

Evolutionary Aspect of Miltefosine Transporter Proteins in *Leishmania major*

Ritika Kabra¹ and Shailza Singh^{1*}

National Centre for Cell Science, NCCS Complex, Ganeshkhind, SP Pune University
Campus, Pune-411007, INDIA

*corresponding author, singhs@nccs.res.in

Abstract

Transporter proteins, P-glycoprotein (P-gp) and P4ATPase-CDC50, are responsible for the transport of Miltefosine drug across cell membrane of a protozoan parasite *Leishmania major*. Mutations or change in activity of these proteins may lead to emergence of resistance in the parasite. Owing to the structural and functional importance of these transporter proteins, in this paper, we have tried to decipher the evolutionary divergence of these Miltefosine transporter proteins across different forms of life including Protists, Fungi, Plants and Animals. We retrieved 96, 207, and 189 sequences of P-gp, P4ATPase and CDC50 proteins respectively, across diverse variety of organisms for the conserved analysis. Phylogenetic trees were constructed for these three transporter proteins based on Bayesian posterior probability inference. The evolutionary analysis concluded that these proteins remain highly conserved throughout the species diversity but still substantial differences in the proteins for host (*Homo sapiens*) and parasite (*L. major*) were observed which have led in targeting these Miltefosine transporter proteins in a parasite specific manner.

Keywords: Evolutionary divergence, phylogenetic analysis, transporter proteins, *Leishmania major*

Abbreviations: ABC: ATP binding cassette; ATP: Adenosine Tri Phosphate; CFTR: Cystic Fibrosis Transmembrane conductance Regulator; gGAPDH: glycosomal glyceraldehyde phosphate dehydrogenase; HMM: Hidden Markov Model; MalK: Maltose/maltodextrin import ATP-binding protein; MCMCMC: Metropolis Coupled Markov Chain Monte Carlo; MSA: Multiple Sequence Alignment; NBD: Nucleotide binding domain; NCBI: National Centre for Biotechnology Information; P-gp: P-glycoprotein; SDP: positions determining specificity; SSU

rRNA: small subunit ribosomal RNA; TMD: Transmembrane Domain; TMH: TransMembrane Helices; TMS: TransMembrane Segments.

1. Introduction

Leishmaniasis, a neglected tropical disease, is the second largest parasitic killer in the world only after malaria. It is prevalent in more than 90 countries all over world with an estimate of 2 million new cases every year (Herwaldt 1999; India National vector Borne Diseases Control Programme 2017). As for treatment regimen, various drugs available are antimonials, Miltefosine, sitamaquine, Paromomycin, amphotericin B, etc. Of these, Miltefosine is the only one to be administered orally, hence widely used. The irregular and extensive use of Miltefosine in the past few years have resulted in development of resistance against this drug (Singh et al. 2012; Ghorbani and Farhoudi 2018). Miltefosine is being transported inside parasite through P4ATPase-CDC50 protein complex while the efflux is carried out by an ATP binding cassette (ABC) transporter protein transporting phospholipid, i.e., P-glycoprotein (P-gp). Structurally, P4ATPase consists of 4 domains, a transmembrane domain (TMD) having 10 transmembrane helices (TMH), an actuator domain (A-domain), a phosphorylating domain (P-domain) and a nucleotide binding domain (NBD). CDC50 has 2 TMH with a large exocyttoplasmic loop. Structural composition of P-gp includes 2 TMDs each having 6 TMH and 2 NBDs (Dawson and Locher 2006; López-Marqués et al. 2015). Alteration in the functional activity of these two proteins have resulted in the development of resistance by the Leishmanial parasite against Miltefosine (Croft et al. 2006; Singh and Mandlik 2015).

The transporters proteins play a pivotal role in the survival of intracellular pathogens inside hosts. They mediate the entry of essential nutrients inside cell, uptake and efflux of ions, maintain homeostasis by regulating concentration of metabolites, and particularly in survival of pathogens by extrusion of drugs, toxins and xenobiotics (Lam et al. 2011). As transporter proteins play such an important role in survival of organisms, hence a lot of studies have been focused on them, especially on the drug transporters. Since most of the transporter proteins are membrane proteins, hence a large fraction of reports have been on the evolution of the transmembrane helices amongst various superfamilies'. Lam *et al.* have proposed a pathway of evolution of transmembrane helices (TMH) in drug – metabolite transporter proteins starting from having 2Transmembrane Segments (TMS) to 4-5TMS and reaching up to 10TMS (Lam et

al. 2011). In another study by MH Saier, a pathway of the evolution of transporter proteins in all organisms was proposed which starts from a simple peptide channel comprising of 1-3 helices to evolving to much complex secondary carriers, primary active transporters as well group translocators comprising of 12-14 helices (Saier 2016).

ATP binding cassette (ABC) transporter proteins constitute one of the largest superfamily of the transporter proteins and are present in all forms of life where gene duplication has played an important role in the evolution of new genes (Jones and George 2004). Gene duplication events have resulted in the evolution of many ABC genes in vertebrates and most of them are associated with various diseases as well as development of resistance (Moitra and Dean 2011). Saurin W. *et al.* retrieved 197 sequences of ABC transporters homologues to Maltose/maltodextrin import ATP-binding protein (MalK protein), a bacterial transporter. The phylogenetic analysis of these transporter proteins showed a clear divergence in the evolution of ABC importers and ABC exporters and subsequently they proposed a model for studying the generation and diversification of ABC systems (Saurin *et al.* 1999). Multigene families like those of ABC transporters having ATP-binding cassette, are subjected to point mutations or duplication events resulting in either few members to be nonfunctional or evolution of new members suggesting that almost all the members have emerged from a common ancestor. Based on this ideology, 68 members of ABC family have been identified from previously known 6 genes in *Dictyostelium discoideum* and other representatives of each family of ABC superfamily. This number is quite more than those present in either yeast or human which are amongst the best characterized genomes (Anjard and Loomis 2002). Evolutionary analysis have helped in establishing the connection of Cystic fibrosis transmembrane conductance regulator (CFTR) protein being a member of ABC-transporter superfamily in spite of being an ATP-dependent chloride channel. Although having the functional divergence, CFTR belongs to ABCC subfamily based on the structural similarity and domain rearrangement (Jordan *et al.* 2008). Igarashi Y. *et al.* have studied 126 prokaryotic ABC transporters on the basis of nucleotide binding domain and classified them into 11 clusters. This have facilitated in understanding functional divergence and identification of new putative functions of ABC genes and how they are related in both eukaryotes and prokaryotes (Igarashi *et al.* 2004). Another large superfamily of membrane transport proteins utilizing ATP-hydrolysis energy is P-type ATPases. Axelsen and Palmgren analyzed 159 P-type ATPase sequences and classified them in 5 major classes based on their substrate similarity instead of sequence

