In silico reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi*

A.C.R. Guimarães[1], T.D. Otto[1,2], M. Alves-Ferreira[1], A.B. Miranda[1] and W.M. Degrave[1]

[1]Laboratório de Genômica Funcional e Bioinformática,
Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, RJ, Brasil
[2]Fundação Ataulpho de Paiva, Rio de janeiro, RJ, Brasil

Corresponding author: A.C.R. Guimarães
E-mail: carolg@fiocruz.br

Genet. Mol. Res. 7 (3): 872-882 (2008)
Received June 2, 2008
Accepted August 11, 2008
Published September 23, 2008

ABSTRACT. *Trypanosoma cruzi* is the epidemiological agent of Chagas' disease, affecting most of Central and South America, constituting a significant health and socio-economic problem. The parasite has a metabolism largely based on the consumption of amino acids, which participate in a diversity of metabolic pathways, leading to many crucial compounds for the survival of this parasite. Study of its enzymes has the potential to disclose new therapeutic targets and foster the development of new drugs. In this study, we employed computational approaches to reconstruct *in silico* the amino acid metabolic pathways of *T. cruzi*, aiming to link genomic information with functional information. For that, protein sequences from 570 EC classes belonging to 25 different

 ©FUNPEC-RP www.funpecrp.com.br

pathways in general amino acid metabolism were downloaded from KEGG. A subset of 471 EC classes had at least one sequence deposited. Clustering of the proteins belonging to each EC class was performed using a similarity-based approach implemented in the tool AnEnPi. Reconstruction of the metabolic pathways comprising the amino acid metabolism of *T. cruzi* was performed by analyzing the output of BLASTP, using as query the dataset of predicted proteins of *T. cruzi* against all sequences of each individual cluster. This approach allowed us to identify 764 *T. cruzi* proteins probably involved in the metabolism of amino acids as well as the identification of several putative cases of analogy. Furthermore, we were able to identify several enzymatic activities of *T. cruzi* that were not previously included in KEGG.

Key words: Amino acid; Metabolism; *Trypanosoma cruzi*; Analogy

INTRODUCTION

The successful completion of several genome projects has led to a new stage in the biological sciences, the post-genome era (Kanehisa and Bork, 2003). One of the most important challenges in this era is the elucidation of cellular functions, which can be viewed as a particular behavior of a complex system of interactions between several proteins (Hieter and Boguski, 1997). Typically, information not directly stored in the genome is necessary for the understanding of a determined situation (Goto et al., 2002). One way to link genomics with biochemistry is the use of EC (Enzyme Commission) numbers representing enzymatic reactions. The assignment of the EC numbers is based on published experimental data on individual enzymes, by the Joint Commission on Biochemical Nomenclature (JCBN) of the International Union of Biochemistry and Molecular Biology (IUBMB) and the International Union of Pure and Applied Chemistry (IUPAC) (Kanehisa, 2003; Kotera et al., 2004). Currently, there are several projects dealing with the compilation, analysis and storage of information about molecular complexes and cellular processes. Among these are EMP, MPW, WIT, UM-BBD, KEGG, MetaCyc, ERGO, and SEED (Popescu and Yona, 2005).

KEGG (Kyoto Encyclopedia of Genes and Genomes) is considered by many one of the most important bioinformatics resources for understanding higher-order functional meaning and utilities of the organism from its genome information. Metabolic networks are represented by wiring diagrams of protein and other gene products responsible for various cellular processes, such as metabolism (Kanehisa et al., 2006). This part of KEGG is supplemented by a set of ortholog group tables. Each reference pathway can be viewed as a network of enzymes or EC numbers. Once genes encoding enzymes are identified in the genome (usually by sequence similarity and/or positional correlation) and their EC numbers are properly assigned, organism-specific pathways can be computationally reconstructed by correlating genes in the genome with gene products (enzymes) in the reference pathways in accordance with their EC numbers. Since metabolic pathways are normally well conserved between most organisms from mammals to bacteria, it is possible

 ©FUNPEC-RP www.funpecrp.com.br

to manually draw one reference pathway and then generate organism-specific pathways using a computational approach (Ma and Zeng, 2003). The quality of this reconstruction largely depends on the quality of the initial assignment of EC numbers (Overbeek et al., 2000).

