Choke point analysis of metabolic pathways in *E.histolytica*: A computational approach for drug target identification

Shailza Singh[1,*], Balwant Kishen Malik [2], Durlabh Kumar Sharma [1]

[1]Center for Energy Studies, Indian Institute of Technology Delhi, Hauz Khas, New Delhi-110016, India; [2]Institute of Genomics and Integrative Biology, CSIR, Mall Road , Delhi-110007, India; Shailza Singh[*]- Email: shailza_iitd@yahoo.com; Phone: 091-11-26591256; Fax: 091-11-26581121; [*]Corresponding author

Abstract:
With the *Entamoeba* genome essentially complete, the organism can be studied from a whole genome standpoint. The understanding of cellular mechanisms and interactions between cellular components is instrumental to the development of new effective drugs and vaccines. Metabolic pathway analysis is becoming increasingly important for assessing inherent network properties in reconstructed biochemical reaction networks. Metabolic pathways illustrate how proteins work in concert to produce cellular compounds or to transmit information at different levels. Identification of drug targets in *E. histolytica* through metabolic pathway analysis promises to be a novel approach in this direction. This article focuses on the identification of drug targets by subjecting the *Entamoeba* genome to BLAST with the e-value inclusion threshold set to 0.005 and choke point analysis. A total of 86.9 percent of proposed drug targets with biological evidence are chokepoint reactions in *Entamoeba* genome database.

Keywords: *Entamoeba*; metabolic pathway; chokepoint

Background:
The amitochondrial protist *Entamoeba histolytica,* lacking both mitochondria and hydrogenosomes [1] is estimated to infect up to 10 percent of the world's population. Fifty million cases of amoebic dysentery and liver abscess are reported each year. [2] Between 40 and 100 thousand people die of amoebiasis each year. [3] Among parasitic infections, amoebiasis caused by *E. histolytica* ranks third worldwide, behind malaria and schistosomiasis, in lethal infections.

With the *Entamoeba* genome essentially complete [4] the organism can be studied from a whole genome standpoint. The knowledge of cellular mechanisms and interactions between cellular components is instrumental to the development of new effective drugs and vaccines. Metabolic pathways illustrate how proteins work in concert to produce cellular compounds or to transmit information at different levels. The view provided by the *E. histolytica* genome sequence is remarkable as it assists in the reconstruction of its metabolism and in the development of new antiamoebic drugs through the identification of its enzymes. The metabolism of *E. histolytica* has been shaped by an influx of bacterial genes through lateral gene transfer. [4]

Metabolic pathway analysis is very useful as it allows determining the overall capacity, i.e., theoretical maximum yield, of a cellular system and studying effects of any genetic modification. However, pathway analysis of large and highly entangled metabolic networks meets the problem of combinatorial explosion of possible routes across the networks. Here we propose a method to cope with this problem by two methods. In the present work, it has been tried to focus on the identification of drug targets by subjecting the *Entamoeba* genome to BLAST with the e-value inclusion threshold set to 0.005 and choke point analysis.

Methodology:
Identification of potential drug targets through two methods
One feature that one can expect a good drug target to have is a lack of similarity to any human enzyme. The *E. histolytica* genome sequence is available from GenBank or from The Institute for Genomic Research website (http://www.tigr.org/tdb/e2k1/eha1/). KEGG [5] pathway database was used as a source of metabolic pathway information. Metabolic pathway identification numbers of the host *H. sapiens* and the pathogen *E.histolytica* were extracted from the KEGG database. Pathways which do not appear in the host but present in the pathogen according to KEGG database annotation have been identified as pathways unique to *E.histolytica* as compared to the host *H. sapiens*. The corresponding protein sequences were retrieved from the KEGG database. They were subjected to a BLASTp [6] search against the non-redundant database

with the e-value inclusion threshold set to 0.005. The search was restricted to proteins from *H. sapiens* through an option available in the BLAST program, which allows the user to select the organism to which the search should be restricted. In the current context, the objective is to find only those targets, which do not have detectable human homologues. Enzymes, which do not have hits below the e-value inclusion threshold of 0.005, were picked out as potential drug targets.