similarity (Axelsen and Palmgren 1998). Moller *et al.* studied the evolutionary divergence of P5 subfamily of P-type ATPases in eukaryotes which acts as secretory pumps in Endoplasmic reticulum (Møller et al. 2008). Similarly substrate specificities based evolutionary analysis was done for maltose transporters in *Thermotoga maritima* (Nanavati et al. 2005). A few research groups have pointed the evolutionary relationship between different subfamilies of the superfamily on the basis of their structure as well as function i.e., their substrate specificity.

Here, we are providing insight on the sequential conservedness of a particular member of a transporter proteins superfamily and their presence in various organisms. To the best of our knowledge, this is the first report of its kind where evolutionary divergence of Miltefosine transporter proteins in *Leishmanial* species is being discussed. Since, Miltefosine is the only orally administered drug available for the treatment of Leishmaniasis, and the protozoan parasite is developing resistance with an alarming rate against this drug. We aim to study the evolutionary relationship of these transporter proteins to find any possible solutions for combating drug resistance. In one of our previous published dataset, we did the conservedness analysis and compared these Miltefosine transporter proteins between host (*Homo sapiens*) and parasite (*L. major*) to identify parasite specific motifs against which peptides were designed in order to allosterically modulate activity of transporter proteins to reverse the resistance effect (Kabra et al. 2020). Using the same data set, we did reanalysis aiming at the evolutionary divergence of these 3 transporter proteins across different kingdoms of life.

2. Methodology

2.1. Sequence retrieval

Sequences for P-gp, P4ATPase and CDC50 were extracted manually from National Centre for Biotechnology Information (NCBI, <https://www.ncbi.nlm.nih.gov/>) in FASTA format. All the annotated, complete and fully characterized sequences available in the database (as of December, 2015) ranging from all five kingdoms of life (Monera, Protista, Fungi, Animalia and Plantae) were retrieved. The sequences of P-gp, P4ATPase and CDC50 proteins of *L. major* (parasite) and Human (host) serves as the basis to identify homologous sequences in other organisms via BLASTp and then apparently their retrieval from NCBI following the accession IDs. As far as, multiple copy or paralogs are concerned we looked for those variants whose substrates were

phospholipids since Miltefosine is a phosphatidylcholine analog and this substrate information was available only in annotated sequences. The sequences ranged from 181 to 1503 amino acids in P-gp; 776 residues to 2061 residues for P4ATPase and from 245 residues to 1537 residues in CDC50 protein. Multiple Sequence Alignment (MSA) for each protein was done using Clustal Omega and the output was saved in Nexus format (Sievers et al. 2011). Clustal omega is a tool, available online at <https://www.ebi.ac.uk/Tools/msa/clustalo/>, for aligning three or more nucleotide or protein sequences. This tool is provided and maintained by EMBL-EBI. It uses either seeded guide trees or Hidden Markov Model (HMM) profile-profile technique to generate MSA.

2.2. Phylogenetic tree construction

Phylogenetic Tree was constructed through MrBayes v3.2.6 (Ronquist et al. 2012) using Metropolis Coupled Markov Chain Monte Carlo (MCMCMC) method. MrBayes is open source software available at <https://nbisweden.github.io/MrBayes/download.html>. It is based on bayesian interference and uses Markov Chain Monte Carlo (MCMC) method to estimate posterior distribution of model parameters. MrBayes accepts multiple sequences aligned in Nexus format as an input file. The amino acid model parameter is set as 'mixed' and default MCMCMC method was used for tree construction considering two runs each having 4 chains. The analysis was terminated when the average standard deviation of split frequency reached below 0.01. It took 2000000, 1150000 and 2900000 generations respectively for P-gp, P4ATPase and CDC50 sequences to get a converged tree with the average standard deviation of split frequency below 0.01. The constructed trees were visualized using FigTree v1.4.2 available at <http://tree.bio.ed.ac.uk/software/figtree/>.

2.3. Identification of specificity determining positions

JDet software was used to analyze aligned sequences and identify the positions determining specificity through Xdet method (Muth et al. 2012). In the graphical user interface, we used predefined Xdet method where the command '`.\conf\Maxhom_McLachlan.metric`' was run. For the Xdet analysis, the threshold cutoffs considered were 0.6 for Xdet and 2.5 for entropy. Using JDet interface, the sequences were represented in logo format.

3. Results

Fully annotated sequences of P-glycoprotein (P-gp), P4ATPase and CDC50 proteins were retrieved from NCBI (as available till 2015) from archaea to mammalia as well as plantae. The sequences homologues to *Leishmania major* were considered. The sequences of these proteins retrieved for various organisms were classified and represented in fig 1. A total of 96, 207 and 189 sequences were retrieved for P-gp, P4ATPase and CDC50 proteins. As can be seen, P4ATPase and CDC50 have been very well characterized in Fungi followed by Chordata, Protozoa and Nematelminthes. In order to study the evolutionary relationship and sequential conservedness amongst different organisms, phylogenetic tree was constructed for these three proteins, viz. P-gp, P4ATPase and CDC50.

