well as the active site residues S118, H119, E397, S399, R304, R535, R523, and R454 were treated as flexible while docking. The solutions were chosen from the cluster with maximum number of solutions exhibiting lowest binding energy and by manual inspection.

3. Results and discussion

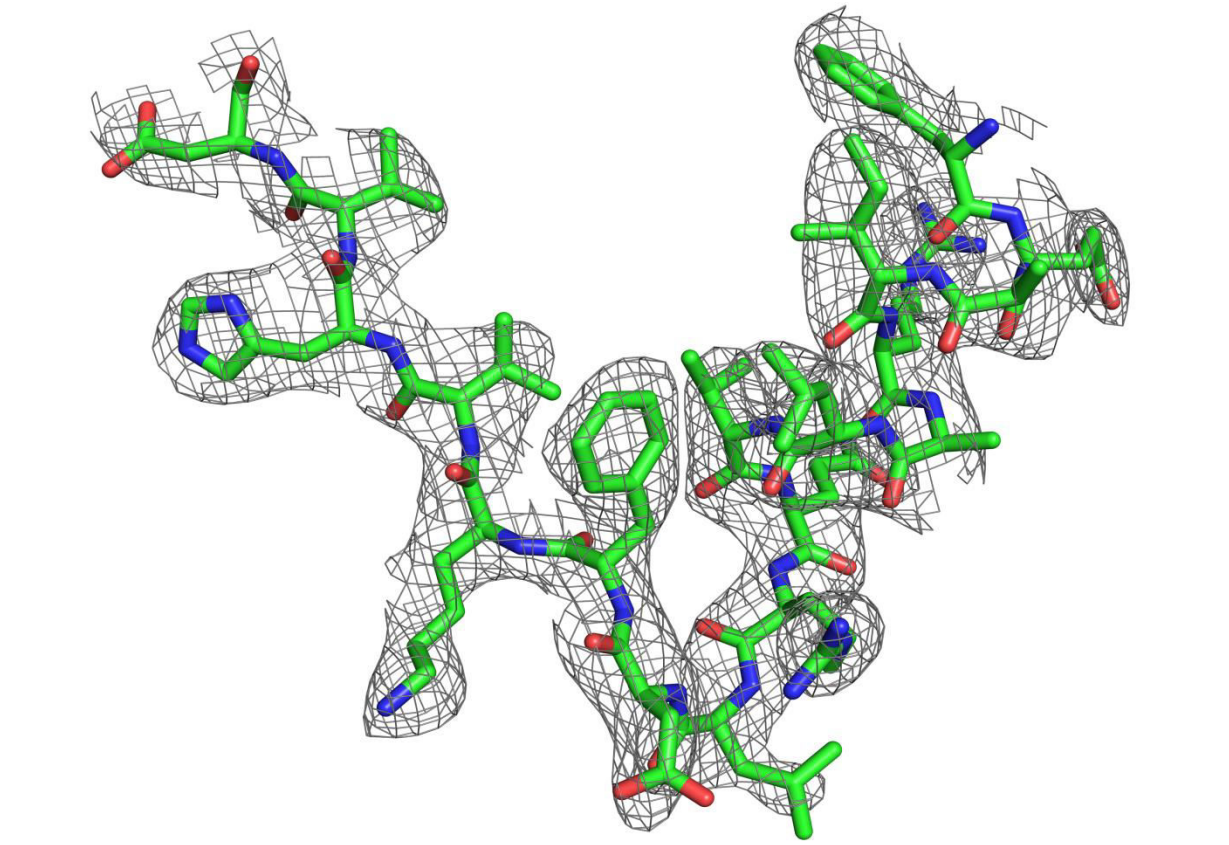
3.1 Description of the contents in the asymmetric unit

Four molecules of LmPGM (Chains A, B, C, D) were found in the asymmetric unit and their mutual Cα RMSD's are shown in **Table 2**. The $2F_o$-F_c electron density maps of a strech of residues contoured at 1.0 σ are shown in **Fig. 1**. The coordinate (**rcsb086012.pdb**) and structure factor files (**rcsb086012-sf.cif)** corresponding to the PDB entry 4QG5 are provided separately with the CD-ROM.

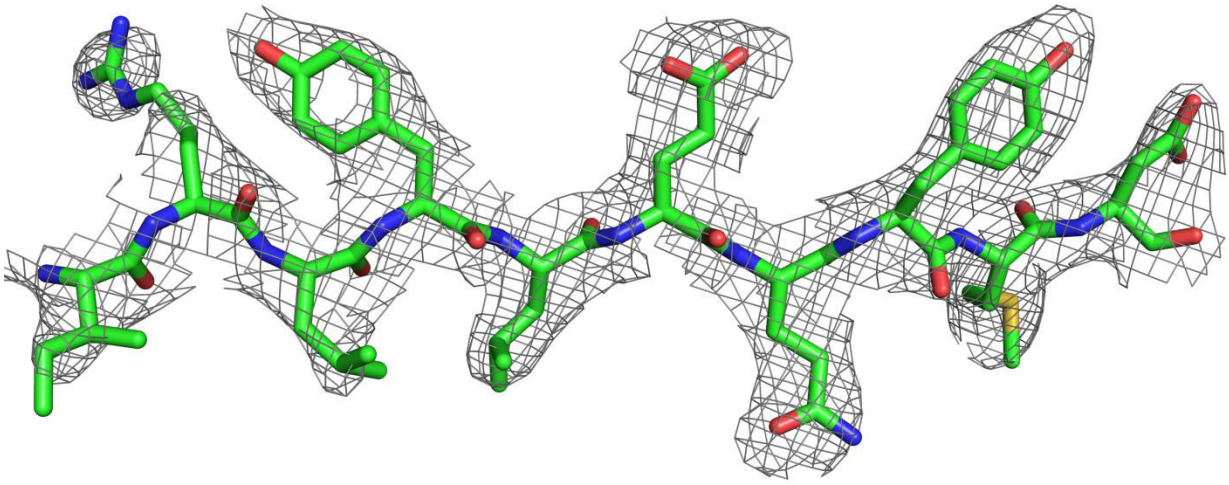
Table 2: The C_α RMSD (Å) between the four chains of LmPGM.

	B	C	D
A	0.58	0.49	0.69
B		0.56	0.45
C			0.63

Fairly good agreement was observed between the subunits with an average C_α RMSD of 0.57 Å (0.05), with a maximum deviation between A, D (0.69 Å) , while the minimum deviation was 0.45 Å between B,D. As expected, structural variability was found in the loops, with



(a)



(b)

Figure 1: The $2F_o$-F_c map contoured at 1.0 σ (a) Electron density map of chain A: residues 203-220 (b) Electron density map of chain A: residues 534-543

very good superposition in regions of regular secondary structure. The relative orientation between the four domains was also conserved in all the four chains. The average Cα rmsd between the search model rabbit phosphoglucomutase and the four chains of LmPGM was found to be 0.872 Å (chain A), 0.773 Å (chain B), 0.800 Å (chain C) and 0.863(chain D). A more detailed description of the inter-domain geometry will be found in the sections given below.

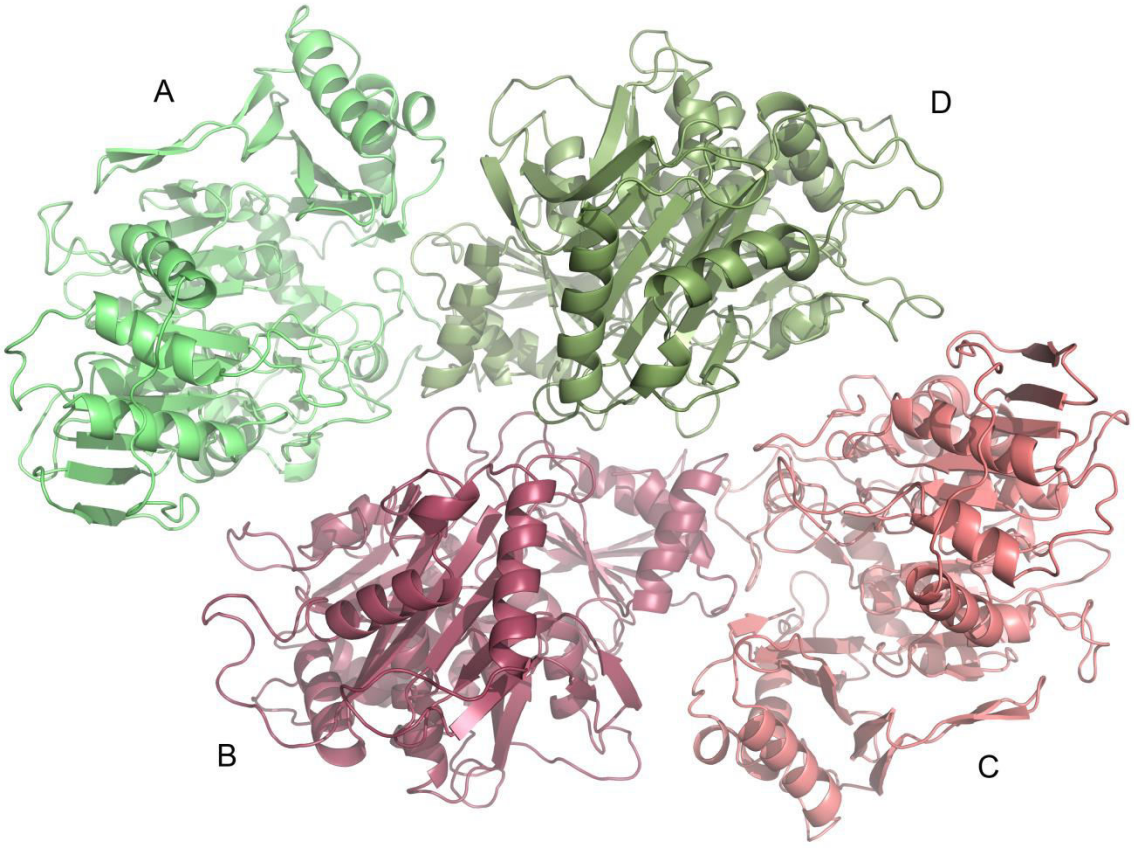


Figure 2: **The arrangement of the four molecules in the asymmetric unit of LmPGM. Molecule A and D, coloured green while molecule B and C, coloured red respectively.**

Interestingly, of the four polypeptide chains, A,D and B,C were related by two independent non – crystallographic axes of about 100° (B - C : 99.8°; A - D : 101.6°). Both the pairs (A, D and B,

C) were of identical inter – subunit geometry as A-D was superimposable on B-C with a C_α rmsd of 0.452 Å (**Fig. 2**). This situation was somewhat reminiscent of the crystal structure of rabbit phosphoglucomutase (OcPGM: which was used as a search model to solve the present structure) wherein two molecules of OcPGM were found in the asymmetric unit (space group $P4_122$) related by a non crystallographic two fold axis. In solution, OcPGM is a monomer with a molecular weight of 61 kD, and all structural considerations confirmed that the pseudo dimeric association was indeed a consequence of crystal packing [11]. Similar to OcPGM, gel filtration studies with LmPGM unambiguously confirmed the monomeric form of the molecule in solution, and its co-elution with bovine serum albumin of molecular weight ~66 kDa under identical experimental conditions (**Fig**. **3**).

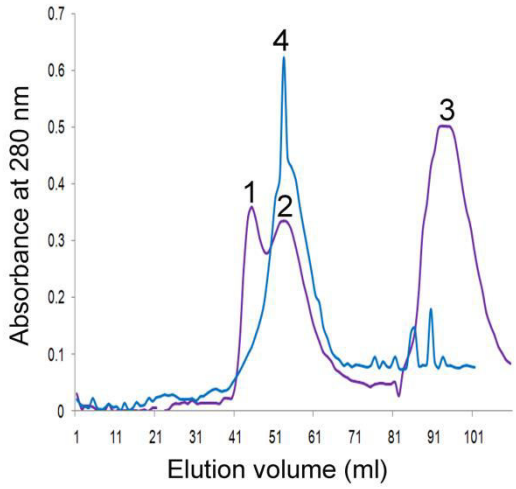


Figure 3: **Chromatograms of LmPGM and the standards on a Sephadex G-100 size exclusion column. The peaks in the figure correspond to: 1. Blue Dextran (MW: 2,000,000), 2. BSA (66 kDa), 3. Lysozyme (14.4 kDa), 4. LmPGM (~64 kDa). The volumes on X-axis represent the elution volume (1ml fractions were collected at a flow rate of 0.5 ml/min) and the data on the Y-axis corresponds to the absorbance measured at 280 nm.**

Thus, the association between subunits A-D and B-C wherein the geometrical relationship between the subunits is related by 100° almost certainly appears to arise from crystal packing. Generally, natural dimers (in solution) are related by 180° axes in the overwhelming majority of cases. However, the interface between A-D (or B-C) and the two subunits of OcPGM (related by the non – crystallographic 2 fold axis) is substantially different with no significant overlap, with the buried solvent accessible area (SAA) of the interface in LmPGM (about 2221.4 $Å^2$) being almost double the area of the interface in the OcPGM dimer (1027.5 $Å^2$). The interaction between subunits (A-D or B-C) is mediated by a helix and its contiguous loops from domain II on one subunit (A,C) with loops present in domain I from the other subunit (B,D : see section below for description of domains). In addition there are significant contacts between a loop in domain IV (A, C: extending from 577 – 588) and Tyr 385 (B, D) also from the same domain. The secondary structural assignment of the molecule is provided in **Supplementary Table 1**.

3.2 Description of the monomer

Each molecule of LmPGM is composed of four domains which pack into a compact 'heart shape' of approximate molecular dimensions 73 Å by 66 Å by 49 Å. All four domains are sequentially contiguous and range from 25 – 187 (domain I), 188 – 313 (II), 314 – 448 (III) and 449 – 589 (IV) respectively. Domains I-III share a conserved topological core ('domain motif') constituted of a four stranded β – sheet flanked by two α helices (**Fig. 4**). The ordering of the strands in the β – sheet are: three parallel β strands with the fourth strand being anti parallel to the rest (↓↓↓↑). Although domains I-III appear to possess a conserved α|β fold (**Fig. 4, 5**), this structural similarity does not appear to be reflected in their sequences, whose mutual identities

fall below 20%. In addition to the conserved core additional secondary structural elements are found in the peripheries of these domains with 2 β strands , 2 α helices (in domain I), 3 α helices (in domain II) and 4 α helices (in domain III). In contrast, the fourth domain (IV) is structurally distinct from the first three, being a member of the TATA – box binding protein like fold superfamily , constituted of a four stranded antiparallel β sheet flanked by two helices and two shorter β strands.

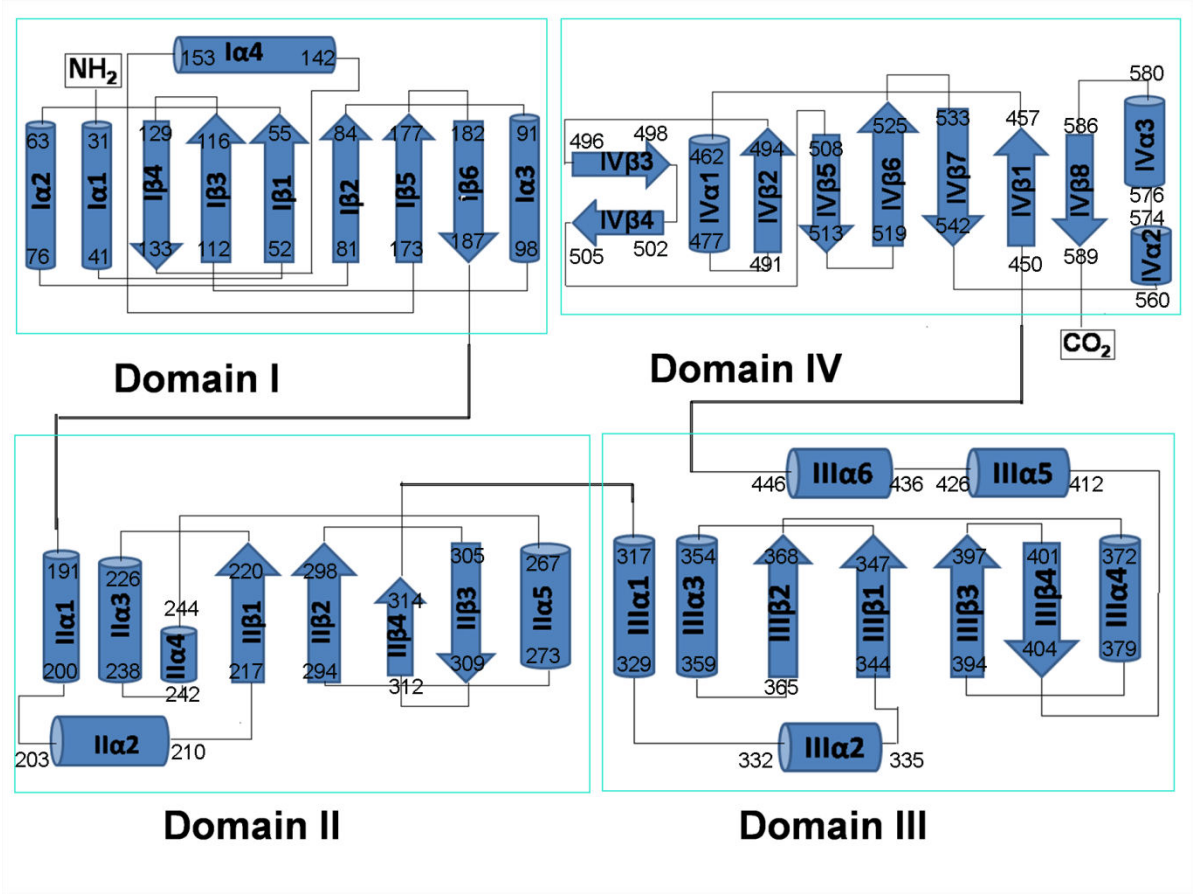


Figure 4: A schematic of the topology of the secondary structural elements of the four domains. Parallel β-strands: arrows pointing in the same direction; α-helices: cylindrical shapes.