Choke point analyses
Cells, whether free-living or as components of an organism, are regularly exposed to stimuli and perturbations to which they have to respond in a coordinated fashion. Such responses may vary widely in the numbers and types of identifiable steps that begin with sensing the stimulus and end with a change in biochemistry or physiology. Metabolic reconstruction of the organism can help in this direction. It is the purpose of this article to introduce alternative measures for this task and hence, the choke point analysis was one such measure. Here the choke point analysis method of Yeh et al. [7] has been adopted. It is being described below:

To identify potential drug targets, a chokepoint analysis of the metabolic network of *E.histolytica* is performed. A "chokepoint reaction" is a reaction that either uniquely consumes a specific substrate or uniquely produces a specific product in the *Entamoeba* metabolic network (Figure 1). As described by Yeh and colleagues [7] it is expected that the inhibition of an enzyme that consumes a unique substrate result in the accumulation of the unique substrate which is potentially toxic to the cell and the inhibition of an enzyme that produces a unique product to result in the starvation of the unique product which potentially cripple essential cell functions. Thus, it is believed that chokepoint enzymes may be essential to the parasite and are therefore potential drug targets.

Chokepoint analysis has several advantages. First, it allows us to test the consistency between experimental data and assumptions about the organization and regulation of the biochemical pathway and of its interdependencies with other processes. Second, it can be used to predict the consequences of various mutations or inhibitors.

Results and discussion:
The targeting of metabolic pathways has several advantages on its own. Each step in the pathway is well validated as an essential function for pathogen growth. The target enzymes from the pathogen which are discarded and which share a similarity with the host proteins ensures that the targets have nothing in common with the host proteins, thereby, eliminating undesired host protein–drug interactions. Metabolic pathway analysis is becoming increasingly important for assessing inherent network properties in reconstructed biochemical reaction networks.
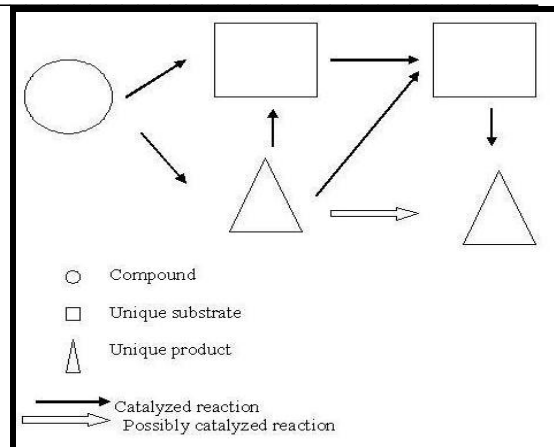

Figure 1: The thick arrows represent reactions that are catalyzed by enzymes, whereas the thin arrows represent reactions that are present (no evidence of the corresponding enzymes). While determining chokepoint reactions, only the catalyzed reaction are considered

Pathways unique to *E. histolytica* when compared to the host *H. sapiens*
An important question to be addressed while choosing potential drug targets is whether the biochemical pathway to be targeted is unique to pathogen. The three pathways addressed in this study: peptidoglycan biosynthesis, streptomycin biosynthesis, and thiamine metabolism which are unique to the pathogen *E.histolytica* are discussed. Enzymes, from these pathways, which do not show similarity to any of the host proteins, are listed in Table 1 (supplementary material).

Peptidoglycan biosynthesis
Entamoeba histolytica was found to grow normally without producing glutathione and the main enzymes of glutathione metabolism, indicating that glutathione is not essential for many eukaryotic processes. This parasitic amoeba is an unusual eukaryote whose special features may help define the crucial functions of glutathione in those eukaryotes that do use it. Since *Entamoeba histolytica* lacks mitochondria and the usual aerobic respiratory pathways, the finding that it grows without glutathione and other evidence support the hypothesis that a primary function of glutathione in eukaryotes involves protection against oxygen toxicity associated with mitochondria and suggest that eukaryotes may have acquired glutathione metabolism at the time that they acquired mitochondria. [8] If this hypothesis is true, then the genes for GSH biosynthesis in eukaryotes should resemble those from alpha-proteobacteria, the modern relatives of the mitochondrial progenitor. [9, 10]

Thiamine metabolism
The electrophoretic patterns of hexokinase and phosphoglucomutase strongly support the redefinition of pathogenic and nonpathogenic *Entamoeba histolytica* as two closely related species *Entamoeba histolytica* and *Entamoeba dispar*. [11] Thiamin pyrophosphokinase (TPK)

Hypothesis

catalyses the transfer of a pyrophosphate group from ATP to vitamin B1 (thiamin) to form the coenzyme thiamin pyrophosphate (TPP). Thus, TPK is important for the formation of a coenzyme required for central metabolic functions. The structure of thiamin pyrophosphokinase suggests that the enzyme may operate by a mechanism of pyrophosphoryl transfer similar to those described for pyrophosphokinases functioning in nucleotide biosynthesis. **[12]**