3.1. P-gp

Fully annotated sequences of P-glycoprotein (P-gp) were retrieved from NCBI (as available till 2015) for 96 organisms ranging from archaea to mammalia as well as plantae. The retrieved organisms were classified as highlighted in the phylogenetic tree shown in fig 2. The Protozoans are shown in red color and includes 17 organisms, 1 organism (*Hydra vulgaris*) belongs to Cnidaria phylum, 4 Nematodes (Orange), 9 Arthropods (Cyan), 6 Molluscs (Pink), 33 Chordates (Blue), 7 Plants (Green), 6 Protists (Purple), 3 Bacteria (Yellow) and 9 Fungi (Grey) were included. A wide variety in the sequence length was observed, with *Loa loa*, a nematode, having the minimum length of 181 residues while the organism having longest P-gp protein was *Trypanosoma brucei* with 1503 amino acid residues.

Although P-gp is sequentially highly conserved, but it was seen to be more conserved amongst chordates and plants when compared to other organisms (fig S1 and S3). Amongst chordates, P-gp in human was observed to be closely related with African monkey (*Chlorocebus aethiops*) and Rhesus Monkey (*Macaca mulatta*). As expected the P-gp protein in Mammals is closely related to Aves which have diverged from an Amphibian, *Xenopus laevis*. The fishes are represented in purple color in fig S1, and *Petromyzon marinus* is found to be diverged from the other 4 organisms representing fish's class. Although, among chordates, *Ophiophagus Hannah*, a reptile, acts as outgroup. Amongst the lower degree of animals, comparatively less conservedness was observed. As represented in fig S2, protozoans were seen clustered in three different clades. One cluster comprising of 6 *Leishmania* species, 2 *Leptomonas* species and *Bodo saltans*, was seen to share similarity with an algae *Phytophthora sojae* suggesting the

transfer of genes horizontally. The P-gp protein of *Leishmania tarentolae* have diverged itself from other species of same genus and was found to share homology with Trypanosomatids. This sister taxa of this clade is a bacteria, *Streptomyces coelicolor*. The third cluster of protozoans consist of plasmodium and toxoplasma species. Even belonging to same order *Trypanosomatidae*, species belonging to genus *Leishmania* and *Leptomonas* share more similarity than *Trypanosomes*. Trypanosomes are more homologues to bacteria and protists and further to other protozoans like *Toxoplasma* and *Plasmodiums*. Similar kind of observations were also made by Hamilton et al. (please cite year) where they studied evolutionary relationship between *Trypanosomes* and *Leishmania* on the basis of few genes like small subunit ribosomal RNA (SSU rRNA), glycosomal glyceraldehyde phosphate dehydrogenase (gGAPDH), spliced leader (SL) RNA gene etc. They observed that although belonging to the same order *Trypanosomatidae*, *Trypanosomes* and *Leishmania* have diverged on the genomic level (Maslov et al. 1996; Hamilton et al. 2004). P-gp also seems to be one of the genetic marker in differentiating these two. In a detailed review by Lukes and group, they considered genome sequence analysis over morphological features in establishing evolutionary relationship between various members of Kinetoplastidae family (Lukeš et al. 2014). Seven species of angiosperms representing the Plantae Kingdom is seen to be conserved with *Sorghum bicolor*, a monocot, being diverged from Eudicots, *Genlisea aurea* being an exception (fig S3). In the evolutionary tree, molluscs are closely related to chordates exception being *Anadara sativa* and *Tegillarca granosa* which are comparatively closer to nematodes. Although nematodes display more homology with arthropods belonging to insecta (*Lygus hesperus*) and arachnida (*Tetranychus cinnabarinus*) class with *Loa loa* being an exception. The phylogenetic tree viewed as a whole for the 96 organisms' displays *Candida albicans*, a fungus, as outgroup. The tree is further subdivided in two clades, one having arthropods (mainly crustaceans and insects) and other clade comprising of rest of the organisms.

For identifying the positions determining specificity (SDPs), JDet was used where *C. albicans* served as the master sequence. Within that, Xdet method was used with a cut off of 0.6 and entropy cutoff as 2.5. From the various position specific residues obtained, we were particularly interested in those residues which directly interacts with P-gp substrate, i.e., Miltefosine. The Miltefosine interacting residues for *L. major* P-gp, which have been previously determined by us (Kabra et al. 2020), are S359, A360, M364, L397, R941, T1008, I1011, Y1012, S1041, I1044

and Q1049. Of these 11 residues, only one residue, L397, was obtained post Xdet analysis. But, in case of protozoan species only, 8 residues (S359, A360, M364, L397, R941, S1041, I1044 and Q1049) were identified as SDPs. The Miltefosine interacting residues for Pgp (all organisms and only protozoans) pre and post Xdet analysis are represented in fig 3.

3.2. P4ATPase

P4ATPase is a flippase belonging to type 4 sub-family of P-type ATPases. It mainly transports phospholipids from outer leaflet of membrane towards inner leaflet (cytosol). Fully annotated sequences for P4ATPase of 207 organisms were retrieved from NCBI, as of 2015. Of these, 80% are fungi and the rest 20% comprises of protista, protozoa, cnidarian, nematodes, arthropods and chordates (fig 4). The minimum and maximum length of P4ATPase protein was observed in fungi with a *Saccharomyces*, *Sugiyamaella lignohabitans*, having 776 residue length and *Taleromyces cellulolyticus*, a *Eurotiomycetes*, having the protein length of 2061 residues. It's evident that this subfamily of P-type ATPases have been well characterized in Fungi. As demonstrated in fig 4, *Trichomonas vaginalis*, a protozoan parasite, was seen as outgroup. The protozoans were clustered together and similar to that of P-gp proteins, *Leishmania* species and *Leptomonas* species were closely related. Amongst, *Leishmania* species, *L. braziliensis* and *L. panamensis* have diverged from other species with *L. amazonensis* and *L. mexicana* being sister taxa and *L. major* and *L. infantum* being sister taxa (fig S4). Although belonging to Trypanosomatidae family, trypanosomes were observed to be deviated from *Leishmania* genus as well as *Leptomonas* genus. Our observation from P-gp and P4ATPase protein backs up the evolutionary relationship established from the small and large subunit rRNAs of different Trypanosomatids as studied by Maslov and group (Maslov et al. 1996). This also validates the consideration of genes/proteins over morphological features while talking about the evolution of these protozoan parasites. In an elaborated review by Lukes and group, they pointed towards more possibility of gene duplication events in trypanosomes having approximately 11000 protein coding genes when compared with *Leptomonas* and leishmanials having approximately 8200 protein coding genes (Lukeš et al. 2018). The three protists in which fully annotated P4ATPase sequence have been characterized and retrieved, *Naegleria gruberi* was found to share more similarity with protozoans compared to the other two protists, *Acanthamoeba castellanii* and *Dictyostelium lacteum*. The Cnidarians and arthropods along with a few chordates (*Bos taurus*