The conformational flexibility of the fourth domain with respect to the rest of the molecule is crucial for the catalytic activity of the protein. The most extended buried solvent accessible area

arising out of inter – domain packing within the monomer is between domains II and III (I-II: 1900 $Å^2$; I-III: 839 $Å^2$; II-III: 2989 $Å^2$ and III-IV: 2540 $Å^2$). Thus given the shape of the molecule the maximum association of any domain is with the domains adjacent to it in sequence.

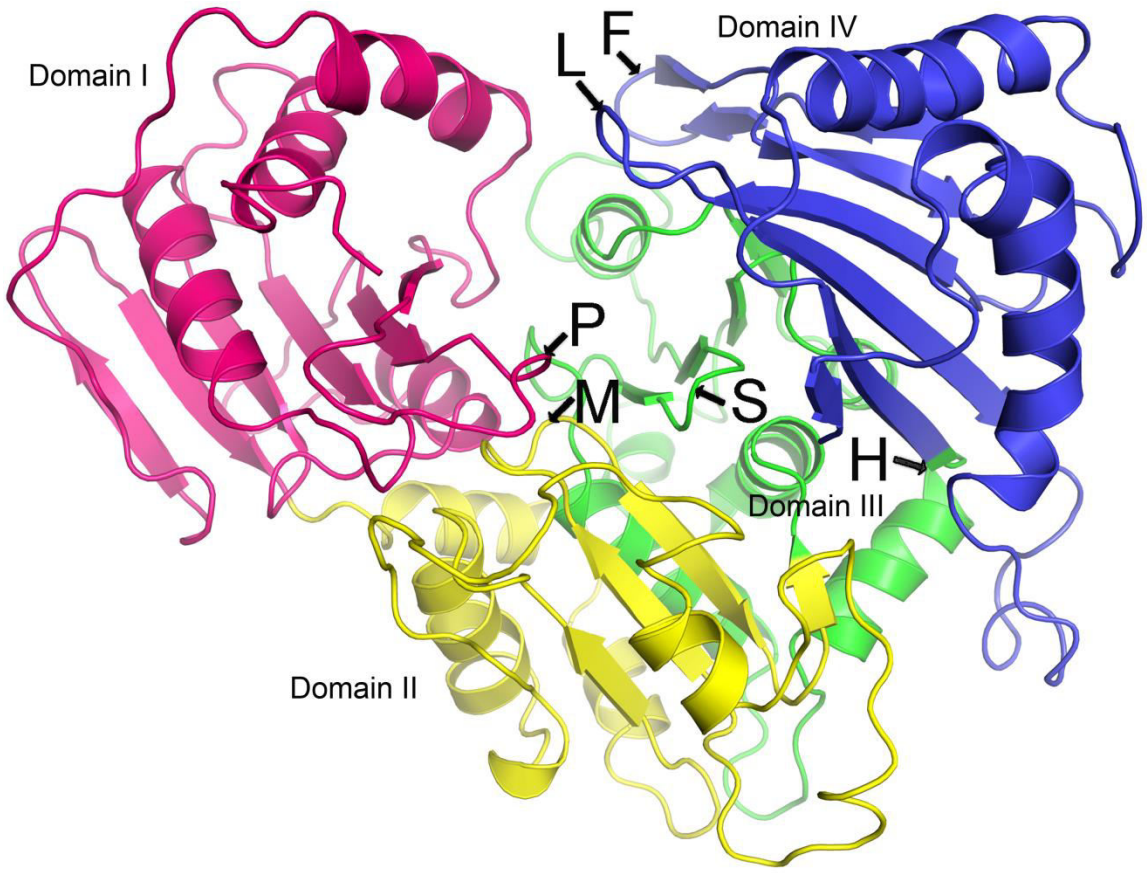


Figure 5: The four domains colored red, yellow, green and blue respectively. L: 'latch' or the 'lid', F: extended flap, P: proximal phosphate transfer site, M: metal binding site, S: the substrate binding loop, H: hinge region.

The principal interaction sites between domains I and II are the flanking helix (Iα3) of the conserved 'domain motif' along with its contiguous loops in domain I in contact with a loop interconnecting strands IIβ2 – IIβ3 of the conserved motif in domain II (nomenclature for helices

and strands as given in **Fig. 4** wherein the character/number of the secondary structural element is preceded by the domain number). In addition, loops between Iβ1-Iα2, Iβ3-Iβ4 and IIα4-IIα5 mediate extensive contacts between the two domains (**Fig. 4**). The inter-domain interface between domains II and III on the other hand, is constituted by helices IIα1, IIα2; strand IIβ3 and the helix IIIα5 along with its adjacent loops. As has been mentioned previously the conformational flexibility of domain IV enables the closure of the extended active site of the molecule. Based on the geometrical position of domain IV with respect to the other domains the molecule can be considered to be in its 'closed' or 'open' state. More will be said of the conformational flexibility of domain IV and its relationship to the catalytic activity of the protein in the succeeding sections. From normal mode analysis (see section on normal modes) it appears that the crystal structure of LmPGM lies at an intermediate conformational state between the 'closed' and 'open' forms (**Fig. 5**). The prominent interactions between domains III and IV in such a conformational state involves helices IIIα3, IIIα4, IIIα6 along with their adjacent loops and beta strands IVβ1, IVβ3, IVβ6 in addition to loops linking IVβ2 – IVβ3, IVβ7 – IVα2 respectively.

In addition to interdomain packing each individual domain also incorporates a unique hydrophobic core. Despite the fact that domains I – III have a conserved 'domain motif' there are significant differences in both the topology and size of the hydrophobic cores in the first three domains (**Table 3**), which however invariably includes interactions which mediate the packing between the α helices and the β – strands within the respective domains. The core in the fourth domain is predominantly constituted of residues located on adjacent β strands (IVβ5 – 7) , with only two residues (Leu 566, Ala 570) contributed by helix IVα2.

.

Table 3: Residues constituing the unique hydrophobic core in each domain.

Domain	Residues constituting the hydrophobic core
I	Phe 34, Val 35, Phe 39, Leu 52, Val 54, Ile 69, Val 78, Val 81, Phe 113, Phe 182
II	Val 219, Val 230, Ile 233, Phe 234, Leu 245, Val 296 ,
III	Leu 320, Leu 323, Phe 333, Leu 394, Phe 418
IV	Val 472, Val 512, Phe 514, Phe 520, Leu 522, Leu 536, Leu 538, Leu 566,Ala 570

.

A comparison of the rabbit phosphoglucomutase, revealed the presence of six major insertion regions, ranging from three residues to a maximum of fourteen residues, in the form of peripheral loops situated away from the active site. The insertion regions were from residues 101-106 (insrt1: RRDADG), 275-288 (insrt2: LLPDGNANPAMKHI), 337-339 (insrt3: SSS), 386-393 (insrt4: GGKDFNPL), 430-435(insrt 5: GTPLVG) and 546-552 (insrt6: TVKSHLA). More will be said obout these insertion regions while discussing possible inhibitor sites on the parasitic enzyme (see section on **Prospects for Inhibitor Design**)

The conformational flexibility of domain IV is a common feature observed within the members of the α-D-phosphohexomutase superfamily, with the whole domain moving via a hinge type motion reported to be essential for the enzyme function [22]. The substrate-P368A/G mutant complex of *P. aeruginosa* PGM (PaPGM) was found to adopt a more open conformation compared to the wild type [23]. P368 was reported to act as the hinge joint and the k_{cat}/K_m of the mutant dropped to 10% that of wild type, affirming the crucial role of the domain motion for the proper functioning of the enzyme. The residue G448 in LmPGM corresponding to P368 in

PaPGM, present at domain III-domain IV junction, was found to be conserved not only within the trypanosomatids, but also in paramecium and rabbit (**Supplementary Fig. S1**). Residues defining the hinge region (447-YGR-449) were located on the small loop connecting the terminal α-helix of domain-III (IIIα6) and the initial β-strand of domain-IV (IVβ1) and were collectively designated as 'H' (**Fig. 4, 5**).

The hairpin flap connecting the sixth (IVβ6) and seventh β-strand (IVβ7) of domain IV appeared to cover the active site from above, like a lid. The structural study of PtPGM revealed that the hairpin flap of domain IV interacts directly with another hairpin flap from domain I, like a latch, via two threonine residues [24]. These two residues were highly conserved among other structurally characterized PGMs and correspond to T15 and T527 in LmPGM. The mutation of the threonine residue from domain I in *A. xylinum* PGM was accompanied by a drop in catalytic efficiency to 0.04% that of wild type [25]. The interaction between these two segments from two different domains was reported to be essential for adopting the closed conformation resulting in proper positioning of the substrate within the catalytic groove. Due to the absence of the initial 24-30 residues from all the four subunits of LmPGM, the hairpin flap of domain IV alone (525-SGTGSSGAT-533) was designated as 'L' (**Fig. 4, 5**).

The extended hairpin flap 'F', connects the second (IVβ2) and fifth β-strands (IVβ5) of the β-sheet (Topology: 256718) from domain IV. The flap contains antiparallel β-strands IVβ3 and IVβ4 (**Fig. 4**) and appears to extend outward to embrace domain III (**Fig. 5**). The flap was present in paramecium, rabbit and bacteria such as *T. thermophilus*, *S. enterica*, *S. typhimurium* but was reduced to a turn, with the absence of the antiparallel β-strands, in archaea (*P. horikoshii*, *S. tokodaii*) and some of the bacteria (*B. melitensis*, *P. aeruginosa*, *B. thilandensis*).

3.3 The active site and specificity pocket

The active site of LmPGM forms a part of a deep cleft extending into the center of the molecule, which is composed of atoms contributed by all four domains. In the absence of the initial 24-30 residues, the solvent accessible area of the cleft was calculated to be 2371.7 $Å^2$. Conventionally, the active site of PGM's (**Table 4, Supplementary Fig. S2**) are analyzed in terms four distinct regions (i) the location of the active site serine (S118) implicated in phosphoryl group transfer (Proximal P loop : 117 - 128), (ii) the metal binding loop (M loop : 299 - 304), (iii) the sugar binding loop (S loop : 397 - 400) also considered to be the determinant of the sugar specificity of the enzyme and (iv) the distal phosphate binding site postulated to interact with the bis-phosphorylated reaction intermediate (Distal P loop : 523 - 535). The enzyme would be expected to contain two phosphate binding sites, one proximal and the other distal from S118, as the reaction intermediate consists of two phosphate groups. Both the proximal P loop incorporating the catalytic serine and the S loop are topologically equivalent, from domains I and III, linking β strands 3 and 4 from the conserved core.

The metal binding loop connects strands 2 and 3 in domain II whereas the distal P loop is contributed by domain IV (linking strands 6 and 7 of the β sheet). Thus, loops from all the four domains are strategically located to constitute the active site of the enzyme. The three dimensional conformations of these polypeptide segments and their location within the cleft correctly position residues involved in the catalytic activity and specificity of the protein. Of central importance in the proximal P loop is the catalytic serine (S118) which acts both as a donor and acceptor of the phosphate group followed by H119 and N120, which together constitute the catalytic triad commonly observed in serine proteases [24]. In LmPGM the electron density about S118 was consistent with its dephosphorylated state (**Fig. 6**).

Table 4: The list of residues constituting the active site of LmPGM. Abbreviations used: Proximal P: Proximal phosphate binding site; M: Metal binding site; S: Specificity loop; Distal P: Distal phosphate binding site.

Residues	LmPGM	OcPGM	PtPGM	PaPGM
Proximal P	117 ASHNPG-GPDADF 128	115 ASHNPG-GPNGD 125	123 LTASHNP-GGKEHGD 136	105 TGSHN-PPD 113
M	299 DGDADR 304	286 FDGD-GDRN 293	307 CDGDA-DRN 314	241 FDGD-GDRV 248
S	397 EESF 400	375 EESF 378	392 EESFG 396	324 GEM-SGH 329
Distal P	523 RLSGT-GSSGATIR 535	502 RLSGTG-SAGATIR 514	515 RLSGTG-CVGATIR 527	421 RASNTT-PVLVLR 432

Examination of the main chain dihedrals (ϕ,ψ) of the proximal P loop and comparison with the main chain conformations in the structures of OcPGM, PtPGM and PaPGM did not appear to indicate the conservation of any pronounced conformational feature. In contrast , the metal binding loop, consisting of the highly conserved motif DGD(A/G)DR was constrained to a fairly rigid three dimensional conformation with the main chain (ϕ_i,ψ_i)'s for the four previously mentioned structures (**Table 5**) in very good agreement (to within 20° in most cases).

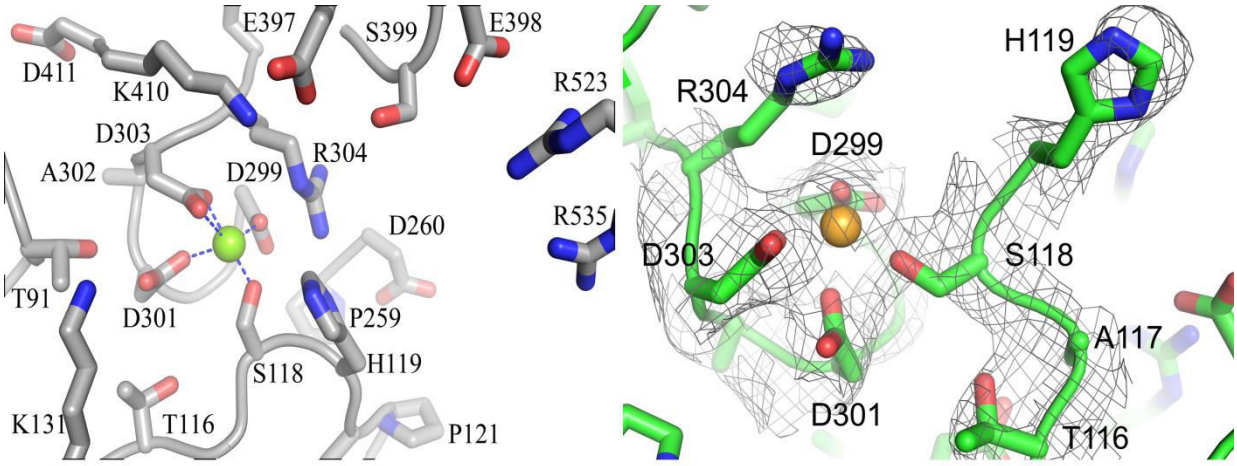


Figure 6: A view of the active site of *L.major* and $2F_o$-F_c map of the active site. (a) The active site of *L.major* containing the catalytic serine (S118), three aspartic acid residues (D299, D301, D303) in contact with the divalent cation Mg^{2+} (green sphere), the specificity loop (E397, S399) and the distal phosphate binding site (R523, R535). (b) The $2F_o$-F_c map of the active site of LmPGM, contoured at 1.0 σ; the cation Mg^{2+} is represented by the central orange sphere.

The conservation of the M-loop conformation is due to the stabilization of the loop geometry by the coordination of D299, D301 and D303 (LmPGM) with the Mg^{2+} ion. The catalytic serine S118 was found to be directed towards the metal binding loop with its O^{γ} in coordination with the divalent metal cation. Other highly conserved residues of the active site include E397, S399 of the S loop postulated to be one of the determinants of substrate specificity due to its interaction with the hydroxyls of the glucose ring. The location of the conserved residue W372 also turned out to be similar to other enzymes harbouring a carbohydrate binding site and therefore, could participate in stacking interactions with the glucosyl group of the substrate [24].

Table 5: The main chain dihedrals (φ, ψ in degrees) of the conserved metal binding site. Abbreviation used: LmPGM: *L. major* PGM, OcPGM: rabbit muscle PGM, PtPGM: *P. tetraurelia* PGM, PaPGM: *P. aeruginosa* PGM.

	LmPGM	OcPGM	PtPGM	PaPGM
ASP	-75.3, 171.7	-82.0, 177.1	-85.4, 179.1	-79.5, 178.2
GLY	-51.5, -46.4	-46.1, -49.0	-39.0, -57.5	-58.7, -31.9
ASP	-108.2, 11.5	-98.0, 13.1	-90.7, 10.5	-110.2, 1.1
ALA/GLY	71.7, 18.5	74.5, -6.7	59.5, 27.9	81.8, 9.8
ASP	-90.7, -18.5	-82.6, 7.7	-100.1, -8.9	-95.5, -1.1
ARG	-125.8, 161.7	-134.7, 154.6	-107.3, 156.7	-115.0, 152.7

In addition, residues R523, S525, T527 of the distal (P) phosphate binding site have been postulated to interact with the phosphate group of the incoming substrate.

3.4 Enzymatic Parameters of LmPGM

The purified enzyme was catalytically active in the absence of the activator glucose-1,6-bisphosphate. The kinetic constants K_m, V_{max} and K_{cat}/K_m were calculated to be 21.35±3.1 μM, 0.55±0.02 $\mu Ms^{-1}\mu M^{-1}$of LmPGM and 2.58X10^4 $M^{-1}s^{-1}$ respectively. The estimated molecular weight (63896 Da) of LmPGM obtained by *ExPASy-ProtParam* tool was used throughout the analysis. Addition of 0.1 μM glucose-1,6-bisphosphate resulted in almost twofold increase in V_{max} (1.02±0.06 μMs^{-1} μM^{-1} of LmPGM) and catalytic efficiency K_{cat}/K_m (4.52 X10^4 $M^{-1}s^{-1}$). The rephosphorylation of the dephospho enzyme by G-1,6-P_2 may have a stimulatory effect on the enzyme, resulting in an increase in its activity.

Table 6: A comparison of kinetic parameters of LmPGM and other reported PGMs towards the formation of G1P in presence of glucose-1,6-bisphosphate.

Enzymes	$K_{m(G1P)}$ μM	$(K_{cat}/K_m)_{G1P}$ $M^{-1}s^{-1}$
L. major PGM	21.4	4.5×10^4
T. brucei PMM	96.0	5.7×10^4
O. cuniculus PGM	63.0	9.0×10^7
P. aeruginosa PMM/PGM	20.0	1.0×10^5
T. gondii PGM	3.53	2.6×10^4

PGM: phosphoglucomutase; PMM: phosphomannomutase, which additionally shows PGM activity.