A metabolic pathway is a set of oriented reactions interacting under given physiological conditions via simple or apparently simple intermediates (Selkov Jr. et al., 1998). Pathways are cellular processes that are associated with a specific functionality in the cell, such as amino acid synthesis and degradation, energy metabolism, signal transduction, molecular oxidation, and others (Popescu and Yona, 2005). The complexity of a cell is a function of its underlying processes. Therefore, there is a strong interest in identifying the active pathways in an organism.

Enzymes that catalyze the same reaction typically show significant sequence and structural similarity. However, in some cases enzymes with the same activity can be associated with different phylogenetic lineages and have different catalytic mechanisms with little structural similarity. Such enzymes are generally believed to have evolved independently, rather than having descended from a common ancestral enzyme, and are appropriately referred to as analogous, as opposed to homologous, enzymes (Fitch, 1970). Sequence comparison alone cannot prove that two sequences are evolutionary unrelated; a common origin can be inferred from protein structure conservation even after sequence conservation has been completely washed out by divergence. The possibility of a common origin can be ruled out only when candidate analogous enzymes have different three-dimensional folds (Galperin et al., 1998). Some enzymatic activities catalyzing conserved metabolic pathways that are central to life are present in almost all organisms. However, enzymes and the mechanisms by which they catalyze the corresponding biochemical reaction may show differences that can be exploited for drug development. Accordingly, differences in sequence structure may be used as a first criterion to elect the protein as a therapeutic target (Silber et al., 2005). However, few studies have been conducted to identify and annotate the occurrence of analogy, which in this study means the independent origin of a certain function in different organisms, and can be interpreted as the absence of detectable similarity and identity between the primary structure of two different proteins sharing the same function or activity.

Trypanosoma cruzi, the etiological agent of Chagas' disease, constitutes a significant health and socio-economic problem in the Americas, with about 18 million people infected, for the most part in Central and South America (Kirchhoff et al., 2006). The disease is primarily transmitted by triatominae insects to different mammalian hosts (El-Sayed et al., 2005). *T. cruzi* has a metabolism largely based on the consumption of amino acids, which participate in a diversity of metabolic pathways, leading to many crucial compounds for its survival. Also, enzymes associated with the metabolism of amino acids are potential targets for drug development (Silber et al., 2005). Therefore, a deeper analysis of the amino acid metabolism of *T. cruzi* may reveal important biochemical features as well as provide new potential targets for the development of drugs against this parasite.

In the present study, we employed a computational approach to reconstruct the pathways involved with the amino acid metabolism of *T. cruzi*, aiming to link genomic information with higher-order functional information by compiling current knowledge of

cellular processes and gene annotations. To accomplish this task, we developed AnEnPi (Otto TD, Guimarães ACR, Miranda AB and Degrave WM, unpublished results), a computational tool that groups protein sequences based on their enzymatic function and the level of sequence similarity, and used it to identify enzymatic functions of the amino acid metabolism of *T. cruzi*. We also present improvements in the annotation of these enzymes by the addition of biological data, in particular the occurrence of analogy.

MATERIAL AND METHODS

Trypanosoma cruzi proteins

We used the release TSK-TSC v5.0 (July 15, 2005) of the predicted proteins of *T. cruzi* - CL Brener genomic sequence (consortium for the sequencing of the *T. cruzi* genome), comprising a total of 19,613 predicted proteins (http://tcruzidb.org/tcruzidb/).

Pathways and enzyme classes

A set of 25 pathways (maps) was downloaded from the KEGG database (http://www.genome.jp/kegg) (release 41, January 2007). This dataset contains a complete biochemical description of the pathways related to the amino acid metabolism observed in different organisms. These descriptions were used as templates for the reconstruction of the correspondent pathways in *T. cruzi*. Functions comprising a certain pathway were extracted from these descriptions as a collection of EC numbers. Each pathway is associated with a set of proteins, usually a list of enzyme families with their EC numbers. To assign *T. cruzi* proteins to EC numbers we used all protein sequences related to the aforementioned maps of all organisms, as available in the KEGG database.

Clustering

Protein sequences of KEGG containing less than 100 amino acid residues were discarded. Each of the remaining sequences was individually compared with all other proteins annotated with the same EC number using the BLASTP program (Altschul et al., 1990). A pair of sequences possessing a similarity score above a cut-off of 120 (corresponding to an E-value $<10^{-6}$) was grouped by single-linkage clustering. This score is based on a previous analysis performed by Galperin et al. (1998) where it was observed that this cut-off was able to separate sequences with different three-dimensional structures. After grouping, proteins belonging to different clusters of the same enzymatic activity (EC number) were defined as analogous. Identification of analogous enzymes was based on the fact that under the IUBMB Nomenclature Commission rules, each EC number specifies one particular reaction. Therefore, analogous enzymes were identified as a pair of proteins with the same EC number but with no detectable sequence similarity with each other (Otto TD, Guimarães ACR, Miranda AB and Degrave WM, unpublished results). Figure 1 shows the methodology workflow used in this study. This methodology is implemented at http://bioinfo.pdtis.fiocruz.br/AnEnPi/ for utilization by the scientific community and detailed information about this can be comprehended following the documentation contained in its homepage.