and *Macaca mulatta*) were observed to be diverged from a common ancestor with *Trichuris trichiura* (a nematode) being an exception (fig S5). The rest of the chordates follow an independent lineage with P4ATPase in humans sharing similarity with rodents. As foresaid, Fungi P4ATPases have been well characterized and out of sequences belonging to 166 fungal species being retrieved; 117 belongs to Ascomycota division, 39 to Basidiomycota, 3 each to Zygomycota and Chytridiomycota and 1 each to Entomophthoromycota, Mucoromycota, Glomeromycota and Deuteromycota (fig S6). The evolutionary divergence suggests Ascomycots and Basidiomycots have diverged from a common ancestor and are closely related to Glomeromycots. Amongst these fungal species, *Rozella allomycis*, a Chytridiomycot, was believed to have been evolved first in comparison to other members of same division like *Spizellomyces punctatus* and *Gonapodya prolifera*.

For the SDPs in P4ATPase protein, *Trichomonas vaginalis* served as the reference sequence. K121, S365, V372 and N870 residues of *L. major* P4ATPase are interacting with Miltefosine (Kabra et al. 2020). All the four residues were identified as SDPs post Xdet analysis whether all 207 organisms or just 15 protozoans were considered. These residues in sequence logo format are shown in fig 5. It was observed that residues 'K', 'S' and 'N' were highly conserved in all organisms while 'V' shared place with 'M' and 'L' (majorly in protozoans). This conservedness of Miltefosine interacting residues was expected since P4ATPase proteins are specifically lipid transporters and Miltefosine is an analog of phosphatidylcholine.

3.3. CDC50

CDC50 acts as β -subunit required for the functioning as well as membrane localization of P4ATPase (López-Marqués et al. 2015). The heterodimeric complex of P4ATPase-CDC50 proteins together is responsible for the influx of their substrates towards the cytosol. As of 2015, fully annotated sequences of CDC50 protein belonging to 189 organisms were retrieved from NCBI. Similar to P4ATPase, CDC50 proteins are also very well characterized in Fungi with fully annotated sequences retrieved from 117 organisms comprising of 62% of the total organisms considered for phylogenetic analysis (fig 6). Unlike P-gp and P4ATPase, in terms of sequence length CDC50 protein was observed to be much conserved with a range from 245 residues (*Drosophila sechellia*, an arthropod) to 612 residues (*Trichinella sp. T9*, a nematode) with *Amazona aestiva*, a chordate, being an exception with protein length of 1537 residues. As

observed in P4ATPase, in case of CDC50 proteins also a protozoan was found to be acting as outgroup, i.e., *Trypanosoma grayi*. The rest of the protozoans, specifically kinetoplastids, have evolved together with CDC50 protein of *L. major* being diverged from other leishmanial species (fig S7). Arthropods, nematodes and molluscs were observed to have diverged from a point with chordates being the other branch of bifurcation which further seems to have diverged from platyhelminth, *Schistosoma japonicum*. *Stegodyphus mimosarum*, an African spider, shares more sequence similarity with a mollusc, *Crassostrea gigas*, rather than with the members of same phylum and believed to have evolved earlier than other arthropods and nematodes (*Trichinella* species). Another arthropod, *Artemia sinica*, deviates from other members by being homologous to *Trichinella* species (fig S8). Of the two protists for which annotated CDC50 sequences were available, *Capsaspora owczarzaki*, have evolved much earlier and diverged from the group of taxas which branch into platyhelminthes, nematodes, arthropods and chordates; while all these groups have been diverged from the other protist, *Thecamonas trahens* (fig S9). Coming to chordates, sequences of CDC50 proteins from 24 chordates were retrieved and their evolutionary relationship is shown in fig S10. The Mammals and Aves form the sister clades. Unlike, P-gp and P4ATPase, CDC50 protein of humans shares similarity with a Sumatran orangutan, *Pongo abelii*, which is also a critically endangered species. Although belonging to same genus, *Xenopus tropicalis* and *Xenopus laevis*, two different species of aquatic frog, an amphibian, have diverged a lot with the former sharing similarity with *Scleropages formosus*, a fish, and the later acting as an outgroup to other chordates (fig S10). Of the 117 fungal species, 88 belongs to Ascomycota division, 25 to Basidiomycota, 2 Zygomycota and 1 each to Chytridiomycota and Entomophthoromycota (fig S11). Amongst them, *Gonapodya prolifera*, a Chytridiomycot, was perceived to be outgroup. Of the two zygomycots, *Mucor ambiguous*, shares more resemblance with entomophthoromycot, *Conidiobolus coronatus* rather than its fellow member i.e., *Lichtheimia corymbifera*. The sister clade of these three comprises of basidiomycots.

In case of *L. major*'s CDC50 protein, F151, Q159, M235 and W236 interacts directly with Miltefosine (Kabra et al. 2020). When all organisms were considered, none of these residues fit in the criteria of Xdet analysis, but when only protozoans were considered, all these miltefosine interacting residues acted as SDPs post Xdet analysis (fig 7). Unlike P4ATPase, in CDC50 proteins, Miltefosine interacting residues are far diverged when all organisms were considered

and hence no residues lied in Xdet analysis. In contrast, these residues showed their dominance in protozoans.