The K_m value of LmPGM in the direction of formation of G1P was comparable to that of PaPGM but lower than OcPGM (**Table 6**) [26, 27]. The results indicate that LmPGM requires almost one third of substrate concentration than that of its mammalian counterpart in order to achieve maximum velocity. The catalytic efficiency of LmPGM was similar to *T. brucei* phosphomannomutase which additionally shows PGM activity and one of the isoforms of PGM (PGM II) from *T.gondii* but much lower than OcPGM [5, 28, 29].

3.5 Conformational variability of domain IV

Normal mode analysis (NMA) has emerged as a powerful tool in studying various dynamical properties of proteins such as conformational changes upon ligand binding, opening and closing of membrane channels, modeling the dynamics of ribosome function and viral capsid maturation [19]. For most proteins, superposition of about two low frequency normal modes and examination of their perturbed amplitudes along the direction of the modes generally provides information with regard to the functionally relevant motion of the proteins [30]. For proteins exhibiting low frequency collective motion, a single low frequency normal mode maybe

sufficient to indicate the crucial conformational flexibility responsible for the progress of the catalytic reaction.

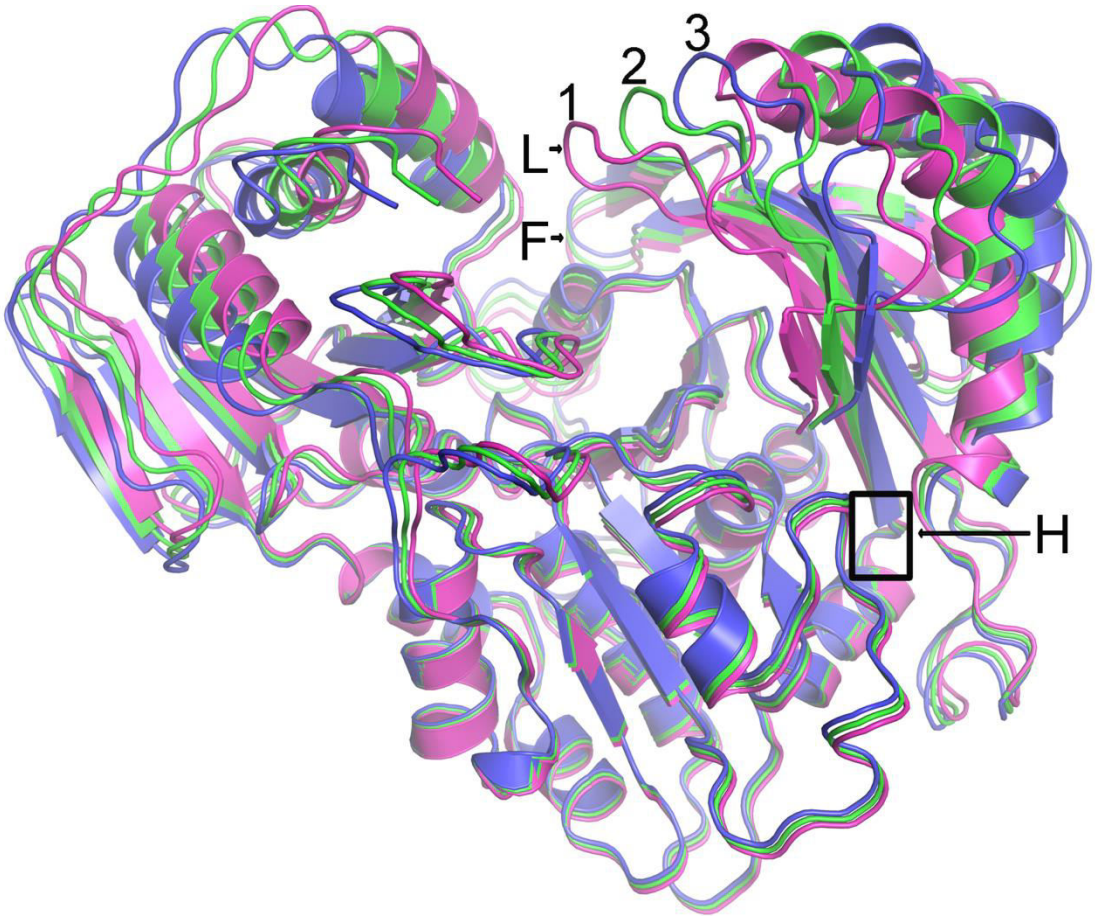


Figure 7: The conformational variability of domain IV approximated by normal mode analyses: 1. "closed" LmPGM (model-11): colored pink, 2. "semi-open" enzyme (model-6): colored green and 3. "open" (model-1) enzyme colored blue. Abbreviations used: L: latch, F: extended flap, H: hinge region.

For the lowest frequency mode (Mode 7), a total of 11 perturbed models of LmPGM were generated by elNémo. All normal mode analyses were performed with molecule A, lacking 25 residues from the N-terminal region. Model-1 appeared to adopt a more open and model -11 a relatively closed conformation compared to the original crystal structure. The crystal structure

(Model 6: C_α RMSD = 0.0 with respect to subunit A of the crystal structure) had also the closest resemblance to the dephospho-enzyme-intermediate complex of rabbit (C_α RMSD = 0.818; PDB code: 1C47; Baranidharan *et al*., unpublished work) which is a semi-open conformation to facilitate the reorientation of the intermediate prior to the transfer of the alternate phosphoryl group back to the catalytic serine. Therefore, model-1, 6 (crystal structure), and 11 probably spans the open, semi-open and the closed conformation adopted by the enzyme respectively (**Fig. 7**).

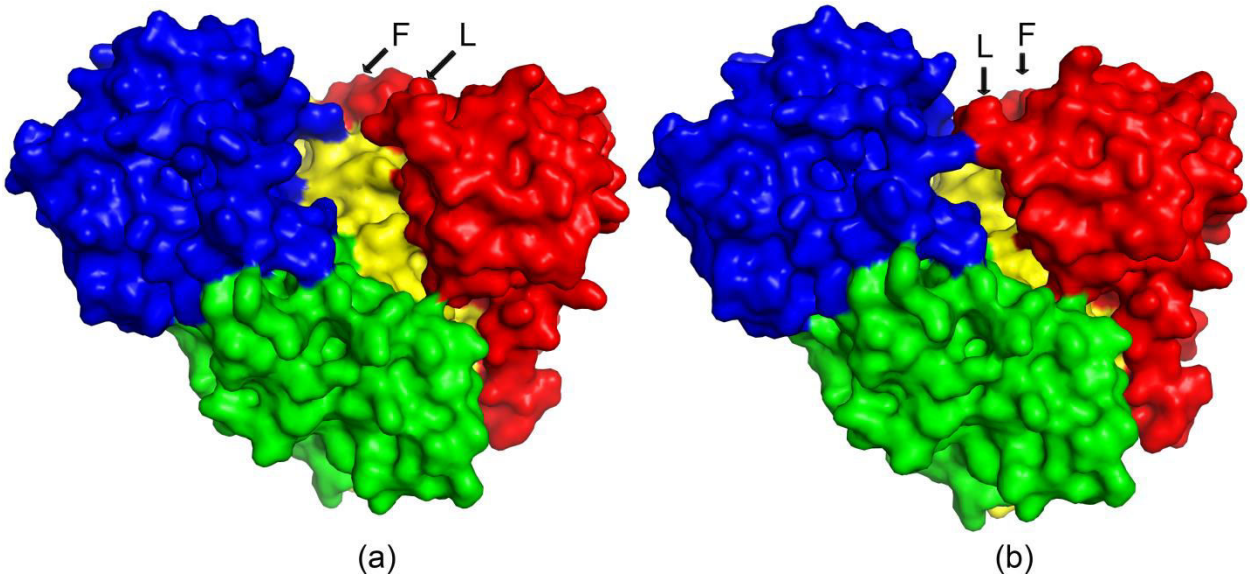


Figure 8: The closure of the active site modeled by normal mode analyses. (a) "open" (model-1) enzyme (b) "closed" LmPGM (model-11). The four domains are colored blue to red. The latch (L) and the extended flap (F) region are highlighted.

Leishmania spp. exhibited a high degree of sequence conservation for the residues of the hinge region (**Supplementary Fig. S1**). Compared to other domains, domain IV exhibited maximum conformational flexibility via the hinge joint. The least affected were domains II and

III. The domain motion analysis by the program *DynDom* [31] revealed that the switching between the "closed" and "open" forms of LmPGM involved ~10° rotation of the fourth domain. The motion of the fourth domain causes the tip of the loop **L** to move towards the central cleft by ~ 10Å and together with flap **F** actuates the closure of the lid over the active site (**Fig. 7, 8; Supplementary movie -** provided separately with the CD-ROM). Examination of the main chain dihedrals of the hinge region for the 'open' and 'closed' forms of the enzyme reveals maximum changes in the (ϕ,ψ)'s in residues (Ala445, Asp456, Tyr457) which lie at the extremities of the hinge bending loop (**Table 7**). On the other hand the flap F appears to be relatively rigid not exhibiting any significant changes in its main chain conformation.

Table 7: The main chain dihedrals (φ, ψ in degrees) of the hinge-region-residues of the open, semi-open and closed LmPGM respectively, previously generated by normal mode analyses.

	Open	Semi-Open	Closed
ALA-445	-76.3, -27.7	-72.5, -27.8	-66.8, -27.9
THR-446	-89.4, -49.8	-89.4, -49.8	-89.3, -49.8
TYR-447	-103.1, -8.7	-103.1, -8.0	-103.1, -6.5
GLY-448	92.7, 165.2	91.0, 165.2	88.2, 165.2
ARG-449	-109.0, 138.6	-109.0, 138.6	-109.1, 138.7
ASN-450	-100.8, 100.3	-100.9, 103.1	-100.9, 106.4
TYR-451	-67.5, 126.7	-69.6, 126.8	-71.6, 126.7
TYR-452	-127.7, 143.3	-127.8, 143.4	-127.6, 143.3
SER-453	-165.6, 163.9	-165.6, 162.0	-165.6, 160.3
ARG-454	-140.4, 129.0	-141.0, 129.2	-142.3, 128.8
TYR-455	-102.4, 136.9	-102.8, 137.0	-102.2, 136.8
ASP-456	-122.5, 116.8	-122.5, 121.8	-122.3, 128.7
TYR-457	-102.8, 99.6	-107.8, 99.8	-114.7, 99.6

The results obtained by normal mode analysis were in close agreement with the domain motion observed in other PGMs. The "semi-open" form of the enzyme (model-6) probably

allows for the required ~180° rotation of the intermediate and the placement of the alternative phosphate group close to the dephosphorylated-S118, enabling phosphoryl transfer. The motion of the fourth domain is critical from the functional point of view, since the residues interacting directly with the incoming substrate are properly positioned only when LmPGM adopts a closed conformation [31].

Studies indicate that domain motion resulting from substrate binding may have evolved under selective pressure, opting for paths that offer relatively low energy barriers [19]. There are instances of hierarchical organization of domain motions within some superfamilies (e.g. Ras-superfamily) where the primary motion associated with the switching function (switch between the active and inactive state of the enzyme) are retained by all the members of the superfamily. Additional family specific motions play a crucial role in the determination of unique functional characteristic of the individual members [32]. Conformational motions were found to be conserved not only within the phosphoglucomutase family but also for the entire superfamily of α-D-phosphohexomutases [31]. Being a member of the phosphohexomutase superfamily, the variability of domain IV modeled by NMA, could be crucial for the proper functioning of LmPGM.

3.6 Modelling of the LmPGM-substrate/intermediate complex

The rabbit and the bacterial PGMs (especially *P.aeruginosa*), are the most structurally and functionally well characterized enzymes. A detailed study of the reported enzyme-substrate/ intermediate/product complexes indicate that, the phosphorylated enzyme, upon binding the substrate adopts a closed conformation, essential for the proper positioning of the substrate within the active site. After the transfer of the phosphoryl group from the catalytic serine to the

substrate, the enzyme adopts a semi open conformation, thereby facilitating the ~180° rotation of the intermediate, glucose-1,6-bisphosphate, to reposition the alternate phosphate group close to the catalytic serine without exposing it completely to the solvent. Once again the enzyme adopts a closed conformation, facilitating the successive transfer of the alternate phosphate group of the intermediate back to the catalytic serine leading to the formation of the product [33].

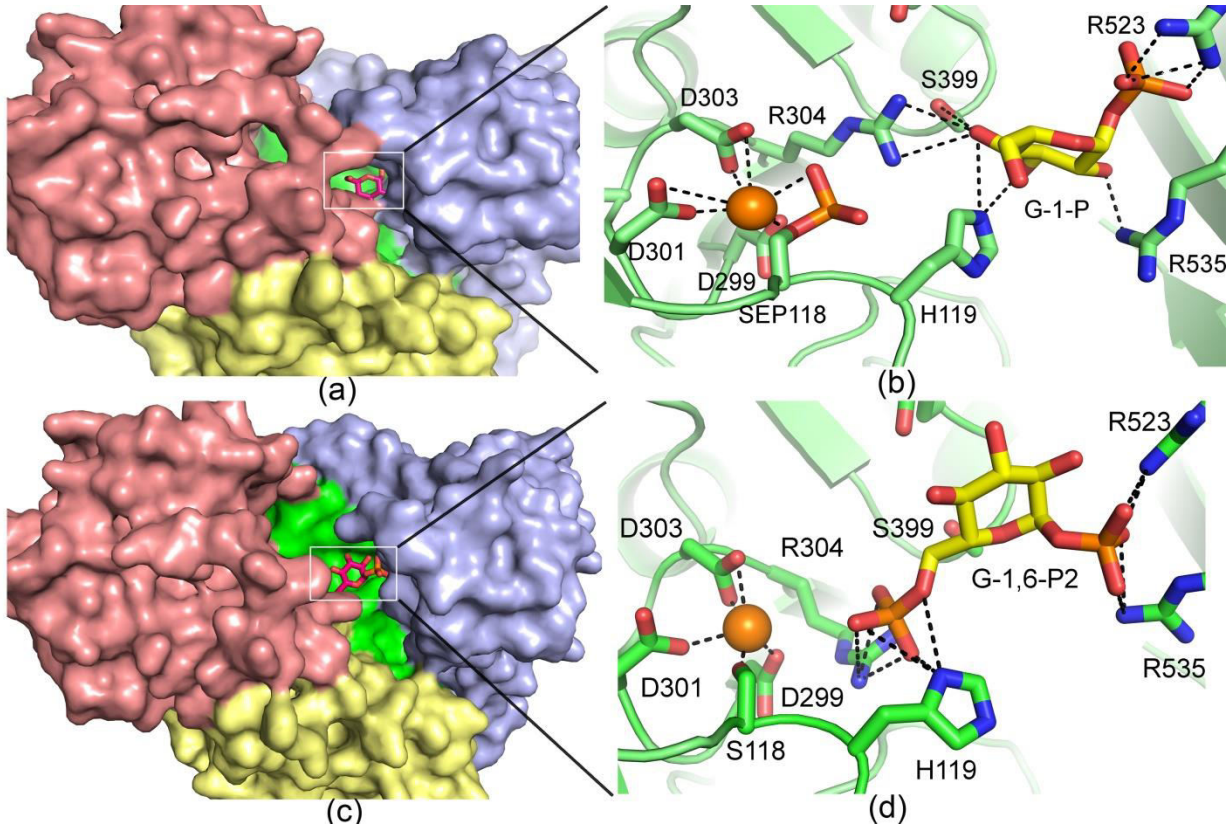


Figure 9: The models of the enzyme-substrate/intermediate complex generated using *AutoDock*. (a) The substrate glucose-1-phosphate modeled to the active site of "closed" LmPGM. The four domains of the enzyme are colored red to blue. (b) A close up view of the active site. The cation Mg^{2+} is colored orange. (c) The reaction intermediate, glucose-1,6-bisphosphate docked to the active site of "semi-open" LmPGM. The four domains are colored red to blue. (d) A close up view of the active site. The cation Mg^{2+} is colored orange.

The substrate gucose-1-phosphate and the intermediate glucose-1, 6-bisphosphate were docked into the active site of the "closed" and "semi-open" enzymes respectively, generated previously by normal mode analysis. The docking solutions for both the substrate and the intermediate were found to be buried deeply within the cleft with only 18% and 25% of the respective total surface area, accessible to the solvent.

The enzyme adopts a close conformation upon substrate binding, resulting in an overall decrease in the volume of the active site (**Fig. 9*a***) [33]. The rotation of the fourth domain and the subsequent lid closure due to the repositioning of the loop **L** and the flap **F** almost shields the substrate from solvent molecules. In this case the substrate was properly positioned to accept a phosphoryl group from the catalytic serine (**Fig. 9*b***). The intermediate phosphorylated catalytic serine SEP118 was in contact with the divalent cation Mg^{2+} which in turn, was found to interact with the three aspartic acid residues D299, D301, D303 from the metal binding loop. The residues R523 and R535 of domain IV were found to interact with the phosphate group of the incoming substrate as was previously observed in other bacterial PGM-substrate complexes [21]. The residue S399, reported to be responsible for the specificity of the enzyme towards its substrate [11], was found to interact with the hydroxyl group of the glucose ring. The other two residues R304 and H119 were also found to be situated in the vicinity of the substrate.

The model of the enzyme-intermediate complex shows that the phosphate attached to C1 hydroxyl group of glucose retains its contact with the two arginine residues R523 and R535 (**Fig. 9*d***), which were previously reported to be essential for the proper positioning of the substrate and intermediate. The alternate phosphate group of the intermediate was present in the vicinity of S118 and other catalytic residues. The semi-open conformation adopted by the enzyme facilitates the reorientation of the intermediate (**Fig. 9*c***) [34]. Studies indicate that the ~180° reorientation

of the intermediate occurs without its release from the enzyme active site [33]. It may, therefore, be possible that the partial opening of the loop **L** prevents the escape of the intermediate by retaining the key contacts between the intermediate and the residues S525 and T527 at the same time provides enough room for the reorientation of the intermediate.