Streptomycin biosynthesis
Glycosylphosphatidylinositol (GPI)-anchored molecules such as cell surface Gal/GalNAc lectin and proteophosphoglycans of the protozoan parasite *Entamoeba histolytica* are thought to be involved in pathogenesis. **[13]** The cells of *E.histolytica* contained inorganic phosphate, pyrophosphate, nucleoside diphosphates, nucleoside triphosphates, NAD(P), phosphocholine, phosphoethanolamine, cytidine 5'-diphosphocholine and cytidine 5'- diphosphoethanolamine . The latter four compounds may act as intermediates in the salvage pathway for the synthesis of phosphatidylethanolamine and phosphatidylcholine. **[14]**

Carbohydrate metabolism
Studies have shown that *E.histolytica* grow or survive better in media containing glucose than in glucose deficient media, which is consistent with carbohydrates being the main energy substrates for anaerobic protists. Carbohydrates are the main source of energy for the parasite. The pathway involved in the fermentation of glucose includes some unusual enzymes that use pyrophosphate (PPi) rather than a nucleoside triphosphate as the phosphate donor. **[15]** Many of the genes encoding enzymes of the glycolytic pathway have been cloned and sequenced. Both D- glucose and D-galactose are transported and actively metabolized by *E.histolytica.* **[15]**

Lipid metabolism
Little is known about lipid metabolism in *Entamoeba*. Three phospholipase activities A1, A2 and L1 have been detected. **[16]** *Entamoeba* can also synthesise cholesterol and isoprenoids. **[17]** The only detectable polyamine present is putrescine at 9.5mM. **[18]** As it is present only in trace amounts in the growth medium, the amoeba must therefore synthesize it.

Protein metabolism
Protein metabolism in *Entamoeba* is not clearly elucidated. Amino acid utilization studies **[19]** showed that certain amino acids such as arginine, threonine, leucine, glutamine, phenylalanine and isoleucine could serve as an alternative energy source in the absence of glucose. The organism synthesizes cysteine, as cysteine synthase. **[20]**

Choke point analyses
If an enzyme catalyzes at least one chokepoint reaction, it is classified as a potential drug target. All the potential metabolic drug targets are listed in Table 2 (supplementary

material). To assess the usefulness of identifying chokepoint enzymes for proposing drug targets, chokepoints and nonchokepoints against proposed drug targets from the literature is compared. A complete literature search for proposed amoebiasis drug targets is attempted that were metabolic enzymes and met the criteria discussed above. Of the 23 proposed targets with biological evidence, 20 are chokepoints in Table 2 (supplementary material). A total of 86.9% of proposed drug targets with biological evidence are chokepoint reactions in *Entamoeba* genome database. Chokepoints may not be essential. One reason could be that they create unique intermediates to an essential product which are not essential themselves and finally, there could be chokepoint reactions that are not essential due to other pathways that achieve the same metabolic goal within the organism. One example could be blocking the reaction that has no deleterious effects on the parasite. **[7]**

Due to the high percentage of enzymes identified as choke points, one additional criteria observed in addition to being a choke point enzyme for identifying potential metabolic drug targets is that an enzyme not having isozymes would make it more likely to be a good drug target. This is because one enzyme would be easier to inhibit than a family of enzymes. As for example, methionine γ - lyase in *E. histolytica* has two isozymes **[40]** and according to the criteria outlined above, it has not been included in the list of potential drug targets identified in Table 2 (supplementary material).

There are further aspects on which the list of potential drug targets can be narrowed down. The drug should adversely affect the parasite but not the human host which means that if the drug target has a homologous enzyme in human, it should not be essential or have differential inhibition in human. In other way, it can be said that potential drug targets should be expressed in the human stages of the parasite. The two most promising concepts for pathway analysis focused here are closely related. Assessing metabolic systems by the set of extreme pathways can, in general, give misleading results owing to the exclusion of possibly important routes. A full assessment of the proposed listed steps will require intense further effort. It is to be expected that some experiments may be significant which will stimulate the next phase of amendments and refinements. As it stands, it is hoped to serve the scientific community as a starting point for further data collection and experimentation in concert with, and based on, pathway analysis.