4. Discussion

The transporter proteins holds utmost position in any cell or organisms as they not only help in the uptake of nutrients, small molecules or ions but are also responsible for the extrusion of toxic substances, xenobiotics, drugs etc. Since different transporters are specific for their substrates hence, these proteins have remained highly conserved in the due course of evolution. Here in this project, we have tried to study the evolutionary relationship of Miltefosine transporter proteins in *Leishmania major* with the homologous proteins present in other organisms. P4ATPase and CDC50 proteins are responsible for the uptake of Miltefosine drug while P-gp is responsible for the extrusion of drug outside parasite. On analyzing the completely annotated sequences of these proteins amongst four kingdoms of life, i.e., Protista, Fungi, Animalia and Plantae; we found that these proteins have shown quite a high degree of conservedness across different classes of organisms. Most of the organisms belonging to a phylum or division were found to be clustered together with very exceptional cases of horizontal gene transfer. Through these analysis it was realized that Type IV – P-type ATPase have been studied and characterized more than the ABC transporter, P-glycoprotein (P-gp) especially amongst fungal species. The identification of SDPs through Xdet analysis confirmed that drug binding residues were more conserved in P4ATPase rather than in its accessory protein CDC50 or in P-gp. But when a particular class of organisms was considered, like in protozoans, the three Miltefosine transporter proteins, i.e., P-gp, P4ATPase and CDC50; were found to exhibit more conservedness. Similar observation was observed through phylogenetic analysis too. In spite of being kinetoplastids, Trypanosomes and Leishmanias were observed to have diverged from each other sequentially. But they were found to possess same substrate specificity as various Trypanosome strains like *T. cruzi*, *T. copemani*, etc. were found to be susceptible of Miltefosine (Saraiva et al. 2002; Luna et al. 2009; Botero et al. 2017). In nutshell, these phylogenetic analysis as well as finding of SDPs not only helped in understanding the evolutionary pattern of a particular protein in an organism but can also help in targeting a protein in an organism specific manner to either increase or decrease their functional activity. Using this concept, the sequential and positional conservedness of P-gp, P4ATPase and CDC50 proteins between Human (host) and *Leishmania major* (parasite) have helped in

designing parasite specific peptides that allosterically modulate activity of Miltefosine transporter proteins resulting in reversal of Miltefosine resistant parasites to the drug sensitive form (Kabra et al. 2020). Similar strategy can be implied for other proteins in same or different organisms. Thus, the study of evolutionary relationship of a probable target protein across diverse organisms may provide probable and long-lasting solutions in curing a disease or fighting against resistance.

Acknowledgement

Ritika Kabra would like to acknowledge her SRF (Senior Research Fellowship) from Department of Biotechnology (DBT/JRF/14/AL/22/3544), Ministry of Science and Technology, Government of India. Authors would also like to thank the Director, National Centre for Cell Science (NCCS) Pune, for supporting the Bioinformatics and High Performance Computing Facility (BHPCF) at NCCS Pune, India.

Authors Contribution:

All authors contributed to study design and conception. Ritika Kabra: data curation, methodology implementation and writing the original draft. Shailza Singh: Initialization and conceptualization, data curation, supervision, writing and editing the original draft. All authors agree with the submission of the final version of the manuscript.

References

- Anjard C, Loomis WF (2002) Evolutionary analyses of ABC transporters of *Dictyostelium discoideum*. *Eukaryot Cell*. <https://doi.org/10.1128/EC.1.4.643-652.2002>
- Axelsen KB, Palmgren MG (1998) Evolution of substrate specificities in the P-type ATPase superfamily. *J Mol Evol*. <https://doi.org/10.1007/PL00006286>
- Botero A, Keatley S, Peacock C, Thompson RCA (2017) In vitro drug susceptibility of two strains of the wildlife trypanosome, *Trypanosoma copemani*: A comparison with *Trypanosoma cruzi*. *Int J Parasitol Drugs Drug Resist*. <https://doi.org/10.1016/j.ijpddr.2016.12.004>

- Croft SL, Sundar S, Fairlamb AH (2006) Drug resistance in leishmaniasis. *Clin Microbiol Rev* 19:111–26. <https://doi.org/10.1128/CMR.19.1.111-126.2006>
- Dawson RJP, Locher KP (2006) Structure of a bacterial multidrug ABC transporter. *Nature* 443:180–185. <https://doi.org/10.1038/nature05155>
- Ghorbani M, Farhoudi R (2018) Leishmaniasis in humans: drug or vaccine therapy? *Drug Des Devel Ther* 12:25–40. <https://doi.org/10.2147/DDDT.S146521>
- Hamilton PB, Stevens JR, Gaunt MW, et al (2004) Trypanosomes are monophyletic: Evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. *Int J Parasitol*. <https://doi.org/10.1016/j.ijpara.2004.08.011>
- Herwaldt BL (1999) Leishmaniasis. *Lancet* 354:1191–1199. [https://doi.org/10.1016/S0140-6736\(98\)10178-2](https://doi.org/10.1016/S0140-6736(98)10178-2)
- Igarashi Y, Aoki KF, Mamitsuka H, et al (2004) The evolutionary repertoires of the eukaryotic-type ABC transporters in terms of the phylogeny of ATP-binding domains in eukaryotes and prokaryotes. *Mol Biol Evol* 21:2149–60. <https://doi.org/10.1093/molbev/msh226>
- India National vector Borne Diseases Control Programme (2017) Accelerated Plan for Kala-azar Elimination 2017. 1–94
- Jones PM, George AM (2004) The ABC transporter structure and mechanism: perspectives on recent research. *Cell Mol Life Sci* 61:682–99. <https://doi.org/10.1007/s00018-003-3336-9>
- Jordan IK, Kota KC, Cui G, et al (2008) Evolutionary and functional divergence between the cystic fibrosis transmembrane conductance regulator and related ATP-binding cassette transporters. *Proc Natl Acad Sci U S A*. <https://doi.org/10.1073/pnas.0806306105>
- Kabra R, Ingale P, Singh S (2020) Computationally designed synthetic peptides for transporter proteins imparts allostericity in Miltefosine resistant *L. major*. *Biochem J* 477:2007–2026. <https://doi.org/10.1042/BCJ20200176>
- Lam VH, Lee J-H, Silverio A, et al (2011) Pathways of transport protein evolution: recent

advances. *Biol Chem* 392:5–12. <https://doi.org/10.1515/BC.2011.018>

López-Marqués RL, Poulsen LR, Bailly A, et al (2015) Structure and mechanism of ATP-dependent phospholipid transporters. *Biochim Biophys Acta - Gen Subj* 1850:461–475. <https://doi.org/10.1016/j.bbagen.2014.04.008>