3.7 Prospects for inhibitor design

Although conventional drug designing approaches generally target the functional site of the pathogenic enzyme other alternative approaches could also be adopted. For example, binding sites other than the native active site could be targeted such as protein-protein interaction sites, allosteric inhibition or exploitation of protein packing defects [35, 36, 37]. Other approaches include the design of drugs which involves the binding of two tethered partner molecules (which constitutes the drug) to two separate pockets resulting in an overall increase in total binding affinity [35]. To explore such possibilities, the program *CASTp* [38] was used to find surface pockets outside the functional site of LmPGM, as the native catalytic site of the molecule exhibits a high degree of sequential and structural conservation with respect to its mammalian homologue.

One of the pockets found by the program was located close to the hinge region, at the juncture of domains II, III and IV, involving residues from all the three domains. Residues from the two β strands of domain II (IIβ3, IIβ4); one helix (IIIα1), the turn (IIIT2) linking the first β-strand and third helix, loop S and the hinge region H of domain III; the two strands (IVβ1, IVβ8) and one helix (IVα2) of domain IV contributed to the formation of the pocket. The walls of the pocket were lined by two insertions unique to the parasite (domain II : 275-288; insrt2 - LLPDGNANPAMKHI & domain IV : insrt6 - 546-552; TVKSHLA) which however appeared

to adopt multiple conformations in the four molecules present in the asymmetric unit. Some variability in case of the pocket was also observed in features such as solvent accessible surface area (~ 240-270 $Å^2$), molecular surface area (~ 496-573 $Å^2$) and the volume (~90-102 $Å^3$) in the 'closed' and 'open' forms of the enzyme generated by NMA. The transient opening of pockets have been observed in other enzymes during molecular dynamic simulation and some of them were reported to bind small molecules [39].

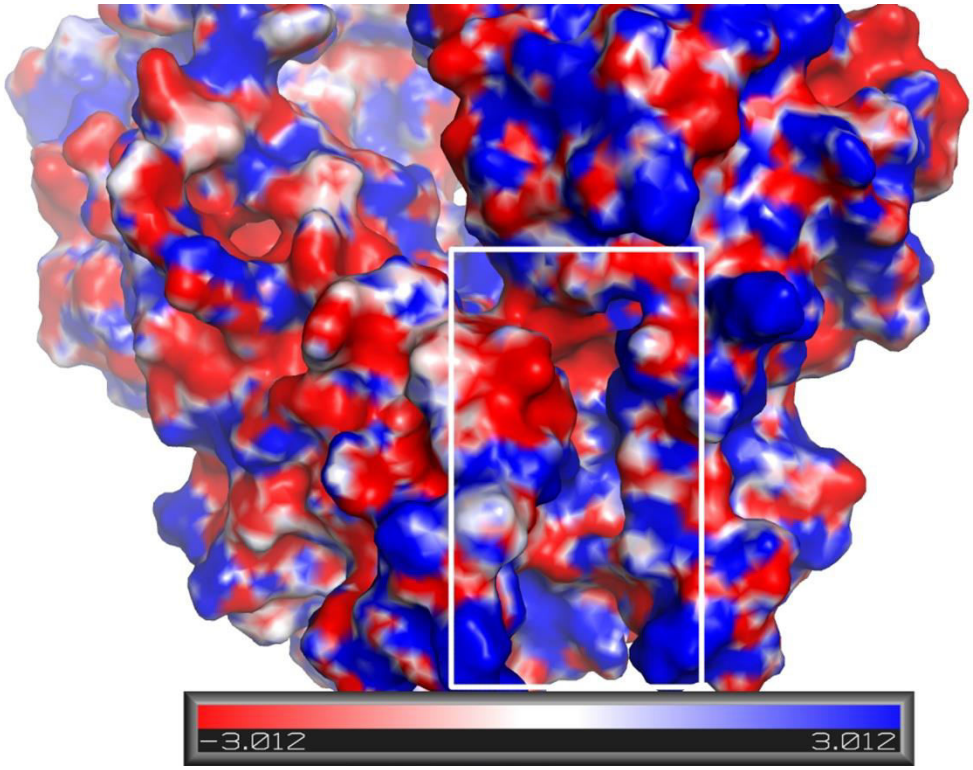


Figure 10: The surface electrostatic potentials of the enzyme (+3 kcal/(mol.e) in blue to -3 kcal/(mol.e) in red). The pocket is highlighted by the rectangular box.

Although ligands can stably bind to surfaces with area larger than 600 $Å^2$ by hydrophobic interactions alone, additional interaction features like surface roughness becomes critical for smaller sites [35]. Surface roughness allows more Van der Waals contacts within a small volume

and facilitates highly specific interactions. Thus along with CASTp , another program *HotPatch* [40] was used with two individual properties: concavity and surface roughness. The program finds surface patches of unusual physicochemical properties present on the surface of the protein. Similar results were obtained using either concavity or surface-roughness features with the highest scoring residue patches located in the close proximity of the pocket region and four residues from the patches turned out to be pocket-residues (P316, E397, E398, F400).

The residues lining the pocket differed substantially from its mammalian counterpart, owing to the shortening of the two loops in the latter. The pocket contained both polar and hydrophobic residues. The electrostatic potential map generated by the program *PBEQsolver* [41] indicated the presence of a negatively charged region at the top of the pocket (**Fig. 10**), while the bottom was lined with three hydrophobic residues (W444, F312 and F313).

Inhibitors with clinical utility can only be designed by selectively targeting the pathogenic enzyme. Since the mammalian and the pathogenic phosphoglucomutase share almost identical active site features, therefore the pocket located in the proximity of the hinge region and the juncture of domain II-III- IV may serve as an alternative target site. The pocket exhibits features believed to be critical for the stable binding of small molecules (e.g. surface roughnes, electrostatic potential, hydrophobicity, concavity). A small molecule, appropriately designed to bind to the pocket may restrict the conformational flexibility and prevent the proper functioning of the enzyme.

4. Conclusion

The structure of the enzyme phosphoglucomutase from the pathogenic organism *Leishmania major* was solved by molecular replacement at 3.5Å resolution, with four molecules (A,B,C,D)

in the asymmetric unit , of which the geometry between subunits A,D and B,C were practically identical , due to crystal packing, as the molecule is a monomer in solution. The overall features of the four constituent domains and the active site were fairly well conserved between the LmPGM and the homologous enzyme from rabbit and the variable conformational states of the fourth domain postulated to be essential for catalytic activity was analysed by means of normal mode analysis. Despite the fact that the active site of the leishmanial enzyme is conserved with respect to the mammalian enzyme from rabbit, there could be alternative inhibitor binding sites, in the proximity of the hinge region which show several surface features unique to the parasite.

5. Protein data bank accession code

The coordinates and the structure factor data of LmPGM have been deposited in the RCSB Protein Data Bank (PDB accession code 4QG5).

References

1. Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P, Cano J, Jannin J, Boer MD: **Leishmaniasis worldwide and global estimates of its incidence.** *PLoS One*. 2012, **7**(5): e35671 (doi:10.1371/journal.pone.0035671).

2. Markle WH, Makhoul K: **Cutaneous Leishmaniasis: Recognition and Treatment**. *American Family Physician* 2004 , 69(6): 1455-1460

3. Penha LL, Sant'Anna CB, Mendonça-Previato L, Cunha-e-Silva NL, Previato JO, Lima AP: **Sorting of phosphoglucomutase to glycosomes in Trypanosoma cruzi is mediated by an internal domain.***Glycobiology* 2009, **19**(12): 1462-1472.

4. Penha LL., Mendonça-Previato L, Previato JO, Scharfstein J, Heise NC, De A, Lima AP: **Cloning and characterization of the phosphoglucomutase of Trypanosoma cruzi and functional complementation of a Saccharomyces cerevisiae PGM null mutant.** *Glycobiology* 2005, **15**(12): 1359-1367.

5. Bandini G, Mariño K, Güther MLS, Wernimont AK, Kuettel S, Qiu W, Afzal S, Kelner A, Hui R, Ferguson MAJ: **Phosphoglucomutase is absent in Trypanosoma brucei and redundantly substituted by phosphomannomutase and phospho-N-acetylglucosamine mutase.** *Molecular Microbiology* 2012, **85**(3): 513-534.

6. Quick CB, Fisher RA., Harris H: **A Kinetic Study of the Isozymes Determined by the Three Human Phosphoglucomutase Loci *PGM_1, PGM_2* and PGM_{3L}.** *Eur. J. Biochem.* 1974, **42**: 511-517.R

7. Azmi K, Schonian G, Schnur LF, Nasereddin A, Ereqat S, Abdeen Z: **Development of assays using hexokinase and phosphoglucomutase gene sequences that distinguish strains of Leishmania tropica from different zymodemes and microsatellite clusters and their application to Palestinian foci of cutaneous leishmaniasis.** *PLOS Neglected Tropical Diseases* 2013, **7**(9): e2464.

8. Bradford MM: **A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding**. *Anal. Biochem.* 1976, **72**: 248–254.

9. Kabsch W: **XDS**. *Acta Cryst.* D 2010, **66**: 125-132.

10. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS: **Overview of the CCP4 suite and current developments.** *Acta Cryst.* D 2011, **67**: 235-242.

11. Liu Y, RayJr WJ, Baranidharan S: **Structure of Rabbit Muscle Phosphoglucomutase Refined at 2.4 Å Resolution.** *Acta Cryst.* D 1997, **53**: 392-405.

12. Stein N: ***CHAINSAW*: a program for mutating pdb files used as templates in molecular replacement.** *J. Appl. Cryst.* 2008, 41: 641-643.

13. McCoy AJ., Grosse-Kunstleve RW., Adams PD, Winn MD, Storoni LC, Read RJ: **Phaser crystallographic software.** *J Appl Cryst.* 2007, 40: 658-674.

14. Adams PD, Grosse-Kunstleve RW, Hung LW, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC: **PHENIX: building new software**

for automated crystallographic structure determination. *Acta Cryst*. D 2002, **58**: 1948-1954.

15. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD: **Towards automated crystallographic structure refinement with phenix.refine.** *Acta Cryst.* D 2012, **68**: 352-367.

16. Emsley P, Cowtan K: **Coot: model-building tools for molecular graphics.** *Acta Cryst.* D 2004, 60: 2126-2132.

17. Chen VB, Arendall III WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC: **MolProbity: all-atom structure validation for macromolecular crystallography.** *Acta Cryst.* D 2010, **66**: 12-21.

18. DeLano WL: **The PyMOL molecular visualization system.** 2002. http://www.pymol.org.

19. Suhre K, Sanejouand YH: **ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement.** *Nucleic Acids Res.* 2004: W610-4. (http://www.sciences.univ-nantes.fr/elnemo).

20. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ: **AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility**. *J. Computational Chemistry* 2009, **16**: 2785-2791.

21. Regni C, Naught L, Tipton PA, Beamer LJ: **Structural basis of diverse substrate recognition by the enzyme PMM/PGM from *P.aeruginosa.*** *Structure* 2004, **12**: 55-63.

22. Luebbering, EK, Mick J, Singh RK, Tanner JJ, Mehra-Chaudhary R, Beamer LJ: **Conservation of functionally important global motions in an enzyme superfamily across varying quaternary structures.** *J. Mol. Biol*. 2012, **423**(5): 831-846.

23. Schramm AM, Mehra-Chaudhary R, Furdui CM., Beamer LJ; **Backbone Flexibility, Conformational Change, and Catalysis in a Phosphohexomutase from *Pseudomonas aeruginosa.*** *Biochemistry* 2008, **47**: 9154-9162.

24. Müller S, Diederichs K, Breed J, Kissmehl R, Hauser K, Plattner H, Welte W: **Crystal structure analysis of the exocytosis-sensitive phosphoprotein, pp63/parafusin (phosphoglucomutase), from Paramecium reveals significant conformational variability.** *J Mol Biol.* 2002, **315**(2): 141-153.

25. Mehra-Chaudhary R, Mick J, Tanner JJ, Beamer LJ: **Quaternary structure, conformational variability and global motions of phosphoglucosamine mutase**. *FEBS Journal*. 2011, 278: 3298-3307.

26. Naught LE, Tipton PA: **Kinetic mechanism and pH dependence of the kinetic parameters of Pseudomonas aeruginosa phosphomannomutase/ phosphoglucomutase.** *Archives of Biochemistry and Biophysics* 2001, **396**(1): 111-118.

27. Joshi JG, Handler P: **Phosphoglucomutase I: purification and properties of phosphoglucomutase from Escherichia coli.** *J. Biol. Chem.* 1964, **239**: 2741-2751.

28. Imada M, Kawashima S, Kanehisa M, Takeuchi T, Asai T: **Characterization of alpha-phosphoglucomutase isozymes from Toxoplasma gondii.** *Parasitology International* 2010, 59: 206-210.

29. Ray WJJr., Post CB, Puvathingal JM: **Reaction of the isosteric methylenephosphonate analog of alpha-D-glucose 1-phosphate with phosphoglucomutase. Induced-fit specificity revisited.** (1993). *Biochemistry*. **32**(1), 38-47.

30. Krebs WG, Alexandrov V, Wilson CA, Echols N, Yu H, Gerstein M: **Normal mode analysis of macromolecular motions in a database framework: developing mode concentration as a useful classifying statistic.** *Proteins: Structure, Function, and Genetics* 2002, 48: 682-695.

31. Hayward S, Berendsen HJC: **Systematic Analysis of Domain Motions in Proteins from Conformational Change; New Results on Citrate Synthase and T4 Lysozyme**. *Proteins: Structure, Function and Genetics* 1998, **30:** 144-154.

32. Wynsberghe AWV, Cui Q: **Conservation and variation of structural flexibility in protein families.** *Structure* 2010, 18: 281-283.

33. Regni C, Schramm AM, Beamer LJ: **The reaction of phosphohexomutase from Pseudomonas aeruginosa: structural insights into a simple processive enzyme.** *J Biol Chem.* 2006, **281**(22): 15564-71.

34. Naught LE, Tipton PA: **Formation and reorientation of glucose 1,6-bisphosphate in the PMM/PGM reaction: transient-state kinetic studies.** *Biochemistry* 2005, 44(18): 6831-6.

35. Pettit FK, Bowie JU: **Protein surface roughness and small molecular binding sites.** *J Mol Biol*. 1999, **285**: 1377-1382.

36. Scheer JM, Romanowski MJ, Wells JA: **A common allosteric site and mechanism in caspases.** *Proc Natl Acad Sci* 2006, 103(20): 7595-7600.

37. Fernández A: **Incomplete Protein Packing as a Selectivity Filter in Drug Design.** *Structure* 2005, 13(12): 1829-1836.

38. Dundas J, Ouyang Z, Tseng J, Binkowski A, Turpaz Y, Liang J: **CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues.** *Nucleic Acid Research* 2006, 34: W116-W118.

39. Wells JA., McClendon CL: **Reaching for high-hanging fruit in drug discovery at protein–protein interfaces.** *Nature* 2007, **450:** 1001-1009.

40. Pettit FK, Bare E, Tsai A, Bowie JU: **HotPatch: A Statistical Approach to Finding Biologically Relevant Features on Protein Surfaces**. *J Mol Biol.* 2007, **369**(3): 863–879.

41. Jo S, Vargyas M, Vasko-Szedlar J, Roux B, Im W: **PBEQ-solver for online visualization of electrostatic potential of biomolecules.** *Nucl. Acids Res.* 2008, **36**: W270-275.

Supplementary materials:

Supplementary Table 1: The structural element types and their designation.

Structural elements	Type	Residues	Designation	Structural elements	Type	Residues	Designation
1	lp	25-30	IL1	41	tn	329-331	IIIT1
2	α	31-41	Iα1	42	α	332-335	IIIα2
3	ex-lp	42-51	IXL1	43	lp	336-343	**Insrt4**
4	β	52-55	Iβ1	44	β	344-347	IIIβ1
5	lp	56-62	IL2	45	tn	348-353	IIIT2
6	α	63-76	Iα2	46	α	354-359	IIIα3
7	tn	77-80	IT1	47	tn	360-364	IIIT3
8	β	81-84	Iβ2	48	β	365-368	IIIβ2
9	lp	85-90	IL3	49	tn	369-371	IIIT4
10	α	91-98	Iα3	50	α	372-379	IIIα4
11	ex-lp	99-111	**Insrt1**	51	ex-lp	380-393	**Insrt5**
12	β	112-116	Iβ3	52	β	394-397	IIIβ3
P	**ex-lp**	**117-128**	**P**	**S**	**hp-F**	**398-400**	**S**

14	β	129-133	Iβ4	54	β	401-404	IIIβ4
15	lp	134-141	IL4	55	lp	405-411	IIIL1
16	α	142-153	Iα4	56	α	412-426	IIIα5
17	ex-lp	154-172	**Insrt2**	57	ex-lp	427-435	**Insrt6**
18	β	173-177	Iβ5	58	α	436-446	IIIα6
19	hp-tn	178-181	IHP1	**H**	**tn**	**447-449**	**H**
20	β	182-187	Iβ6	60	β	450-457	IVβ1
21	lp	188-190	IIL1	61	lp	458-461	IVL1
22	α	191-200	IIα1	62	α	462-477	IVα1
23	tn	201-202	IIT1	63	ex-lp	478-490	IVXL1
24	α	203-210	IIα2	64	β	491-494	IVβ2
25	lp	211-216	IIL2	65	tn	495	IVT1
26	β	217-220	IIβ1	66	β	496-498	IVβ3
27	tn	221-225	IIT2	**F**	**hp-tn**	**499-501**	**F**
28	α	226-238	IIα3	68	β	502-505	IVβ4
29	tn	239-241	IIT3	69	lp	506-507	IVL1

30	α	242-244	IIα4	70	β	508-513	IVβ5
31	ex-lp	245-266	IIXL1	71	hp-flap	514-518	IVHF1
32	α	267-273	IIα5	72	β	519-525	IVβ6
33	ex-lp	274-293	**Insrt3**	**L**	**hp-flap**	**526-532**	**L**
34	β	294-298	IIβ2	74	β	533-542	IVβ7
M	**hp-F**	**299-304**	**M**	75	ex-lp	543-559	**Insrt7**
36	β	305-309	IIβ3	76	α	560-574	IVα2
37	hp-tn	310-311	IIHP1	77	tn	575	IVT2
38	β	312-314	IIβ4	78	α	576-580	IVα3
39	tn	315-316	IIT4	79	lp	581-585	IVL2
40	α	317-329	IIIα1	80	β	586-589	IVβ8

The abbreviation used are: **lp** : Loop, **α** : α-helix, **β** :β-strand, **ex-lp** : Extended loop, **tn** : turn, **hp-tn** : hair pin-turn, **hp-flap** : hairpin flap, **P** : Phosphate binding site, **M** : Metal binding site, **S** : Specificity loop, **L** : Latch region, **H**: Hinge region, **F**: anti-parallel flap.