 ©FUNPEC-RP www.funpecrp.com.br

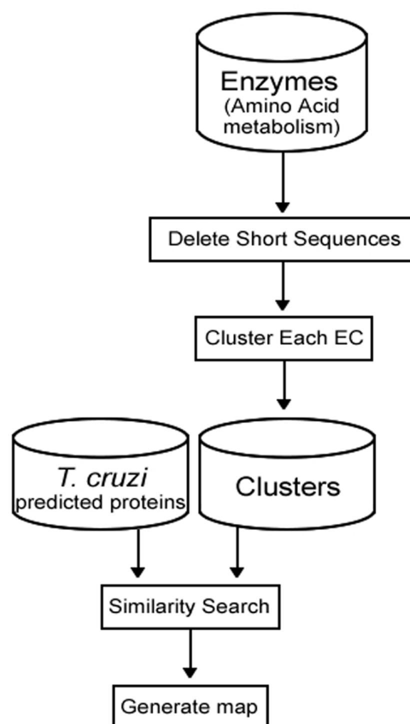

Figure 1. Workflow of methodology. EC = Enzyme Commission number.

Trypanosoma cruzi pathway reconstruction

The reconstruction of the metabolic pathways involved with the amino acid metabolism of *T. cruzi* was performed using the BLASTP program, using as query the dataset of predicted proteins of *T. cruzi* against the sequences of each individual cluster, employing different threshold values. Maps were generated using a tool available at KEGG (http://www.genome.jp/kegg/tool/color_pathway.html).

RESULTS

Clustering of enzymatic activities

From a subset of KEGG comprising 25 maps representing the pathways involved with the metabolism of amino acids, we were able to obtain a set of 471 non-redundant EC numbers with at least one sequence available. However, only 435 EC numbers have a complete (four-digit) EC number. After clustering, a total of 1384 clusters were obtained. Analysis of these clusters allowed us to identify 222 enzymatic activities with putative analogous proteins, corresponding to 51% of the enzymatic functions with a complete EC number in this particular

 ©FUNPEC-RP www.funpecrp.com.br

metabolism (Table 1). As expected, no similarity between sequences from different clusters was detected (Figure 2). There are two different main clusters produced for EC:1.1.1.42 (isocitrate dehydrogenase) where no similarity sequence is detected and two sub-clusters generated in cluster 1. The presence of sub-clusters is due to the low similarity among some sequences in the cluster and groups of sequences with more similarity.

Table 1. Data obtained after clustering for enzymatic functions of amino acid metabolism.

Data	Amount
Maps in KEGG	25
ECs with at least one sequence in KEGG	471
ECs with a complete classification (four-digit EC number)	435
Enzymatic activities with putative cases of analogy	222

KEGG = Kyoto Encyclopedia of Genes and Genomes; EC = Enzyme Commission number.