Lukeš J, Butenko A, Hashimi H, et al (2018) Trypanosomatids Are Much More than Just Trypanosomes: Clues from the Expanded Family Tree. *Trends Parasitol* 34:466–480. <https://doi.org/10.1016/j.pt.2018.03.002>

Lukeš J, Skalický T, Týč J, et al (2014) Evolution of parasitism in kinetoplastid flagellates. *Mol. Biochem. Parasitol.*

Luna KP, Hernández IP, Rueda CM, et al (2009) In vitro susceptibility of *Trypanosoma cruzi* strains from Santander, Colombia, to hexadecylphosphocholine (miltefosine), nifurtimox and benznidazole. *Biomedica*. <https://doi.org/10.7705/biomedica.v29i3.15>

Maslov DA, Lukes J, Jirku M, Simpson L (1996) Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: Implications for the evolution of parasitism in the trypanosomatid protozoa. *Mol Biochem Parasitol*. [https://doi.org/10.1016/0166-6851\(95\)02526-X](https://doi.org/10.1016/0166-6851(95)02526-X)

Moitra K, Dean M (2011) Evolution of ABC transporters by gene duplication and their role in human disease. *Biol. Chem.*

Møller AB, Asp T, Holm PB, Palmgren MG (2008) Phylogenetic analysis of P5 P-type ATPases, a eukaryotic lineage of secretory pathway pumps. *Mol Phylogenet Evol* 46:619–34. <https://doi.org/10.1016/j.ympev.2007.10.023>

Muth T, García-martín JA, Rausell A, et al (2012) JDet: Interactive calculation and visualization of function-related conservation patterns in multiple sequence alignments and structures. *Bioinformatics* 28:584–586. <https://doi.org/10.1093/bioinformatics/btr688>

Nanavati DM, Nguyen TN, Noll KM (2005) Substrate specificities and expression patterns reflect the evolutionary divergence of maltose ABC transporters in *Thermotoga maritima*. *J*

Bacteriol. <https://doi.org/10.1128/JB.187.6.2002-2009.2005>

Ronquist F, Teslenko M, Van Der Mark P, et al (2012) MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61:539–542. <https://doi.org/10.1093/sysbio/sys029>

Saier MH (2016) Transport protein evolution deduced from analysis of sequence, topology and structure. *Curr Opin Struct Biol* 38:9–17. <https://doi.org/10.1016/j.sbi.2016.05.001>

Saraiva VB, Gibaldi D, Previato JO, et al (2002) Proinflammatory and cytotoxic effects of hexadecylphosphocholine (miltefosine) against drug-resistant strains of *Trypanosoma cruzi*. *Antimicrob Agents Chemother*. <https://doi.org/10.1128/AAC.46.11.3472-3477.2002>

Saurin W, Hofnung M, Dassa E (1999) Getting in or out: Early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. *J Mol Evol*. <https://doi.org/10.1007/PL00006442>

Sievers F, Wilm A, Dineen D, et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539. <https://doi.org/10.1038/msb.2011.75>

Singh N, Kumar M, Singh RK (2012) Leishmaniasis: Current status of available drugs and new potential drug targets. *Asian Pac J Trop Med* 5:485–497. [https://doi.org/10.1016/S1995-7645\(12\)60084-4](https://doi.org/10.1016/S1995-7645(12)60084-4)

Singh S, Mandlik V (2015) Structure based investigation on the binding interaction of transport proteins in leishmaniasis: insights from molecular simulation. *Mol BioSyst* 11:1251–1259. <https://doi.org/10.1039/C4MB00713A>