Figure S1: Sequence conservation of the residues of the hinge region. The conserved residue G448, predicted to serve as the hinge joint is highlighted. The multiple sequence alignment of PGMs was performed with *T-coffee*.

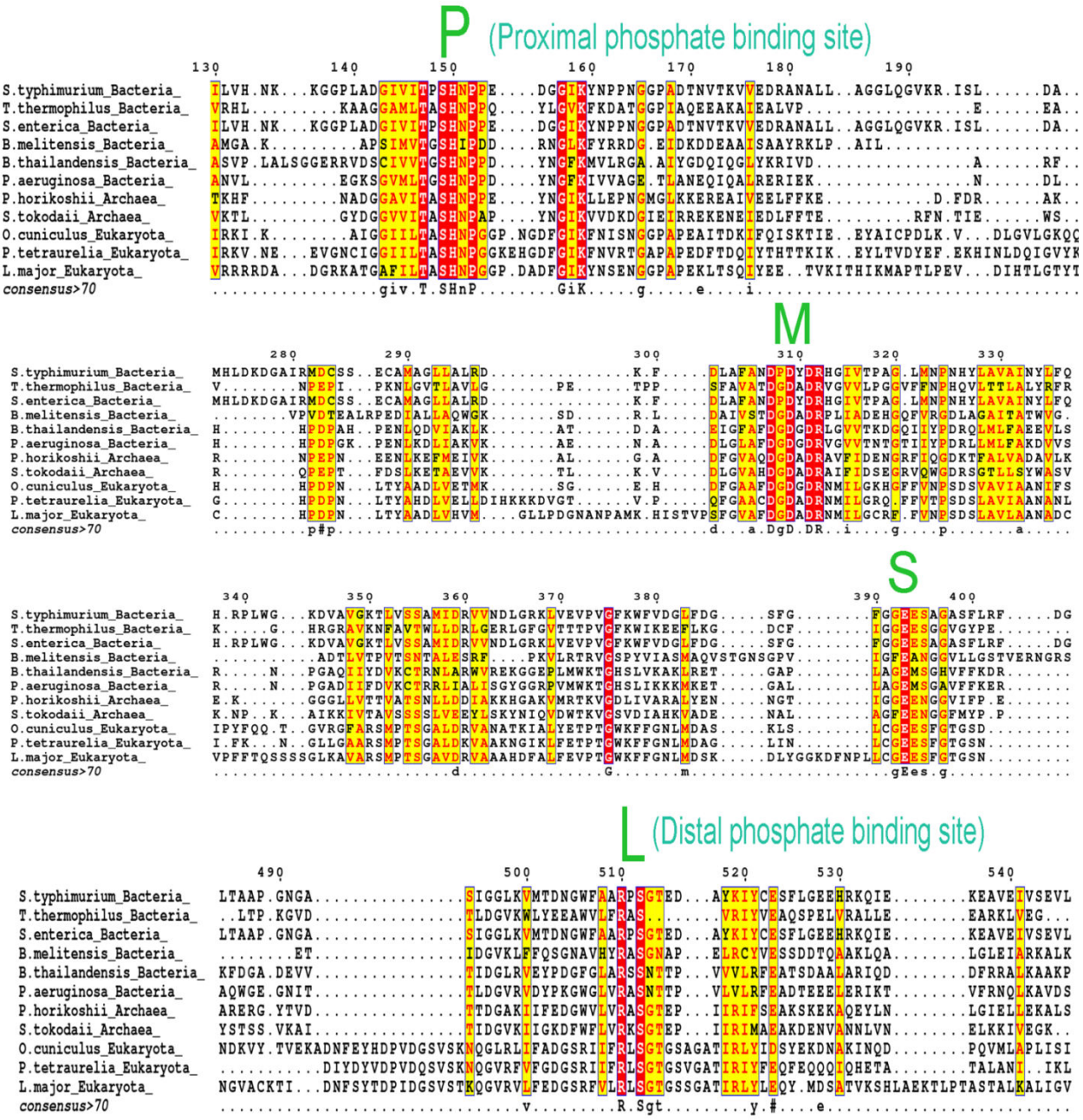


Figure S2: The sequence alignment of structurally characterized PGMs using *T-coffee* followed by the generation of the output file by *ESPript 3.0*. The conserved key residues are highlighted. Abbreviation used: P – proximal phosphate binding site, M – metal binding site, S – specificity loop, L: latch/ distal phosphate binding site.

Appendix

1. Purification and crystallization trials of seven enzymes from several human pathogens.

1.1 ATP-dependent DNA helicase II UvrD1from *M. smegmatis*

The pathogen *Mycobacterium tuberculosis* causes TB in human while *Mycobacterium smegmatis* is a non-virulent sister species of the pathogenic parasite. In mycobacteria, the enzyme UvrD1 functions as a DNA dependent ATPase and a Ku dependent DNA helicase that translocates along the single stranded DNA and unwinds the double strand during DNA repair and recombination [1].

The recombinant plasmid construct was obtained from Dr. K. M. Sinha, Institute of Molecular Medicine. The plasmid construct containing the UvrD1 gene (2352 bp) from *M. smegmatis* was transformed into *E. coli* Bl-21(DE3) cells followed by overnight incubation on LB-agar/kanamycin plates at 37° C. A single colony was used to inoculate the LB media containing kanamycin and incubated at 37°C until OD_{600} attained a value of ~0.6. IPTG (isopropyl-β-D-thiogalactopyranoside) was then added to a final concentration of 0.5mM, followed by overnight incubation. The cells were harvested by centrifugation at 5000g for 15 minutes and suspended in lysis buffer [50mM Tris-HCl, pH 7.5, 250mM NaCl, 10% sucrose, 2 mM phenyl methane sulphonyl fluoride (PMSF) and 1X protease inhibitor cocktail]. Lysozyme was added to reach a final concentration of 200µg/ml. Subsequent to sonication, the homogenized lysate was centrifuged at 14000g for 45 minutes and the supernatant was loaded onto a pre-equilibrated Ni-NTA column. The column was washed with 50mM Tris-HCl, pH 8.0,

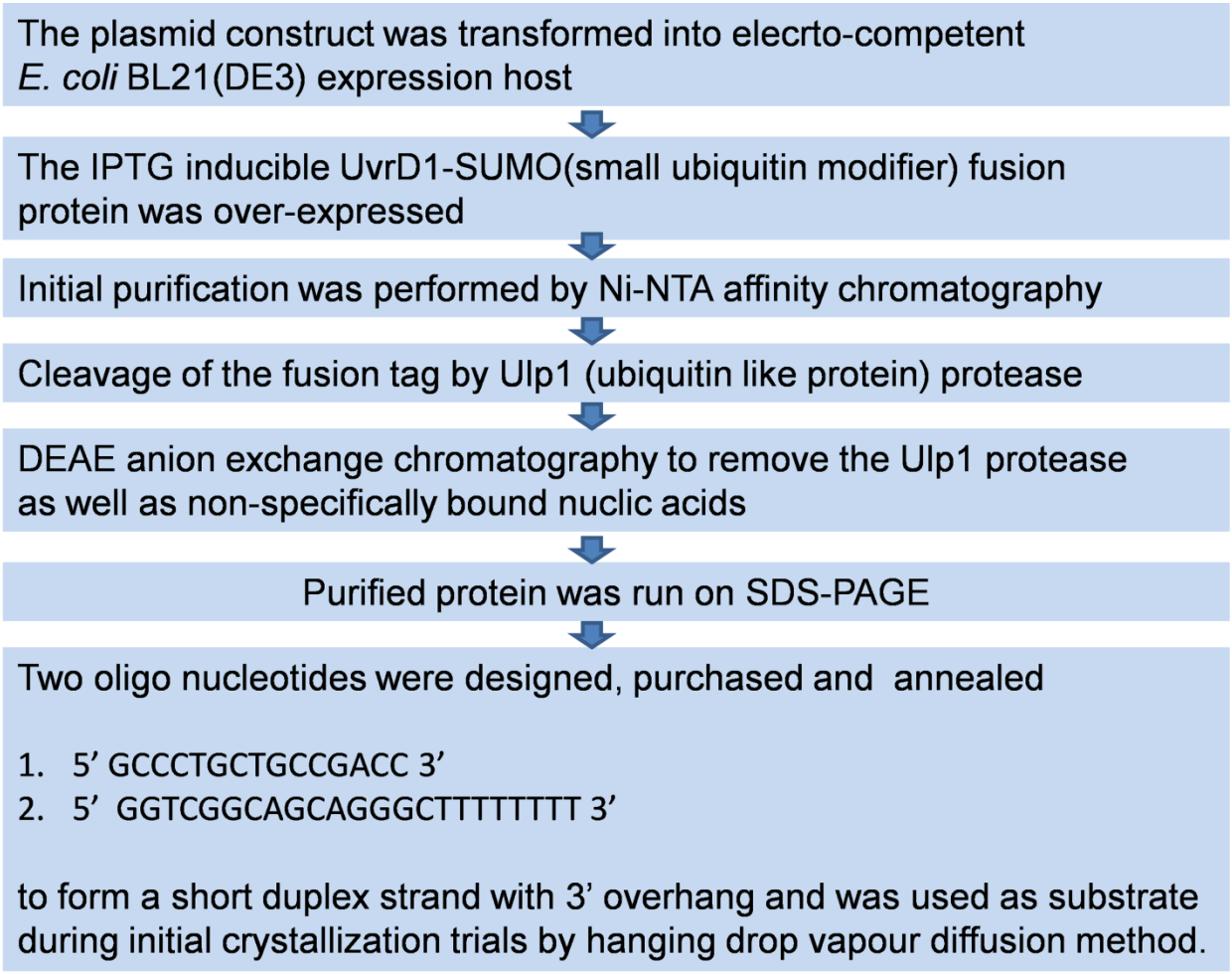


Figure 1: A flowchart of the purification steps of the enzyme UvrD1 from *M. smegmatis*.

250 mM NaCl, 5% glycerol followed by another wash with 20mM imidazole. The protein was eluted with 250mM imidazole. The protein containing fractions were pooled and dialysed against 50mM Tris-HCl, pH 8.0, 250mM NaCl, 5% glycerol, 1 mM EDTA. The protein was then loaded onto pre-equilibrated DEAE anion exchange column and the protein was recovered in the flowthrough fractions. In order to remove the leader sequence (His_{10} tag) the Ulp1 protease was added at a fixed ratio of UvrD1:Ulp1=100:1. The reaction mixture was dialyzed against 50mM

Tris-HCl, pH 7.5, 150mM NaCl, 5% glycerol, 0.02% NaN_3 and was applied to the pre-equilibrated Ni-NTA column. UvrD1 was recovered in the flowthrough fractions and the enzyme concentration was estimated by Bradford method. A flowchart of the purification steps is shown in **Fig. 1**. The purified enzyme was then run on SDS-PAGE and was found to exhibit a single band of expected molecular weight (**Fig. 10*a***).

Two oligo nucleotides were designed (i) 5' GCCCTGCTGCCGACC 3' (ii) 5' GGTCGGCAGCAGGGCTTTTTTTT 3' and purchased. They were annealed to form a short duplex strand with 3' overhang and were used as substrate-mimic during initial crystallization setups. Initial crystallization screens were set up at 20°C using the hanging drop vapour diffusion method. Drops containing 7-9µl of protein solution at a concentration of 10mg/ml and 1-3 µl of crystallization buffer were sealed over a 1 ml reservoir.

1.2 The non-homologous end joining protein Ku from *Mycobacterium tuberculosis*

In mycobacteria, Ku plays a crucial role in DNA break repair via non-homologous end joining. Double strand breaks are repaired either by homology directed repair (HDR), where a homologous sequence is used as a template; or non-homologous end joining (NHEJ) which generally involves alteration of sequence at the double strand break site [2]. The efficiency of UvrD1 increases many fold in presence of its binding partner Ku which is assumed to serve as a processivity factor for the unwinding of the double strand [1].

The recombinant plasmid construct was obtained from Dr. K. M. Sinha, Institute of Molecular Medicine, India. The plasmid construct containing the Ku gene (822 bp) from *M. tuberculosis* was transformed into *E. coli* Bl-21(DE3) cells and followed by overnight incubation

on LB-agar/ampicillin plates at 37° C. A single colony was used to inoculate the LB media containing ampicillin and incubated at 37°C until OD_{600} attained a value of ~0.6.

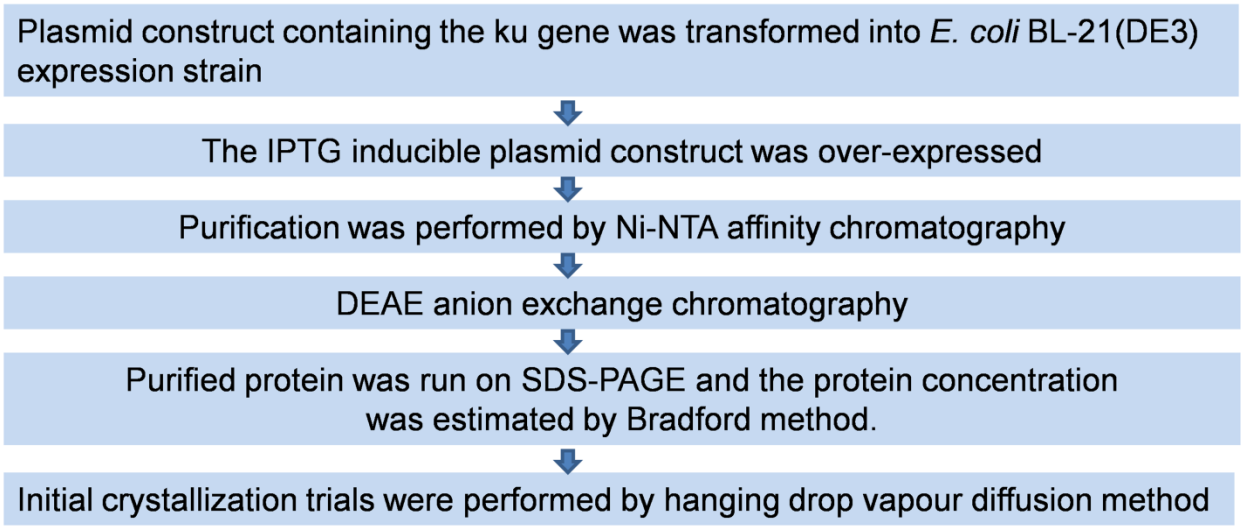


Figure 2: A flowchart of the purification steps of the enzyme Ku from *M. tuberculosis*.

IPTG (isopropyl-β-D-thiogalactopyranoside) was then added to a final concentration of 0.2mM followed by overnight incubation at 17°C. The cells were harvested by centrifugation at 5000g for 15 minutes and suspended in lysis buffer (50mM Tris-HCl, pH 7.5, 250mM NaCl, 10% sucrose, 0.1% triton X-100). Lysozyme was added to reach a final concentration of 200µg/ml. Subsequent to sonication, the homogenized lysate was centrifuged at 14000g for 45 minutes and the supernatant was loaded onto a pre-equilibrated Ni-NTA column. The column was washed with 50mM Tris-HCl, pH 8.0, 0.05% triton X-100, 10% glycerol and eluted with 100 and 200 mM imidazoles. Elution fractions containing protein were pooled and dialysed against 50mM Tris-HCl, pH 8.0, 1mM EDTA, 0.1% triton X-100, 10% glycerol followed by another round of dialysis agaist buffer containing 50mM Tris-HCl, pH 8.0, 1mM EDTA, 0.1% triton X-100, 10% glycerol, 250 mM NaCl. The dialyzed protein was then applied to a pre-equilibrated DEAE

anion exchange column and was eluted by varying the concentration of NaCl from 0.25-1M. The protein was finally dialyzed against 50mM Tris-HCl, pH 8.0, 100 mM NaCl, 2% glycerol, 0.02% NaN_3. The purified enzyme was then run on SDS-PAGE and exhibited a single band of expected molecular weight (**Fig. 10*b***). A flowchart of the purification steps is shown in **Fig. 2.**

Initial crystallization screens were set up at 20°C using the hanging drop vapour diffusion method. Drops containing 7-9µl of protein solution at a concentration of 10mg/ml and 1-3 µl of crystallization buffer were sealed over a 1 ml reservoir.

1.3 Aldose reductase from *Leishmania donovani*

The cytotoxic and mutagenic compound methylglyoxal is catabolized either by glyoxalase pathway (which is dependent on glutathione in eukaryotes except trypanosomes where it is dependent on trypanothione) or an alternate pathway involving aldose reductase [3]. Aldose reductase catalyses the formation of acetol from methylglyoxal in a NADPH dependent two step reaction.