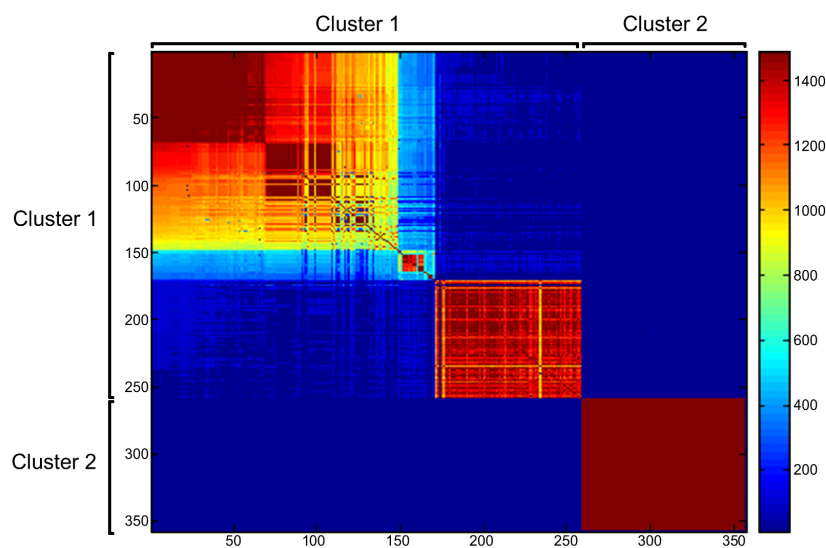

Figure 2. A graph depicting the similarity scores of all pairwise comparisons between the sequences belonging to EC:1.1.1.42 (isocitrate dehydrogenase). The order of the sequences along the axis is arbitrary. The scale of the similarity score is presented as a vertical bar on the right of the figure. Two main clusters can be observed.

Amino acid metabolism reconstruction of *Trypanosoma cruzi*

The utilization of BLASTP allowed us to identify several *T. cruzi* proteins involved in the metabolism of amino acids. A critical point in this procedure is the cut-off employed. Table 2 displays the number of identified proteins, the number of the corresponding enzymatic activities, the number of cases of analogy, and the number of groups obtained with different threshold values. Obviously, the more rigorous the cut-off, the less proteins were identified. However, the higher cut-off used in this analysis (e^{-20}) is usually accepted by annotators as a positive result.

Table 2. Number of proteins, enzymatic activities, cases of analogy and clusters found in the amino acid metabolism of *Trypanosoma cruzi* using different E-values as cut-off.

E-value	Proteins	Enzymatic activities	Cases of analogy	Groups
$<e^{-20}$	764	229	54	290
$<e^{-40}$	528	192	43	221
$<e^{-80}$	377	136	31	145

Table 3 summarizes our data on the reconstruction of the amino acid metabolic pathways of *T. cruzi*, using two different E-values as cut-off (e^{-20} and e^{-80}). In all cases, we were able to identify more proteins when comparing our results to the previously annotated subset of *T. cruzi* proteins present in KEGG or in the tables provided by the consortium for the sequencing of the *T. cruzi* genome. For example, from a total of 34 activities participating in the urea cycle and metabolism of amino group pathway, we identified 18 (E-value = e^{-20}) and 12 (E-value = e^{-80}) enzymatic activities while KEGG registers only 7 activities and the consortium, 4.

Analogous enzymes were found in 22 pathways. For instance, three clusters were produced after grouping of the sequences available in KEGG for the enzyme enoyl-CoA hydratase (EC:4.2.1.17). In the genome of *T. cruzi*, 18 sequences can be found annotated with this enzymatic activity; they were all placed in the same cluster. Since they share some degree of sequence similarity, they probably have a common ancestor; however, they do not display any detectable similarity with sequences from the other two clusters. In this case, we can say that *T. cruzi* enoyl-CoA hydratase sequences are analogous to sequences from the other two clusters.

Table 3. Data description of the computational reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi.*

Metabolic pathways	EC	Consortium	Tcr-KEGG	(e^{-20})	(e^{-80})	Analogy
Glutamate metabolism	36	7	16	22	21	12
Alanine and aspartate metabolism	42	9	15	22	16	10
Glycine, serine and threonine metabolism	60	7	12	33	19	23
Methionine metabolism	37	7	9	18	13	11
Cysteine metabolism	23	2	5	10	7	6
Valine, leucine and isoleucine degradation	33	8	14	19	14	10
Valine, leucine and isoleucine biosynthesis	18	4	4	9	4	5
Lysine biosynthesis	29	2	0	11	5	8
Lysine degradation	53	6	9	13	10	10
Arginine and proline metabolism	75	7	7	26	12	15
Histidine metabolism	40	5	6	14	11	10
Tyrosine metabolism	70	4	6	21	12	15
Phenylalanine metabolism	45	2	3	14	7	9
Tryptophan metabolism	60	6	11	22	15	16
Phenylalanine, tyrosine and tryptophan biosynthesis	31	3	4	8	6	4
Urea cycle and metabolism of amino groups	34	4	7	18	12	8
Beta-alanine metabolism	32	4	6	12	6	7
Taurine and hypotaurine metabolism	17	0	0	3	1	2
Aminophosphonate metabolism	15	2	4	12	5	7
Selenoamino acid metabolism	22	4	7	15	13	10
Cyanoamino acid metabolism	19	2	4	7	5	4
D-glutamine and D-glutamate metabolism	12	0	0	2	1	0
D-arginine and D-ornithine metabolism	8	0	0	1	0	0
D-alanine metabolism	6	1	0	1	0	0
Glutathione metabolism	27	4	6	13	11	6

EC, total number of activities registered in KEGG; Consortium, number of activities identified by the consortium for the sequencing of the *T. cruzi* genome; Tcr-KEGG, number of activities registered in KEGG for *T. cruzi*; e^{-20}, number of activities found with E-value = e^{-20}; e^{-80}, number of activities found with E-value = e^{-80}; Analogy, cases of enzymatic activities with analogy.