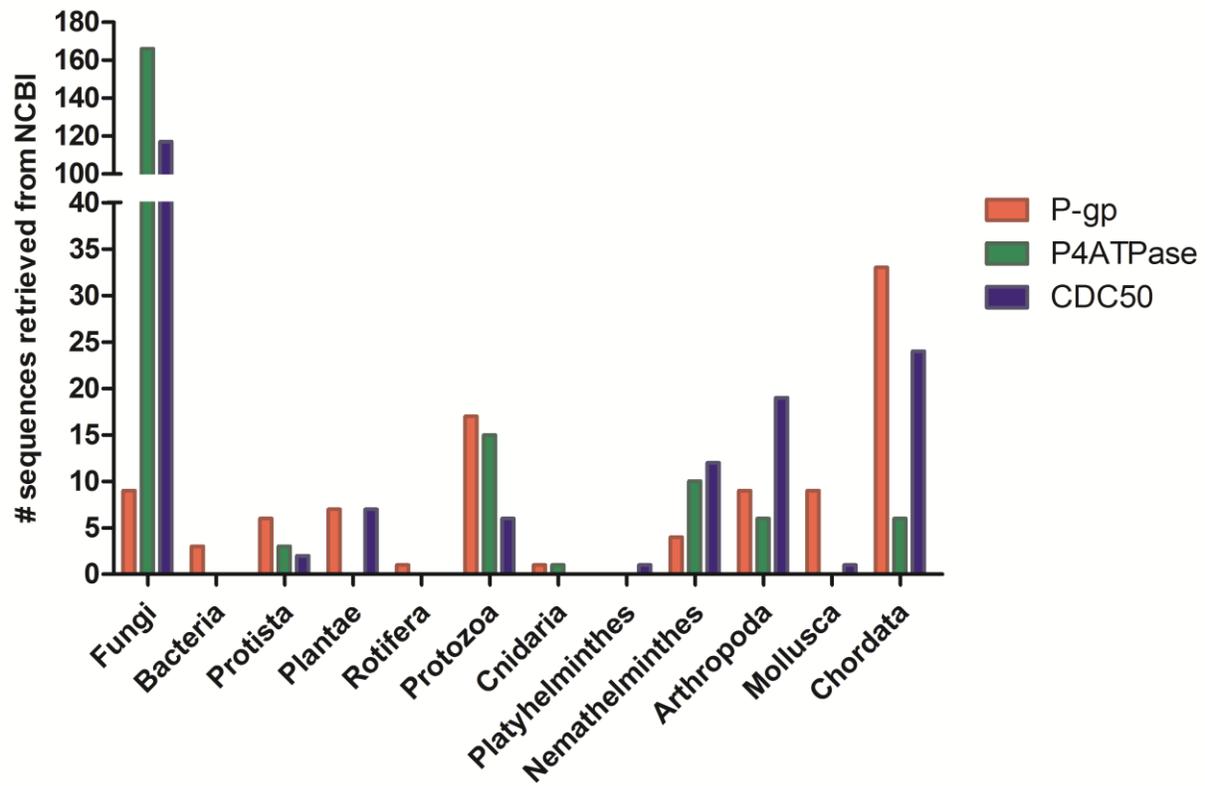


Fig 1: Classification of P-gp, P4ATPase and CDC50 sequences retrieved from NCBI.

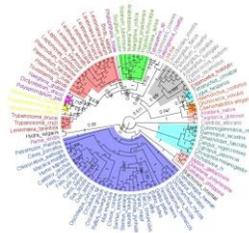


Fig 2: Phylogenetic tree construction and evolutionary analysis of P-gp proteins of 96 organisms. Color code: Red – Protozoa; Blue – Chordata; Grey – Fungi; Green – Plants; Cyan – Arthropoda; Pink - Mollusca; Orange – Nematoda; Yellow- Bacteria; Purple – Protista.

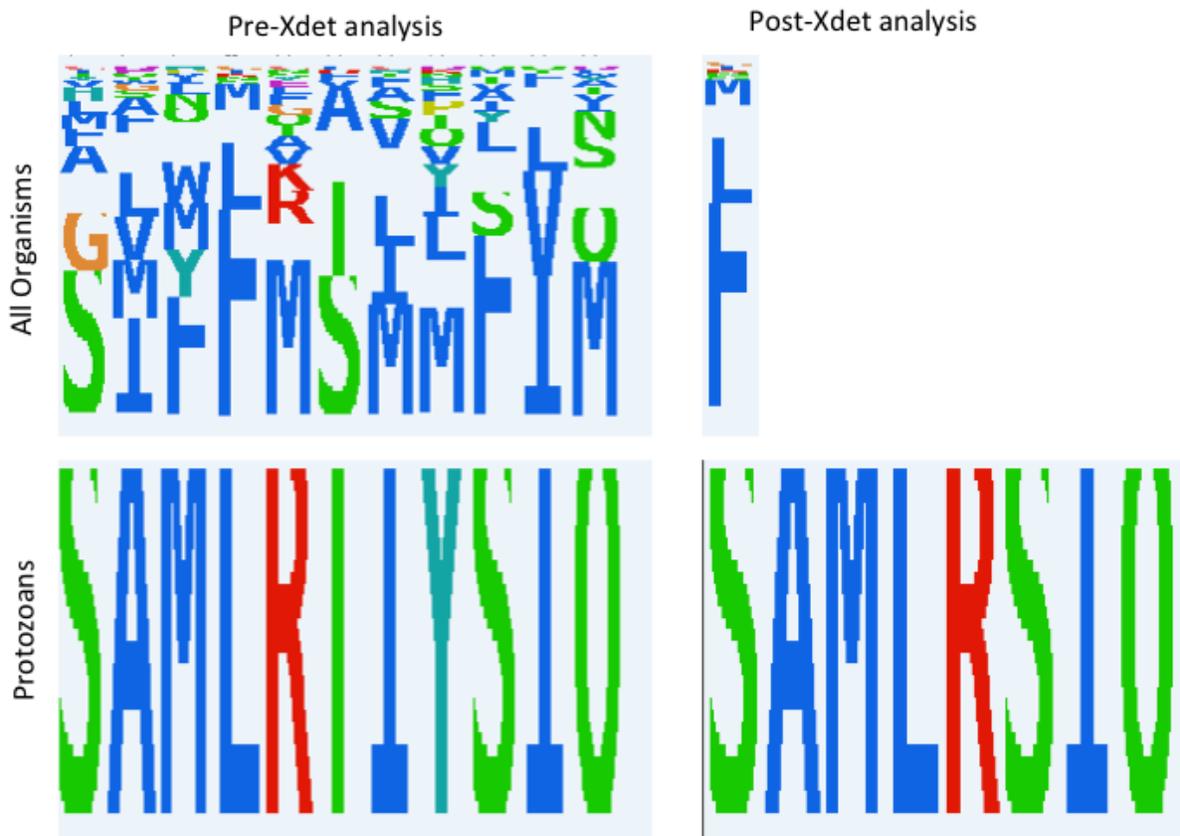


Fig 3: P-gp residues interacting with Miltefosine in sequence logo format. The upper panel shows sequence logo of P-gp from all 96 organisms while lower panel includes 17 protozoans. The left panel shows all the 11 residues interacting with Miltefosine. The right panel shows only those Miltefosine interacting residues that are also specificity-determining positions obtained through Xdet analysis with xdet cutoff as 0.6 and entropy cutoff as 2.5.

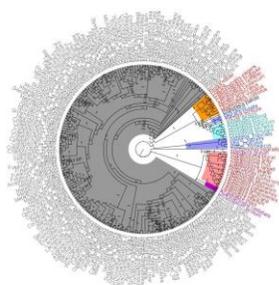


Fig 4: Phylogenetic tree construction and evolutionary analysis of P4ATPase proteins of 207 organisms. Color code: Red – Protozoa; Blue – Chordata; Grey – Fungi; Cyan – Arthropoda; Orange – Nematoda; Purple – Protista.