The recombinant plasmid construct was obtained from Prof. M. Rentala, School of life sciences, J. N. U, India. In this construct the aldose reductase gene was cloned into the BamH I and Hind III site of the pMAL-c2 vector. The vector was transformed into *E. coli* BL21(DE3) strains followed by overnight incubation on LB-agar/ampicillin plates at 37° C. A single colony was used to inoculate the LB media containing ampicillin and incubated at 37°C until OD_{600} attained a value of ~0.6. IPTG (isopropyl-β-D-thiogalactopyranoside) was then added to a final concentration of 0.2mM, followed by overnight incubation. The cells were harvested by centrifugation at 5000g for 15 minutes and suspended in lysis buffer (20mM Tris-HCl, pH 7.4, 200mM NaCl, 1mM EDTA, 2 mM phenyl methane sulphonyl fluoride (PMSF) and 1X protease

inhibitor cocktail). Lysozyme was added to reach a final concentration of 200µg/ml. Subsequent to sonication, the homogenized lysate was centrifuged at 14000g for 45 minutes and the supernatant was loaded onto a pre-equilibrated amylase resin column. The column was washed with 20mM Tris-HCl, pH 7.4, 200mM NaCl, 1 mM EDTA, 2mM PMSF and 1X protease inhibitor cocktail. After another wash with 20mM Tris-HCl, pH 7.4, 1M NaCl, 1 mM EDTA, 2mM PMSF and 1X protease inhibitor cocktail, the protein was eluted with 10mM maltose. To remove the maltose binding protein (MBP) fusion tag, 0.02U/µg of factor Xa protease (Qiagen) was added to the eluted protein and was incubated overnight at 4°C. The reaction mixture was added to pre-equilibrated factor Xa removal resin (Qiagen), mixed gently and incubated for 10 minute at 4°C. The reaction mixture was centrifuged at 1000g for 5 minutes and the supernatant was collected. The supernatant was loaded onto a pre-equilibrated (with 20mM Tris-HCl, pH 8.0, 25mM NaCl) Q-sepharose anion exchange column and the flow through was collected. The column was washed with buffer containing 20mM Tris-HCl, pH 8.0, the concentration of NaCl varied from 25-500 mM and the flowthrough was collected as 1ml fraction. Subsequent SDS-PAGE profile indicated that both the fusion protein and ALR eluted in the same fraction. DEAE anion exchange chromatography using a concentration gradient of NaCl turned out to be futile too and the protein of interest could not be separated from the fusion tag.

In order to overcome the difficulty the aldose reductase gene was subcloned into the vector pET-28a(+) of same reading frame. The pMAL-c2-ALR and pET-28a(+) were double digested with BamH I and Hind III and the digestion product was run on agarose gel. The 855 bp insert corresponding to the ALR gene obtained by the double digestion of pMAL-c2-ALR construct was extracted from the gel and finally ligated to the double digested pET-28a(+).

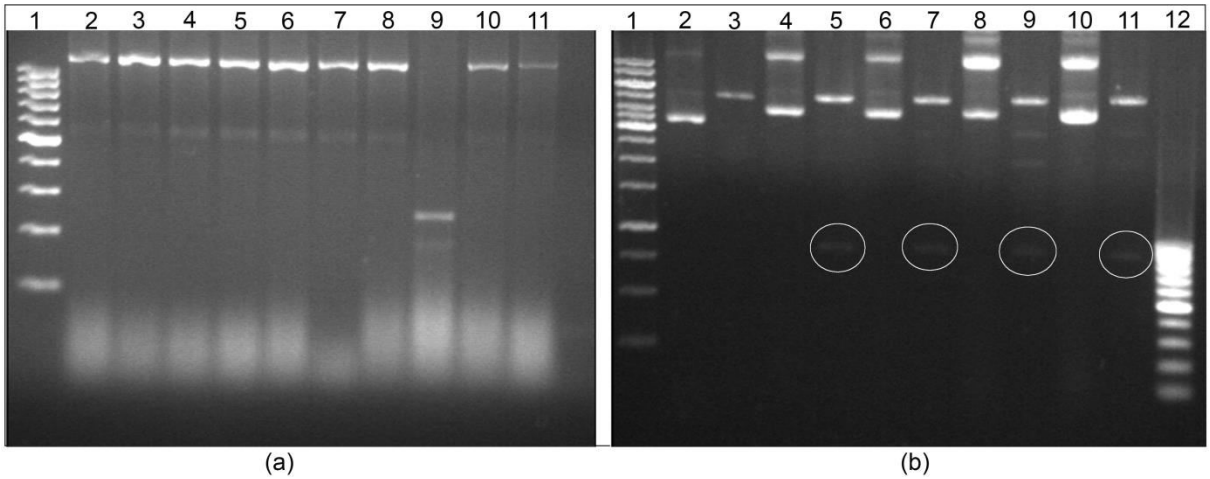


Figure 3: PCR amplification and double digestion of pET-28a-ALR transformed into *E. coli* BL21 (DE3) (a) PCR amplification of the ALR gene from ten randomly selected *E. coli* BL21 (DE3) colonies using T7 forward and reverse primer. Lane 1: 100bp ladder; lane 2-11: PCR amplification of the ALR gene inserted into pET-28a (+); lane 9: Desired product could not be amplified by PCR. (b) Lane 2-3: Undigested and double digested product of pET-28a (+) vector. Lane 4-11: Undigested and double digested product of pET-28a-ALR. Lane 1,12: 1kb and 100 bp ladders respectively.

After purification by ethanol precipitation, the ligated product was transformed into *E. coli* BL21(DE3) strains followed by overnight incubation on LB-agar/kanamycin plates at 37° C. Ten colonies were then selected from the plate and colony PCR was performed using T7 forward and reverse primers which indicated the presence of insert DNA of expected size (**Fig. 3**). The subcloned pET-28a(+) vector was also subjected to automated sequencing in order to confirm the success of the subcloning procedure.

Expression from the construct pET28a (+)-LdALR was induced when the OD_{600} attained a value of ~0.6. IPTG (isopropyl-β-D-thiogalactopyranoside) was then added to a final

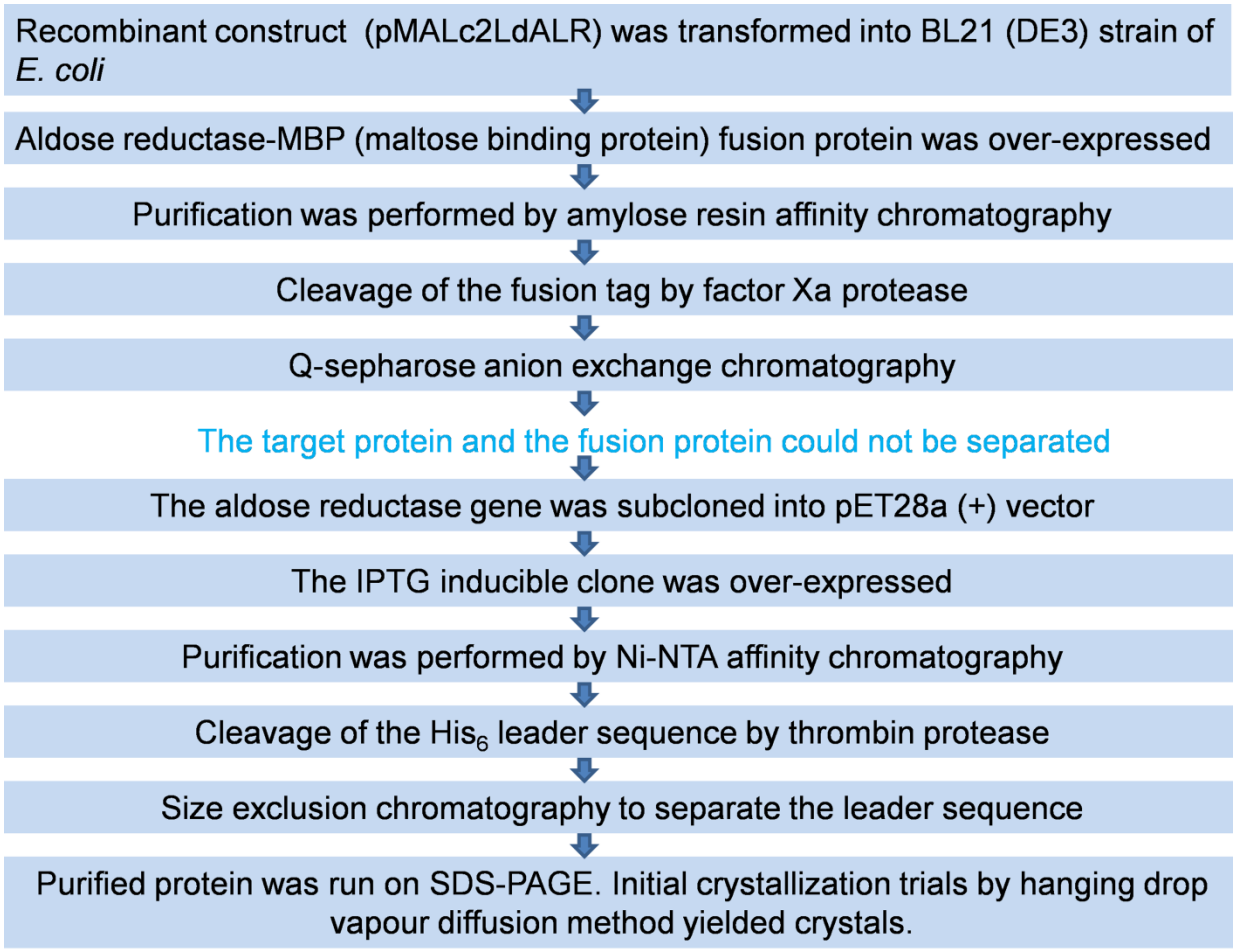


Figure 4: A flowchart of the purification of the enzyme aldose reductase using amylose resin affinity chromatography, subcloning the gene into pET28a(+) vector and finally purification by Ni-NTA chromatography and gel filtration purification steps.

concentration of 1mM followed by overnight incubation. The cells were harvested by centrifugation at 5000g for 15 minutes and suspended in lysis buffer (50mM NaH_2PO_4, pH 8.0, 10mM Imidazole, 2 mM phenyl methane sulphonyl fluoride (PMSF) and 1X protease inhibitor cocktail). Lysozyme was added to reach a final concentration of 200µg/ml. Subsequent to sonication, the homogenized lysate was centrifuged at 14000g for 45 minutes and the supernatant was loaded onto a pre-equilibrated Ni-NTA column. The column was washed with buffer

containing 50mM NaH_2PO_4, pH 8.0, 300mM NaCl, 20mM Imidazole. Finally the protein was eluted with 250mM imidazole. The protein containing fractions were pooled, concentrated using Vivaspin (MW cutoff 10kDa) and dialyzed agains the dialysis buffer (10 mM Tris-HCl, pH 8.0, 100mM NaCl, 0.02% NaN_3). To remove the N-terminal leader sequence (His_6 tag), 0.1U/mg of restriction grade thrombin protease was added to the eluted protein prior to dialyzation. The protein was then loaded onto a pre-equilibrated sephadex G-75 column. The protein containing fractions were pooled together, concentrated and quantified using Bradford method. The purified enzyme was when run on SDS-PAGE, exhibited a single band of expected molecular weight (**Fig. 10*c***). A flowchart of the subcloning and purification steps is shown in **Fig. 4**.

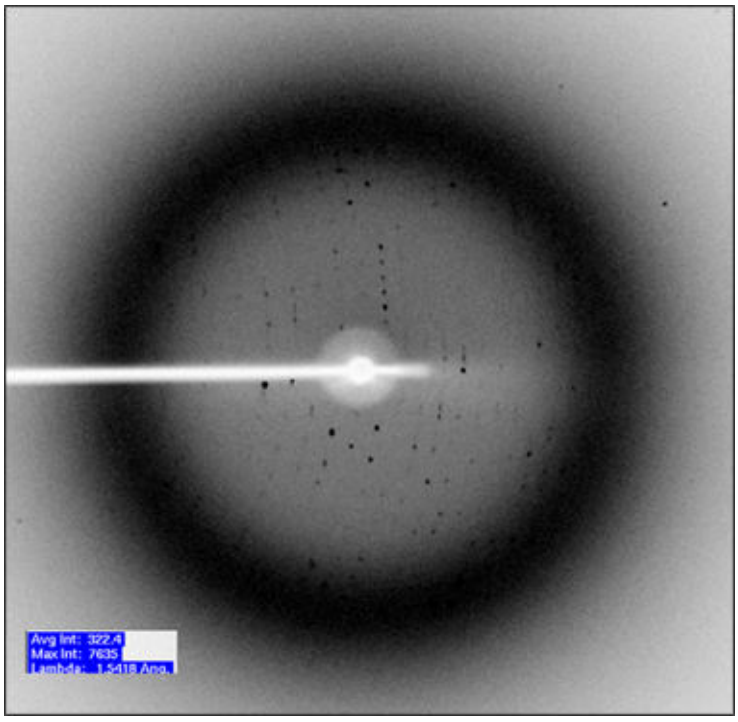


Figure 5: Aldose reductase crystals were found to diffract to maximum resolution of ~7 Å.

Initial crystallization screens were set up at 20°C using the hanging drop vapour diffusion method using ammonium sulfate as the precipitant. Drops containing 7-9µl of protein solution at

a concentration of 10mg/ml and 1-3 µl of crystallization buffer were sealed over a 1 ml reservoir. The protein crystallized in crystallization buffer contained 0.1M NaH_2PO_4, pH 7.0, 10% (Dilution of saturated solution) ammonium sulfate, 25mM NaCl, 1% glycerol, 3mM dithiothreitol and 0.02% NaN_3. The use of micro-seeding technique yielded crystals which were both bigger in size and morphologically better than those obtained by initial trials. Crystals were mounted in capillary tubes and test for diffraction was performed at the inhouse facility (Rigaku rotating Cu-anode X-ray diffractometer) at room temperature. Crystals were found to diffract poorly to a maximum resolution of ~7Å (**Fig. 5**).

1.4 Phosphofructokinase from *L. donovani*

The trypanosomes rely solely on the glycolysis pathway for the generation of ATP. Properties of some of the glycolytic enzymes vary significantly from its host [4]. One example is phosphofrucktokinase (PFK) which although belongs to the inorganic pyrophosphate (PP_i) dependent PFK family, yet uses ATP as the donor of the phosphate group. PFK catalyzes the conversion of fructose 6-phosphate (F6P) to fructose 1,6-bisphosphate (FBP). In *T. brucei* the enzyme is localized inside the glycosome and AMP is the only known activator of this enzyme.

The plasmid construct containing the phosphofrucktokinase gene was purchased from Genscript. The construct was transformed into the *E. coli* BL21(DE3) followed by overnight incubation on LB-agar/kanamycin plates at 37° C. A single colony was used to inoculate the LB media containing kanamycin and incubated at 37°C until OD_{600} attained a value of ~0.6. IPTG (isopropyl-β-D-thiogalactopyranoside) was then added to a final concentration of 0.2, 0.5 and 1mM respectively, followed by overnight incubation. The over expression of the PFK gene could not be ascertained by running the cell lysate on SDS-PAGE. The plasmid construct was then

transformed into two different expression strains of *E. coli*-BL21pLysS and Rosetta. In order to induce expression of the gene, 0.2, 0.5 and 1mM IPTG was added to the culture media followed by incubation. But over-expression of the PFK gene could not be inferred from the SDS-PAGE profle of the cell lysate.

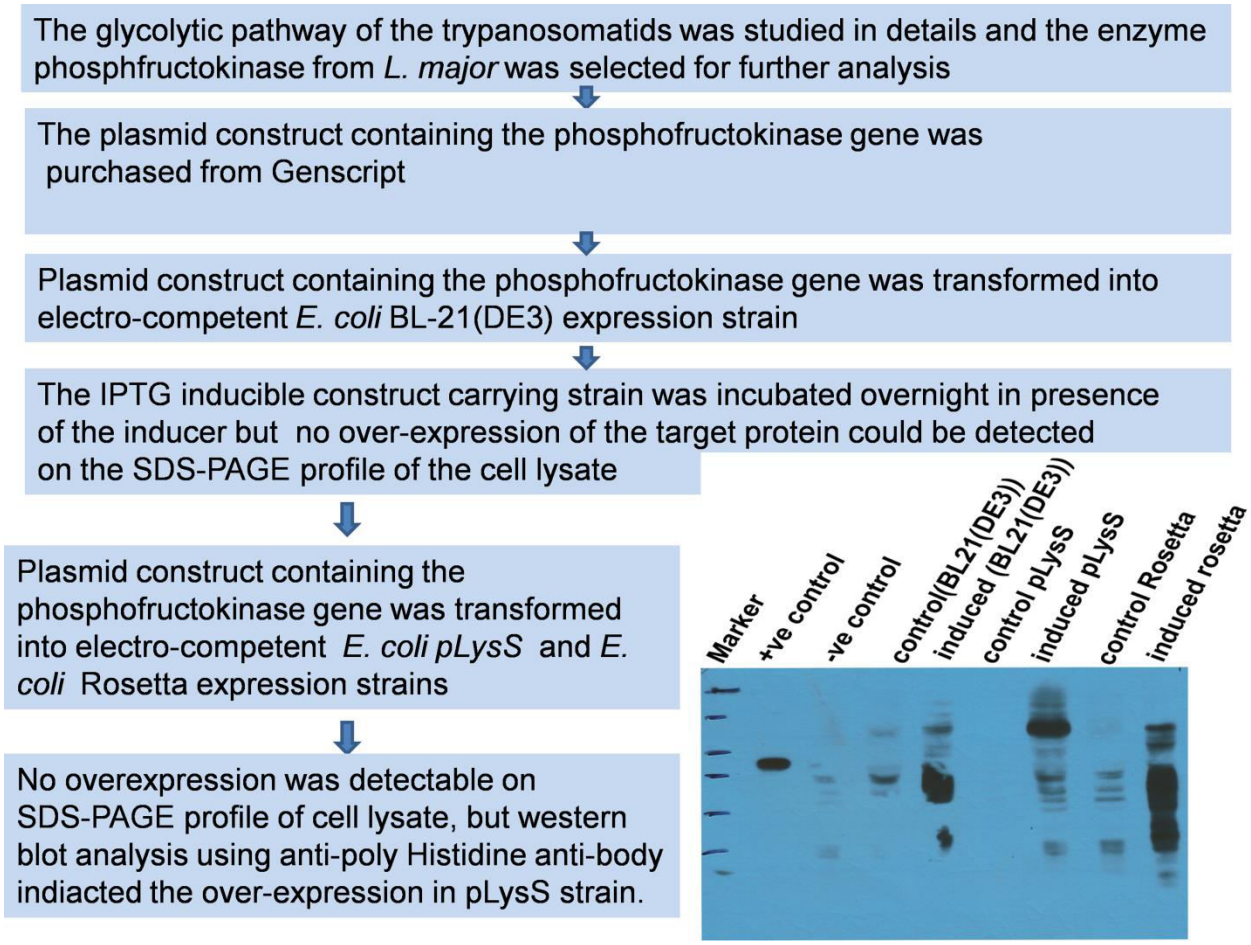


Figure 6: Transformation of the plasmid construct containing the phosphofructokinase gene in three different expression hosts and the exression profile of the gene in those strains as indicated by western blot analysis.