Conclusion
The genome sequence of *E. histolytica* has provided greater in-depth facts revealing that many pathways normally present in eukaryotes are absent from this organism. In addition, by identifying choke point reactions, identification of enzymes has been done that are essential to the parasite's survival. There is an enrichment of drug targets in chokepoints as compared with non-chokepoints.

This leads to the conclusion that the classification of an enzyme as a chokepoint has some bearing on whether or not it would make a good drug target. The capabilities of the parasite to transport an accumulating metabolite out of the cell or a limiting metabolite into the cell have not been considered. It further limits the choke point analysis. The provisional targets which have been cited here needs to be examined further, both computationally and experimentally for these additional features. It can be said that biochemical reaction systems have become an important approach for understanding the functionality of metabolic networks reconstructed by genomic and biochemical data.

References:
[01] M. Muller, *Evolutionary Relationships among Protozoa. Kluwer Academic Publishers, Dordrecht, The Netherlands*, 109 (1998)
[02] J. I. Ravdin, *Clinical Infectious Diseases* 20: 1453 (1995) [PMID: 7548493]
[03] T. F. H. G. Jackson, *Amebiasis. Imperial College Press, London*, 47 (2000)
[04] B. Loftus, *et al.*, *Nature,* 433: 865 (2005) [PMID: 15729342]
[05] M. Kanehisa, *et al.*, *Nucleic Acids Res.*, 30: 42 (2002) [PMID: 11752249]
[06] S. F. Altschul, *et al.*, *Nucleic Acids Res.*, 25: 3389 (1997) [PMID: 9254694]
[07] I. Yeh, *et al.*, *Genome Res.,* 14: 917 (2004) [PMID: 15078855]
[08] R. C. Fahey, *et al.*, *Science,* 224: 70 (1984) [PMID: 6322306]
[09] D. Yang, *et al.*, *Proc Natl Acad Sci.,* 82: 4443 (1985) [PMID: 3892535]
[10] M. W. Gray, *Curr Opin Genet Dev.,* 3: 884 (1993) [PMID: 8118213]
[11] S. Ortner, *et al.*, *Mol Biochem Parasitol.,* 86: 85 (1997) [PMID: 9178270]
[12] L. J. Baker, *et al.*, *Structure,* 9: 539 (2001) [PMID: 11435118]
[13] D. Vats, *et al.*, *Inf. and Immun.,* 73: 8381 (2005) [PMID: 16299336]
[14] J. B. Martin, *et al.*, *Eur. J. Biochem.,* 214: 711 (1993) [PMID: 8319681]
[15] T. Tanaka, *et al.*, *Biochem.Biophys.Res.Comm.,* 236: 611 (1997) [PMID: 9245698]
[16] J. V. Villarreal, *et al.*, *Parasitology Research,* 81: 320 (1995)
[17] H. D. Lujan & L. S. Diamond, *Arch Med Res.,* 28:96 (1997) [PMID: 9033028]
[18] T. Bakker-Grunwald, *et al.*, *J. Eukar. Microbiol.,* 42: 346 (1995)
[19] X. Zuo, & G. H. Coombs, *FEMS Microbiol. Lett.,* 130: 253 (1995) [PMID: 7649448]
[20] T. Nozaki, *et al.*, *Mol. Biochem Parasitol.,* 97: 33 (1998) [PMID: 9879885]
[21] C. E. Garrett, *et al.*, *Mol. Biochem Parasitol.,* 11: 257 (1984) [PMID: 6749182]
[22] K. Becker, *Eur. J. Biochem.,* 267: 6118 (2000) [PMID: 11012663]
[23] J. E. J. Nixon, *et al.*, *Biol. Bull.,* 204**:** 1 (2003) [PMID: 12588739]
[24] R. C. MacFarlane & U. Singh, *Infect Immun.,* 74: 340 (2006) [PMID: 16368989]
[25] S. Muller, *et al.*, *Trends in Parasitol.,* 17: 242 (2001) [PMID: 11323309]
[26] S. Shuman, *Nat. Rev Mol Cell Biol.,* 3: 619 (2002) [PMID: 12154373]
[27] A. K. Bhattacharya, *et al.*, *Current Sci.,* 77: 564 (1999)
[28] H. F. Hassan, & G. H. Coombs, *Mol Biochem Parasitol.,* 19: 19 (1986) [PMID: 2872591]
[29] B. Rosenthal, *et al.*, *J. Bacteriol.,* 179: 3736 (1997) [PMID: 9171424]
[30] S. Muller, *et al.*, *Trends in Parasitol.,* 19: 320 (2003)
[31] G. Ramakrishnan, *et al.*, *Mol Biochem Parasitol.,* 138: 205 (2004)
[32] S. D. Copley & J. K. Dhillon, *Genome Biol.,* 3: 1 (2002) [PMID: 12049666]
[33] H. Mukhtar, *Exp Parasitology,* 26: 135 (1969)
[34] M. Kumagai, *et al.*, *J.Biol.Chem.,* 279: 2316 (2004) [PMID: 14583615]
[35] I. Bruchhaus, and E. Tannich, *Biochem. J.,* 303: 743 (1994)
[36] H. Tachibana, *et al.*, *J Clin Microbiol.,* 29: 2234 (1991) [PMID: 1939576]
[37] E. Ullu, *et al.*, *Cell Microbiol.,* 6: 509 (2004) [PMID: 15104593]
[38] G. Kaur & A. Lohia, *Biochem Biophys Res Commun.,* 320: 1118 (2004) [PMID:15249205]
[39] A. Stechmann, & T. Cavalier-Smith, *Science,* 297: 89 (2002) [PMID: 12098695]
[40] M. Tokoro, *et al.*, *J. Biol. Chem.,* 278: 42727 (2003) [PMID: 12920135]
[41] ftp://ftp.ncbi.nlm.nih.gov/blast/executables/BLAST executables