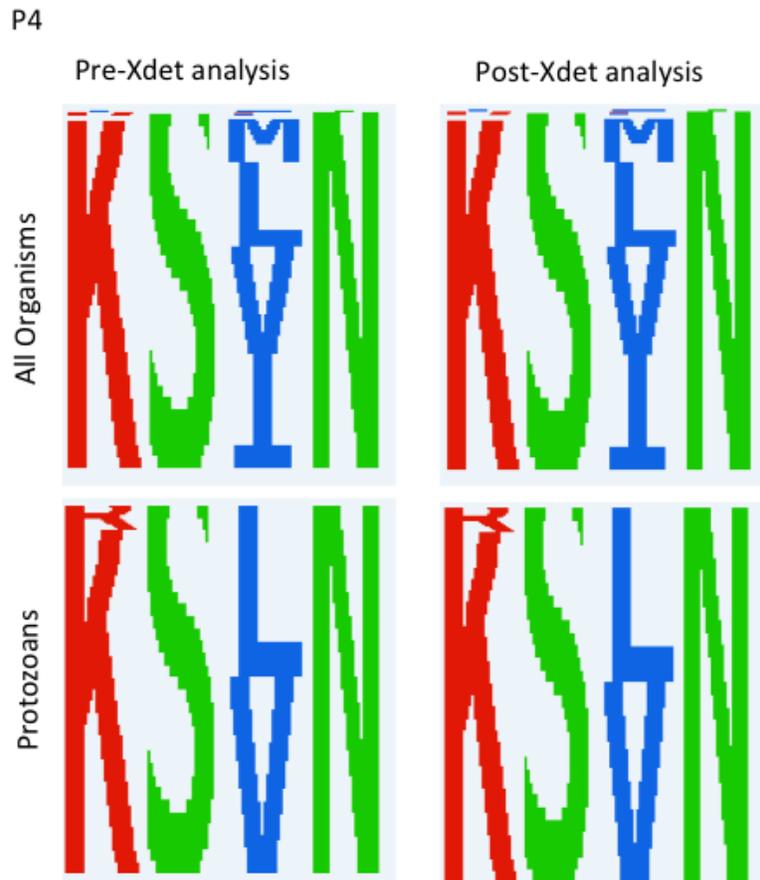


Fig 5: P4ATPase residues interacting with Miltefosine in sequence logo format. The upper panel shows sequence logo of P4ATPase from all 207 organisms while lower panel includes 15 protozoans. The left panel shows all the 4 residues interacting with Miltefosine. The right panel shows only those Miltefosine interacting residues that are also specificity-determining positions obtained through Xdet analysis with xdet cutoff as 0.6 and entropy cutoff as 2.5 (in this case all 4 residues are identified as SDPs).

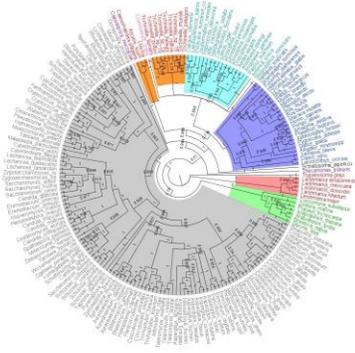


Fig 6: Phylogenetic tree construction and evolutionary analysis of CDC50 proteins of 189 organisms. Color code: Red – Protozoa; Blue – Chordata; Grey – Fungi; Cyan – Arthropoda; Orange – Nematoda; Purple – Protista; Green – Plants; Dark Green: Platyhelminthes and Pink: Mollusc.

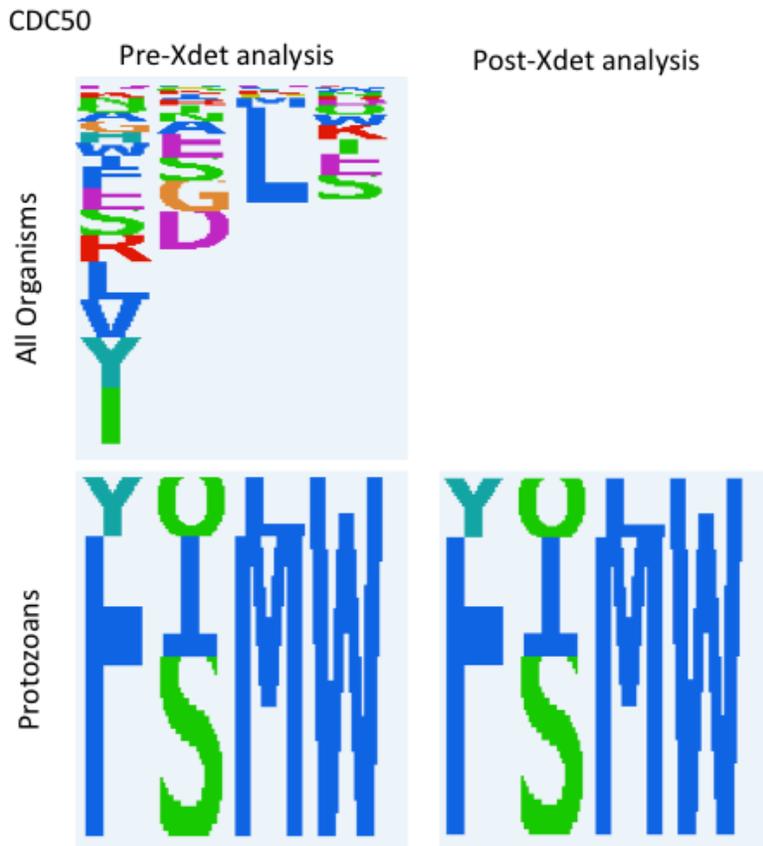


Fig 7: CDC50 residues interacting with Miltefosine in sequence logo format. The upper panel shows sequence logo of CDC50 from all 189 organisms while lower panel includes 6 protozoans. The left panel shows all the 4 residues interacting with Miltefosine. The right panel shows only those miltefosine interacting residues that are also specificity-determining positions obtained through Xdet analysis with xdet cutoff as 0.6 and entropy cutoff as 2.5.