The cell lysates of the three expression strains were separated on SDS-PAGE. A nitro cellulose sheet (6cm×8cm) was soaked in 40ml transfer buffer (20mM Tris-HCl, 192mM glycine, 20%

(v/v) methanol) for 15 minutes. The polyacrylamide gel-membrane sandwitch was placed inside the buffer chamber and the current was set at 300mA for one and half hour for electro transfer. After completion of the transfer, the membrane was stained with Ponceau S and the visible bands corresponding to the protein molecular weight marker were marked with a ball point pen. The memebrane was washed thrice with PBST (1.4mM KH_2PO_4, 8mM Na_2HPO_4, 140mM NaCl, 2.7mM KCl and the pH was adjusted to 7.4 with NaOH followed by the addition of 0.05% tween-20). The membrane was blocked for one hour by the addition of 5% skimmed milk in PBST at room temperature followed by three washes with PBST. The membrane was incubated overnight at 4°C in the presence of primary mouse anti-His antibody (1X PBST, 3% BSA, mouse anti-His antibody at 1000 fold dilution). After three washes with PBST, the blot was incubated in the presence of goat secondary antibody (1X PBST, goat HRP-conjugated secondary antibody at 20000 fold dilution) for two hours. The blot was washed thrice with PBST, followed by the drop wise addition of solution I (1.5M Tris,pH 8.8, 90 mM coumaric acid, 250mM luminol) and II (1.5M Tris,pH 8.8, 30% H_2O_2). The membrane was exposed to X-ray films and the film was finally dipped in developing agent and fixation agent.

The western blot analysis indicated that upon induction, the expressed protein undergoes degradation in both BL21(DE3) and Rosetta strains. In contrast, the PFK gene was found to express poorly in pLysS strains and the protein appeared to remain relatively stable after cell lysis (**Fig. 6**).

1.5 Phosphoglucomutase from *L.major*

Discussed in details in chapter III.

1.6 Phosphoglycerate kinase B from *L.major*

The glycolytic enzyme Phosphoglycerate kinase (PGK) catalyses the first substrate level phosphorylation reaction: conversion of 1,3-bisphosphoglycerate to 3-phosphoglycerate [5]. Structural studies of PGK from *T. brucei* revealed it to be a hinge bending enzyme where the bending leads to the closure of the central cleft, bringing the two ligands within close proximity [6].

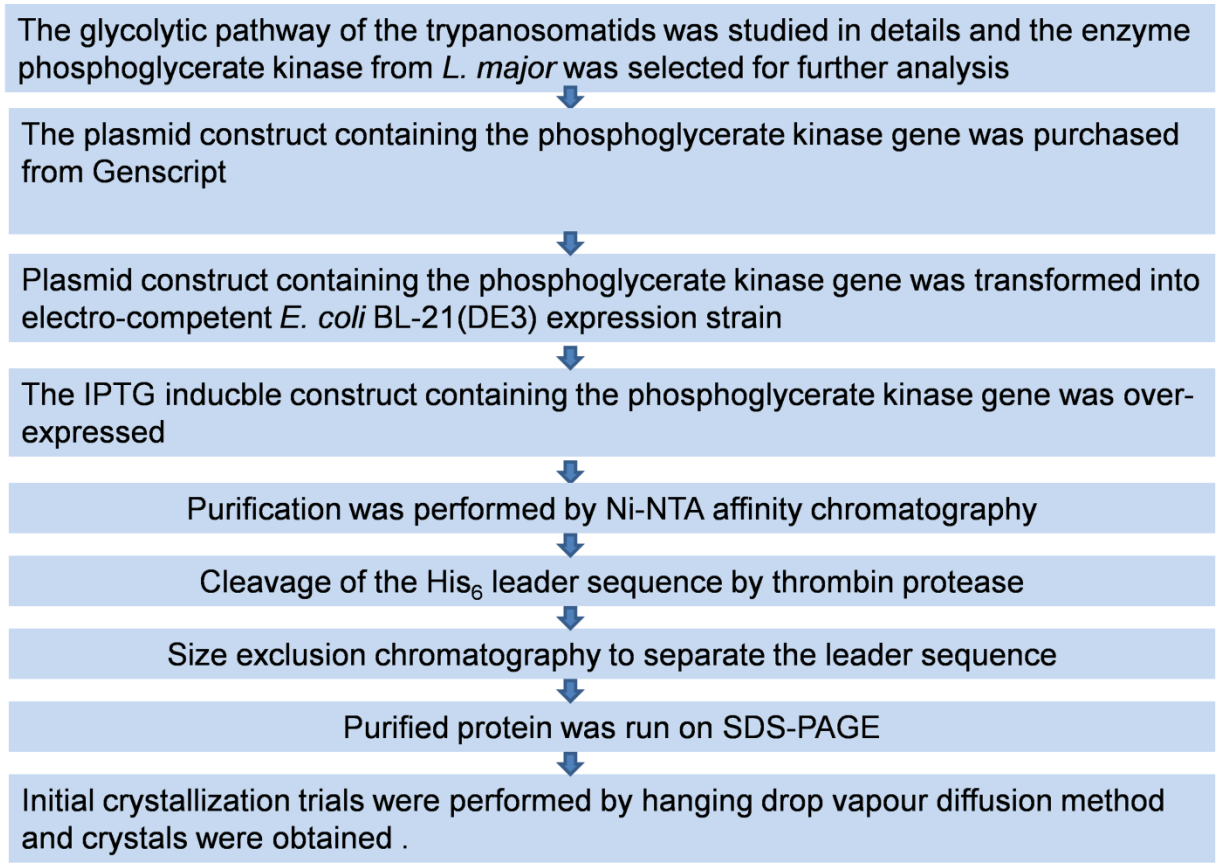


Figure 7: Steps involved in the expression and purification of the enzyme phosphoglycerate kinase B from *L. major*.

The plasmid construct containing the phosphoglycerate kinase B gene was purchased from Genscript. The construct was transformed into the *E. coli* BL21(DE3) cells followed by overnight incubation on LB-agar/kanamycin plates at 37° C. A single colony was used to

inoculate the LB media containing kanamycin and incubated at 37°C until OD_{600} attained a value of ~0.6. IPTG (isopropyl-β-D-thiogalactopyranoside) was then added to a final concentration of 0.75mM respectively, followed by overnight incubation. The cells were harvested by centrifugation at 5000g for 15 minutes and suspended in lysis buffer (25mM Tris, pH 7.5, 400mM NaCl, 10mM Imidazole, 0.01% triton X-100, 10mM β-mercaptoethanol, 10% glycerol, 2 mM phenyl methane sulphonyl fluoride (PMSF) and 1X protease inhibitor cocktail). Lysozyme was added to reach a final concentration of 200µg/ml. Subsequent to sonication, the homogenized lysate was centrifuged at 14000g for 45 minutes and the supernatant was loaded onto a pre-equilibrated Ni-NTA column. The column was washed with buffer containing 25mM Tris, pH 7.5, 400mM NaCl, 20mM Imidazole, 0.01% triton X-100, 10mM β-mercaptoethanol, 10% glycerol. Finally the protein was eluted with 250mM imidazole. The protein containing fractions were pooled, concentrated using Vivaspin (MW cutoff 10kDa) and dialyzed agains the dialysis buffer (25 mM Tris-HCl, pH 7.5, 100mM NaCl, 3mM DTT, 1% glycerol, 0.02% NaN_3). To remove the N-terminal leader sequence (His_6 tag), 0.1U/mg of restriction grade thrombin protease was added to the eluted protein prior to dialyzation. The protein was then loaded onto a pre-equilibrated sephadex G-75 column. The protein containing fractions were pooled together, concentrated and quantified using Bradford method. The purified enzyme exhibited a single band of expected molecular weight on SDS-PAGE (**Fig. 10*e***). A flowchart of the purification method is shown in **Fig. 7**.

Initial crystallization screens were set up at 20°C using the hanging drop vapour diffusion method using polyethyleneglycol (MW ~8000) as the precipitant. Drops containing 2µl of protein solution at a concentration of 10mg/ml and 2 µl of crystallization buffer were sealed over

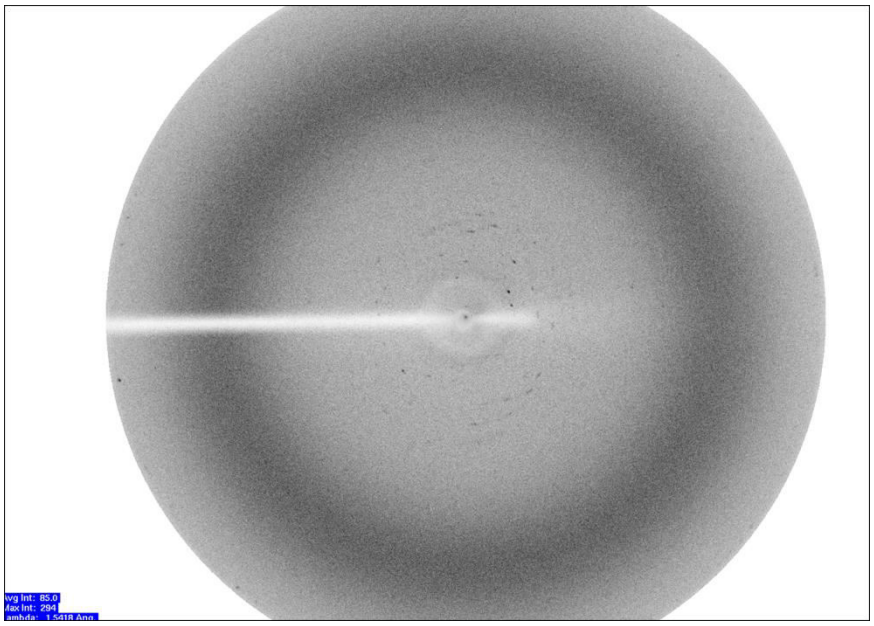


Figure 8: The crystals of phosphoglycerate kinase B diffracted poorly to a maximum resolution of 7Å.

a 1 ml reservoir. The protein crystallized in crystallization buffer containing 50mM Tris-HCl, pH 7.0, 25% (w/v) PEG 8K, 50 mM sodium citrate, 50 mM ammonium acetate. Crystals were mounted in capillary tubes and test for diffraction was performed at the inhouse facility (Rigaku rotating Cu-anode X-ray diffractometer) at room temperature. Crystals were found to diffract poorly to a maximum resolution of ~7Å (**Fig. 8**).

1.7 Fructose-1,6-bishposphate from *L.major*

The gluconeogenic enzyme fructose-1,6-bisphosphate (F-1,6-BP) is responsible for the conversion of fructose-1,6-bisphosphate to fructose-6-phosphate. In trypanosomes, F-1,6-BP is localized within the glycosome like other glycolytic enzymes [7]. The F-1,6-BP null mutant of *L. major* failed grow in the absence of hexose whereas resulted in exhaustion of internal carbohydrate reserves in a glycerol rich media. The null mutant promastigotes engulfed by

macrophages were able to transform into amastigotes but were unable to replicate indicating that they rely heavily on non-glucose carbon sources within the macrophage.

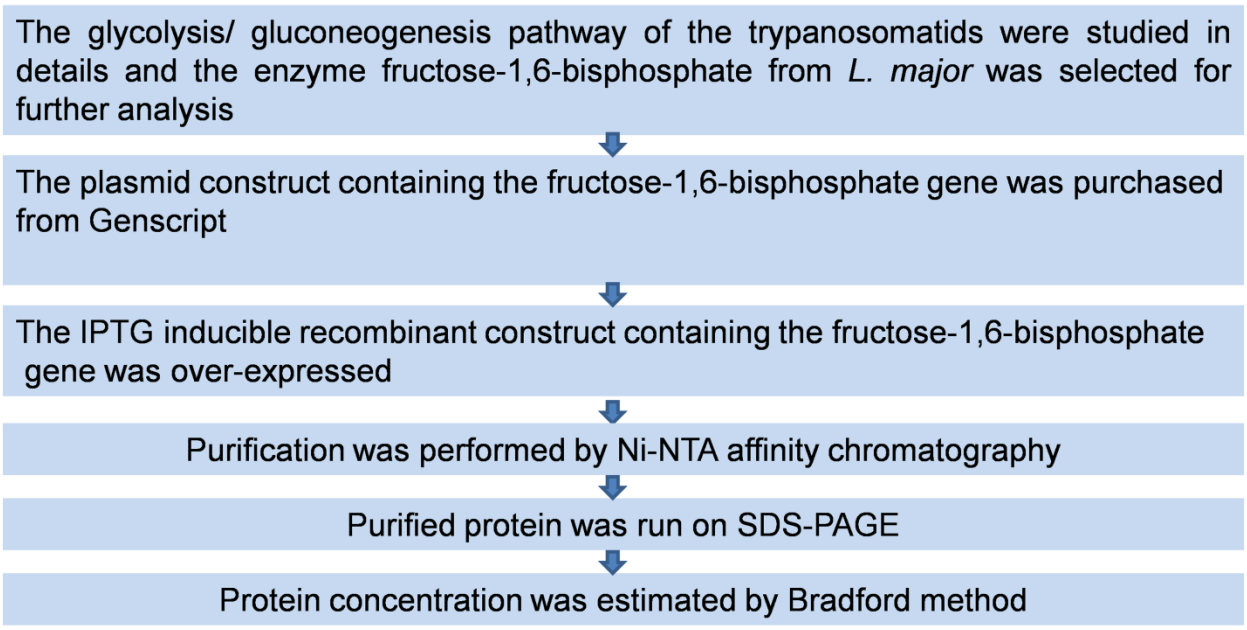


Figure 9: A schematic of the purification steps of the enzyme fructose-1,6-bishposphate from *L.major*.

The plasmid construct containing the fructose-1,6-bishposphate gene was purchased from Genscript. The construct was transformed into the *E. coli* BL21(DE3) cells followed by overnight incubation on LB-agar/kanamycin plates at 37° C. A single colony was used to inoculate the LB media containing kanamycin and incubated at 37°C until OD_{600} attained a value of ~0.6. IPTG (isopropyl-β-D-thiogalactopyranoside) was then added to a final concentration of 0.5 mM respectively, followed by overnight incubation at 30°C.

The cells were harvested by centrifugation at 5000g for 15 minutes and suspended in lysis buffer (25mM Tris, pH 7.5, 400mM NaCl, 10mM Imidazole, 0.01% triton X-100, 10mM β-

mercaptoethanol, 10% glycerol, 2 mM phenyl methane sulphonyl fluoride (PMSF) and 1X protease inhibitor cocktail). Lysozyme was added to reach a final concentration of 200µg/ml. Subsequent to sonication, the homogenized lysate was centrifuged at 14000g for 45 minutes and the supernatant was loaded onto a pre-equilibrated Ni-NTA column. The column was washed with buffer containing 25mM Tris, pH 7.5, 400mM NaCl, 20mM Imidazole, 0.01% triton X-100, 10mM β-mercaptoethanol, 10% glycerol. Finally the protein was eluted with 250mM imidazole. The purified enzyme exhibited a single band of expected molecular weight on SDS-PAGE (**Fig. 10*f***). The protein tends to precipitate on storage (both at 4°C and -80°C after flash freezing in liquid nitrogen).

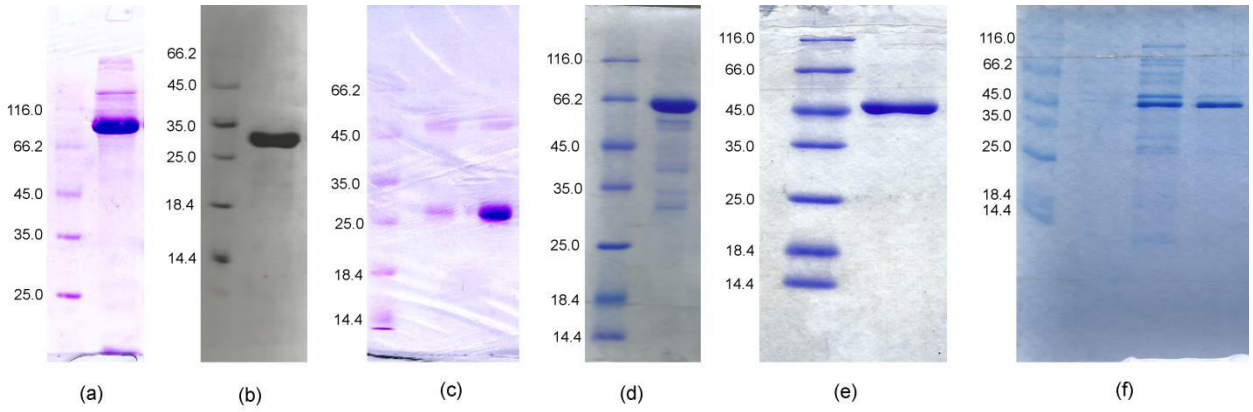


Figure 10: SDS-PAGE profiles of the purified enzymes. (a) UvrD1 (MW: 85kDa) (b) Ku (MW: 30.9 kDa) (c) Aldose reductase (MW: 32kDa) (d) Phosphoglucomutase (63.9kDa) (e) Phosphoglycerate kinase B (MW: 45kDa) (f) Fructose-1,6-bisphosphate (39kDa)

A flowchart of the purification process is shown in **Fig. 9**. The purification and crystallization details of all the seven enzymes from various pathogenic organisms are listed in **Table 1**.

Table 1: Purification and crystallization trials of several enzymes from different pathogenic organisms.