Supplementary material

Genbank accession No.	Gene	Description
Peptidoglycan biosynthesis: KEGG pathway ID ehi00550		
EAL50404	23.t00033	glutamine synthatase,putative
Thiamine metabolism: KEGG pathway ID ehi00730		
EAL50105	28.t00019	myotubularin, putative
EAL44953	226.t0008	inosine triphosphate pyrophosphatase, putative
EAL52121	1.t00126	thiamine pyrophosphokinase,putative
Streptomycin biosynthesis: KEGG pathway ID ehi00521		
EAL50580	19.t00020	hexokinase
EAL44377	273.t00008	L-myo-inositol-phosphate synthase
EAL45227	208.t00004	myo-inositol monophosphatase,putative
EAL51122	11.t00065	dTDP-D-glucose 4, 6-dehyratase,putative

Table 1: Targets from unique pathways, which do not have human homologue

Target enzyme	EC number	Reference	Choke point	In Human
Thymidylate synthase	2.1.1.45	[21]	Yes	Yes
Thioredoxin reductase	1.8.1.9	[22]	No	Yes
Succinate dehydrogenase	1.3.99.1	[23]	No	Yes
Sphingomyelinase	3.1.4.12	[24]	Yes	No
S-adenosylmethionine decarboxylase	4.1.1.50	[25]	Yes	Yes
S-adenosyl-L-homocysteine hydrolase	3.3.1.1	[26]	Yes	Yes
RNA polymerase	2.7.7.6	[27]	Yes	Yes
Purine nucleoside phosphorylase	2.4.2.1	[28]	Yes	No
NADH dehydrogenase (ubiquinone)	1.6.5.3	[29]	Yes	No
Lysophospholipase	3.1.1.5	[16]	Yes	No
Lactoglutathione lyase	4.4.1.5	[30]	Yes	Yes
IMP dehydrogenase	1.2.1.14	[28]	No	Yes
Hypoxanthine phosphoribosyltransferase	2.4.2.8	[28]	Yes	Yes
Histone deacetylase		[31]	Yes	Yes
Gamma-glutamylcysteine ligase	6.3.2.2	[32]	Yes	Yes
Fructose-bisphosphate aldolase	4.1.2.13	[33]	Yes	Yes
Farnesyl diphosphate farnesyltransferase	2.5.1.21	[34]	Yes	Yes
Alcohol dehydrogenase (NAD+)	1.1.1.1	[35]	Yes	No
DNA polymerase	2.7.7.7	[36]	Yes	Yes
Chorismate synthase	4.2.3.5	[37], [38]	Yes	No
Choline transport		[14]	Yes	No
Carbamoyl-phosphate synthase	6.3.5.5	[39]	Yes	Yes
3-oxoacyl-[acyl-carrier protein] synthase	2.3.1.41	[13]	Yes	Yes

Table 2: Drug targets proposed in the literature with biological evidence is given. The reaction and its EC number are given. The "In Human" column denotes whether or not the enzymatic activity has a similar enzyme in human as determined by BLAST alignment with an expectation of less than 0.0001. Of these 23 reactions, 20 (86.9 percent) were identified as chokepoints.