Sl. No.	Enzyme	Organism	Molecular weight of protein (kDa)	Function	Crystals obtained	Remarks
1	ATP-dependent DNA helicase II UvrD1	*M. smegmatis*	85	Helicase involved in DNA break repair and recombination	No	Crystallization trials by hanging drop vapour diffusion method
2	Ku	*M. tuberculosis*	30.9	Non-homologous end joining protein and binding partner of UvrD1	No	Crystallization trials by hanging drop vapour diffusion method

3	Glutathione specific aldose reductase	*L. donovani*	32	Methylglyoxal detoxification		Crystals obtained by hanging drop vapour diffusion method
4	Phosphofructokinase	*L. donovani*	54	Glycolysis	____	poor expression
5	Phosphoglucomutase	*L. major*	63.9	Sugar nucleotide synthesis pathway		Crystals obtained by hanging drop vapour diffusion method
6	Phosphoglycerate kinase	*L. major*	45	Glycolysis		Crystals obtained by hanging drop vapour diffusion method
7	Fructose-1,6-bisphosphate	*L. major*	39	Gluconeogenesis	No	Protein tends to precipitate on storage

2. A Study of in-vitro interactions of anti-leishmanial drugs against promastigotes of *L.major* (strain 5ASKH) to explore the possibility of synergistic effects.

2.1 Introduction

Due to emergence of leishmanial strains resistant to the existing therapies and the toxic side effects of the available drugs, alternative therapeutic strategies such as multidrug treatment are being explored. One rational approach towards lowering the rate of treatment failure and to check the emergence and spreading of resistant strains is currently provided by the combination therapy of antileishmanial drugs. The combinatorial effects of administered drugs may exceed the sum of their individual effects (synergism) or may produce effect less than predicted (antagonism). A simple additive effect is represented by linear additive isoboles where as non-linear additive isoboles indicate either synergism or antagonism. The mathematical formulae used to distinguish cases have been discussed in later sections.

This part of the project focuses on the study of the combinatorial effect of the existing antileishmanial drug paromomycin and the natural product: curcumin against the promastigotes of *L.major*, strain 5ASKH.

Paromomycin, an aminoglycoside antibiotic is currently used for the treatment of leishmaniasis. In bacteria, paromomycin interacts with ribosomal RNA subunits and thereby inhibits protein synthesis. In trypanosomatids, the cationic paromomycin have been suggested to bind to the negatively charged glycocalyx and enter the cell by endocytosis [8]. Its uptake in the mitochondria is regulated by the memebrane potential. In-vivo studies of leishmania promastigotes have shown that both protein synthesis and proliferation were markedly inhibited

by paromomycin and to a much lesser extent by streptomycin and neomycin B [9]. Surface plasmon resonance studies showed that the antibiotic binds strongly to the decoding site of the parasite ribosome [9]. A comparative proteome study of the wild type and paromomycin resistance strains of *L. donovani* revealed the upregulation of several proteins in later, involved in the translational machinery and probably offers a defence mechanism against paromomycin [8]. Another protein which was upregulated in resistance strain was lipophosphoglycan biosynthetic protein, putative (LPG3), responsible for the synthesis of glycoconjugates, a virulence factor. The null mutants are viable but exhibit defective phosphoglycosylation.

Curcumin, a plant phenol, exhibits a wide range of therapeutic effects. High doses of curcumin do not exhibit adverse effects in different animal models [10]. Curcumin exhibits biological and pharmacological properties such as antioxidative, anti-inflammatory, anti-infectious immunomodulatory effects, apoptosis-regulator, antiangiogenic and anticarcinogenic. Besides, it interacts with several molecules such as growth factors, transcription factors and cytokines. Studies of *T. cruzi* infected mice revealed that the macrophage infiltration was reduced by 38% in curcumin treated compared to untreated model. *Trypanosoma cruzi*, responsible for the Chagas disease, causes an acute myocarditis and chronic cardiomyopathy in infected host. Treatment with curcumin resulted in marked reduction in the mRNA levels of inflammatory markers such as TNF-α, IL-19, IL-22, Bcl2 like1, COX-2 and TLR-9 compared to the untreated infected mice [10]. Besides, curcumin was repoted to reduce the mRNA levels of enzymes involved in oxidative signaling such as catalase, peroxidase, superoxide dismutases (SOD), nitric oxide synthases 2 (NOS2), and NOS1 in heart tissue compared to untreated infected mice heart. Curcumin was found to inhibit phospholipase A2 (PLA2), cyclooxygenases (COX) and lipoxygenases (LOX) which are involved in the metabolism of arachidonic acid in

cell membranes resulting in the generation of the messengers of the inflammatory response i.e. eicosanoids [10].

The level of the two protein caveolin-1 (Cav-1) and Cav-3 reduces in the hearts of infected mice. (Cav-1) and Cav-3 null mice are reported to develop cardiomyopathy, a condition similar to that caused by *T. cruzi* infection. Curcumin have been shown to regulate the Cav-1 expression and can inhibit atherosclerosis (narrowing of the arteries) by reducing the mRNA levels of intercellular adhesion molecule (ICAM) and vascular cell adhesion molecule (VCAM), the known inducers of atheroscleroscis.

Quercetin is a polyphenolic flavonoids found in plants such as onion, gingko-biloba and tea [11]. Quercetin is reported to inhibit tumor necrosis factor-α (TNF-α), induced in *T. brucei* infected monocytoid human cell line, by inhibiting the phosphorylation and activation of Jun N-terminal kinase/stress-activated protein kinase which in turn suppresses the AP-1 (activating protein-1) activation, involved in the post transcrictional regulation of TNF-α mRNA [12]. Besides, quercetin was reported to be a specific inhibitor of Cox-2 and NO synthases (NOS-II) and thereby exhibits anti imflammatory and anti oxidant properties. Quercetin was reported to promote time dependent apoptosis of *T. brucei. gambiense* in mice models [12]. In vitro studies show that quercetin is able to suppress *T. gondii* development by inhibiting the synthesis of several enzymes involved in the synthesis of virulence factors susch as Hsp90, Hsp70, and Hsp27 [12].

2.2 Materials and methods

2.2.1 Materials

Media M199, curcumin was obtained from SIGMA. Quercetin was from sisco research laboratories. Fetal bovine serum (FBS) was obtained from GIBCO. Penicillin, streptomycin, 3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium (MTT) was obtained from HIMEDIA.

2.2.2 Parasite culture

L. major (strain 5ASKH) was kindly provided by Dr. Subrata Adak, IICB, India. The promastigotes were maintained at 22°C in media M 199, supplemented with 1% pencillin-streptomycine and 10% heat inactivated fetal bovine serum.

2.2.3 Drug activity assay on *L. major* (5ASKH) promastigotes

Stock solutions of curcumin and quercetin were prepared in DMSO. Stock solutions of paromomycin was prepared in PBS (1.4mM KH_2PO_4, 8mM Na_2HPO_4, 140mM NaCl, 2.7mM KCl and the pH was adjusted to 7.4 with NaOH). IC_{50} of individual drugs were determined from the survivor curve prior to analyzing the combinatorial effects. Final DMSO concentration was maintained below 0.5% which did not exhibit any adverse effect on parasite clearance. Fixed ratio solutions of one drug and the corresponding partner drug were prepared at ratios of 100:0, 60:1, 40:2, 20:3, 1:4, and 0:5. Parasites were seeded at $1x10^6$ parasites/ml and grown for 5days in presence of serial dilution of drugs over five concentrations. Medium and drug dilutions were replaced after three days. Following incubation, the cells were washed twice with equal volume of PBS and finally suspended in 100ul of PBS supplemented with 10ul of MTT (5mg/ml) and incubated for another 3hrs. At the end of incubation, a solubilization solution (acid- isopropanol) was added to dissolve the formazan. The absorbance intensity was measured using test wavelength of 570nm and reference wavelength of 630nm. All experiments were performed in

triplicate. The fixed ratio solutions 100:0 and 0:5 yielded the IC_{50} of the two individual drugs and solutions 60:1, 40:2, 20:3, 1:4 yielded IC_{50} of the two drugs in combination.

2.2.4 Statistical analyses

Statistical analyses were performed with One Way ANOVA (GraphPad Prism 5) and a $P<0.001$ considred stastically significant. Graphical illustrations were performed using GraphPad Prism 5.

2.2.5 Determination of FIC index and determination of nature of interaction

The IC_{50} of the individual drugs and those in combination were determined from the respective survival curves. The fractional inhibitory concentration (FICs) and the sum of FICs (ΣFICs) were calculated as was described by Seifert *et al*. [13]. FIC of the drug was calculated as:

FIC of the drug = (IC_{50} of the drug in combination)/ (IC_{50} of the individual drug)

The sum of FIC was calculated as:

ΣFIC = FIC of drug A + FIC of drug B

The mean ΣFIC values were used to classify the nature of interaction as per the values given in **Table 2**.

Table 2: The mean ΣFIC values indicating the nature of interaction.

Mean ΣFIC≤ 0.5	Synergism
0.5 > Mean ΣFIC ≤ 4	Indifference
Mean ΣFIC > 4	Antagonism

2.2.6 Construction of isobologram

The FICs of drug A was plotted along the X-axis while the FICs of drug B along the Y-axis. The straight line connecting the IC_{50} (FIC_A = 1, FIC_B = 1) of the two drugs represents 'line of additivity' (indifference) which distinguishes between simple additive, synergistic or antagonistic interactions. The three cases are shown graphically in **Fig 11**.

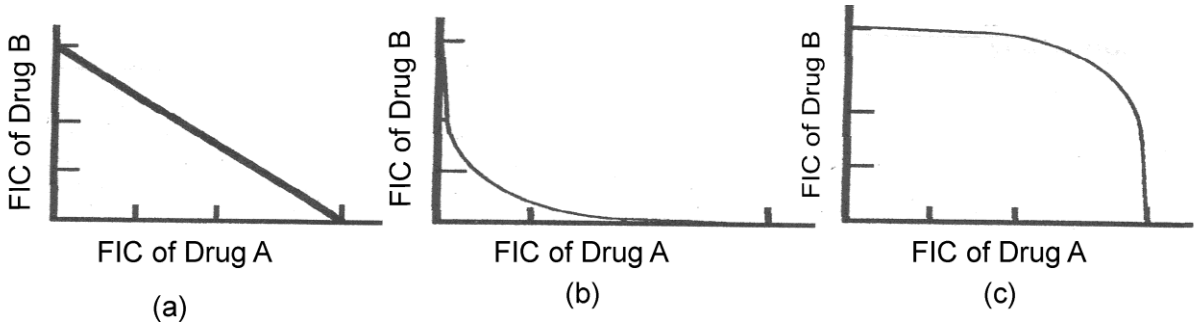


Figure 11: Isobolograms depicting the three possible interactions of a combination of two drugs (a) indifference (b) synergism (c) antagonism.

2.3 Results

In vitro interactions of paromomycin and curcumin were analysed at the IC_{50} level against the promastigotes of *L. major* using modified fixed ratio isobologram method [13]. The individual IC_{50} of the drugs were determined from the survival curves and were estimated to be 204µM for paromomycin (P), 2.15µM for curcumin (C) and 73.12µM for quercetin (Q). In order to study the in-vitro interaction of paromomycin and curcumin, fixed ratio solution of P and C were prepared as 100:0, 60:1, 40:2, 20:3, 1:4, 0:5. Overall ΣFICs of the experiment ranged from 0.942 (for a ratio of P: C= 20:3) to 1.34 (for a ratio of P: C = 60:1) and the mean ΣFIC was calculated to be 1.08. Therefore, the in vitro interaction between paromomycin and curcumin was classified as indifferent according to the values in **Table 2**. The corresponding isobologram is shown in **Fig. 12**.

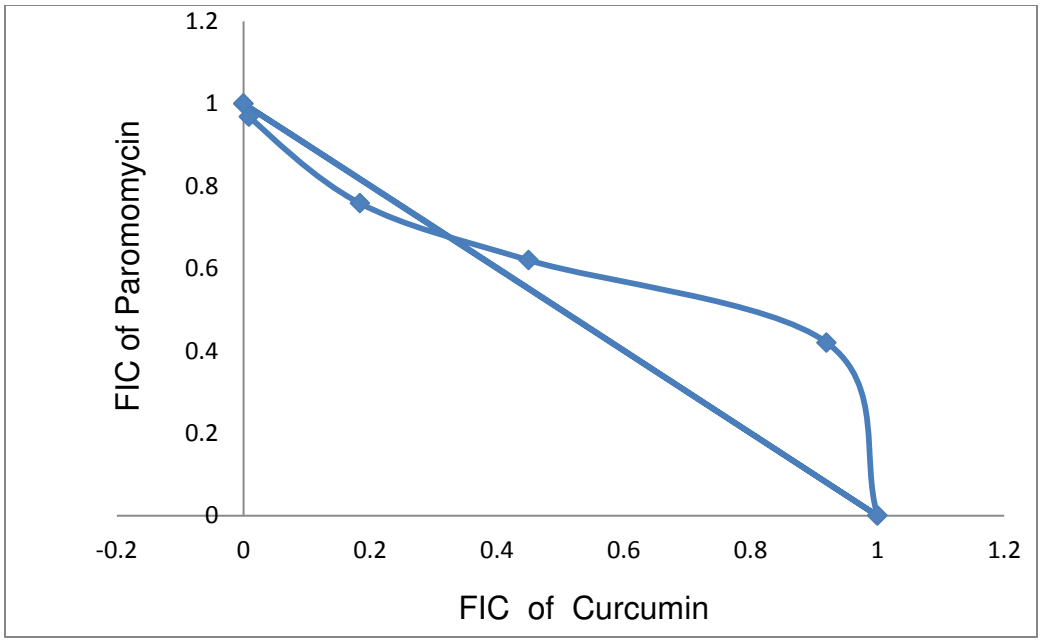


Figure 12: Representative isobologram of in vitro interactions of paromomycin and curcumin against *L.major* (strain 5ASKH) promastigotes at the IC_{50} level.

2.4 Conclusion

In recent years, combination drug therapies have achieved greater therapeutic efficacy than monotherapy for the treatment of AIDS, tuberculosis, malaria and cancer [14]. The drug Malarone, resulting from the combination of atovaquone and proguanil; Lap-Dap- combination of chlorproguanil and dapsone and Coartem- combination of artemether and lumefantrine are being developed or have become available for the treatment of malaria [14]. Therefore, multidrug therapy may result in increased efficacy, reduced dosage and toxicity besides delaying the emergence of drug resistant strains of *Leishmania spp*.

References

1. Sinha KM, Glickman MS, Shuman S: **Mutational analysis of Mycobacterium UvrD1 identifies functional groups required for ATP hydrolysis, DNA unwinding and chemomechanical coupling.** *Biochemistry* 2009, **48**: 4019-4030.

2. Pierce A J, Hu P, Han M: **Ku DNA end binding protein modulates homologous repair of double strand breaks in mammalian cells**. *Genes Dev.* 2001, **15**: 3237-3242.

3. Rath J, Gowri VS, Chauhan SC, Padmanabhan PK, Srinivasan N, Madhubala R: **A glutathione specific aldose reductase of *Leishmania donovani* and its potential implications for methylglyoxal detoxification pathway**. *Gene* 2009, **429**(1-2): 1-9. doi: 10.1016/j.gene.2008.09.037.

4. Martinez-Oyanedel J, McNae IW, Nowicki MW, Keillor JW, Michels PAM, Fothergill-Gilmore LA, Walkinshaw MD: **The First Crystal Structure of Phosphofructokinase from a Eukaryote: Trypanosoma brucei**. *J. Mol. Biol.* 2007, **366**: 1185–1198.

5. Barros-Álvarez X, Cáceres AJ, Michels PAM, Concepción JL, Quiñones W: **The phosphoglycerate kinase isoenzymes have distinct roles in the regulation of carbohydrate metabolism in Trypanosoma cruzi**. *Experimental Parasitology* 2014, **143**: 39–47.

6. Bernstein BE, Michels PA, Hol WG: **Synergistic effects of substrate-induced conformational changes in phosphoglycerate kinase activation**. *Nature* 1997, **385**(6613): 275-8.

7. Naderer T, Ellis MA, Sernee MF, De Souza DP, Curtis J, Handman E, McConville MJ: **Virulence of Leishmania major in macrophages and mice requires the gluconeogenic enzyme fructose-1,6-bisphosphatase**. *PNAS* 2006, 103(14): 5502–5507.

8. Chawla B, Jhingran A, Panigrahi A, Stuart KD, Madhubala R: **Paromomycin affects translation and vesicle-mediated trafficking as revealed by proteomics of paromomycin –susceptible –resistant leishmania donovani**. *PLoS ONE* 2011, 6(10): e26660.

9. Fernández MM, Malchiodi EL, Algranati ID: **Differential Effects of Paromomycin on Ribosomes of *Leishmania mexicana* and Mammalian Cells.** *Antimicrob. Agents Chemother.* 2011, **55**(1): 86-93.

10. Nagajyothi F, Zhao D, Weiss LM, Tanowitz HB: **Curcumin treatment provides protection against *Trypanosoma cruzi* infection.** Parasitol Res. 2012, **110**(6): 2491–2499. doi:10.1007/s00436-011-2790-9.

11. Fonseca-Silva F, Inacio JDF, Canto-Cavalheiro MM, Almeida-Amaral EE: **Reactive Oxygen Species Production and Mitochondrial Dysfunction Contribute to Quercetin Induced Death in Leishmania amazonensis.** *PLoS ONE* 2011, 6 (2): e14666

12. Mamani-Matsuda M, Rambert J, Malvy D, Lejoly-Boisseau H, Dauloue`de S, Thiolat D, Coves S, Courtois P, Vincendeau P, Mossalayi MD: **Quercetin Induces Apoptosis of *Trypanosoma brucei gambiense* and Decreases the Proinflammatory Response of Human Macrophages**. *Antimicrobial agents and chemotherapy* 2004, **48**(3): 924–929. DOI: 10.1128/AAC.48.3.924–929.2004.

13. Seifert K, Munday J, Syeda T, Croft SL: **In vitro interactions between sitamaquine and amphotericin B, sodium stibogluconate, miltefosine, paromomycin and pentamidine against Leishmania donovani**. *J Antimicrob Chemother*. 2011, **66**(4): 850-4. doi: 10.1093/jac/dkq542.

14. Fivelman QL, Adagu IS, Warhurst DC: **Modified fixed-ratio isobologram method for studying in vitro interactions between atovaquone and proguanil or dihydroartemisinin against drug-resistant strains of Plasmodium falciparum.** *Antimicrob Agents Chemother.* 2004, **48**(11): 4097-102.

.