The metabolic pathways involved with the amino acid metabolism of *T. cruzi* were graphically reconstructed using the maps from KEGG as templates. Figure 3 illustrates one of these maps depicting the selenoamino acid metabolism pathway of *T. cruzi*. A comparison between our annotation (Figure 3A) and KEGG (Figure 3B) shows some enzymatic activities identified with AnEnPi and absent in KEGG. All maps and data are available at http://www.dbbm.fiocruz.br/labwim/TcruziAA/.

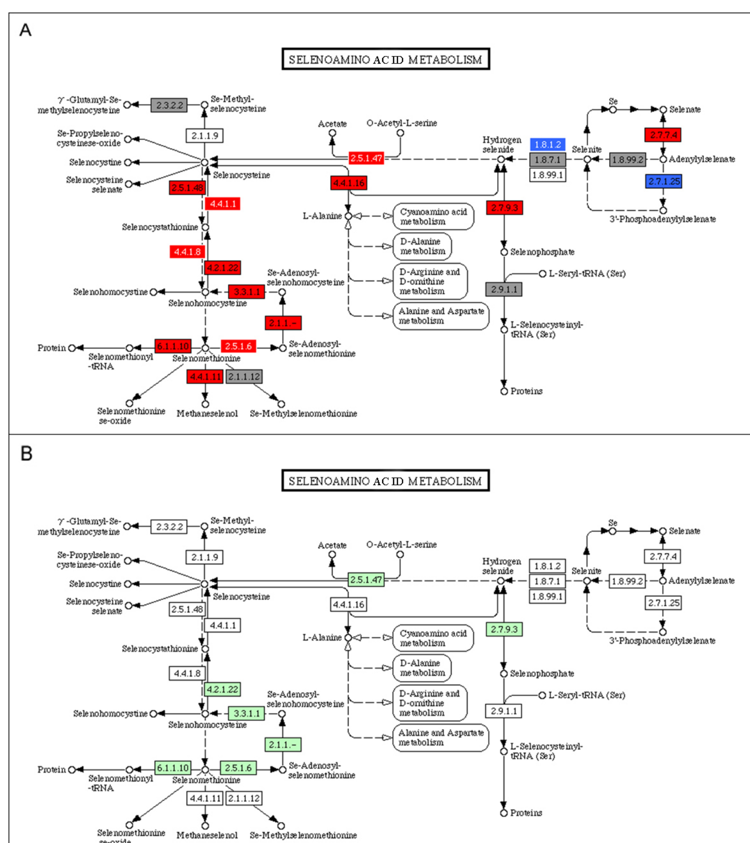

Figure 3. Maps of selenoamino acid metabolism showing the enzymatic functions of *Trypanosoma cruzi* identified by AnEnPi (A) and annotated for *T. cruzi* by KEGG (B). **A.** The background colors of the boxes specifying EC classes represent attributes of the enzymatic activities, respectively: white - no representatives in KEGG; gray - not found in *T. cruzi*; red - found in *T. cruzi* with E-value = e^{-80}; blue - found in *T. cruzi* with E-value = e^{-60}. Font of EC numbers: black - no analogy detected; white - possible cases of analogy within that class. **B.** Green - enzymatic function of *T. cruzi* registered in KEGG.

DISCUSSION

When the reconstruction of a metabolic pathway is accurate, it is likely that the gene products have been correctly identified. On the other hand, two possibilities arise when this reconstruction is incomplete: i) gene function has not been correctly characterized, resulting in

an incorrect assignment of the EC number; ii) incomplete or superficial knowledge of a certain metabolic pathway. Thus, we developed a tool that allows the identification of enzymatic functions using a methodology previously described by Galperin et al. (1998).

The annotation quality can be influenced by assembling mistakes, generating false positives (wrongly identified enzymatic activities) or false negatives (absence of enzymes) because the method used in this study is based on previously annotated functions. Therefore, it is important to use a curated database. For this main reason, we chose to use data from KEGG.

The model of this study is *T. cruzi* due to its socio-economic importance and also the paucity of therapeutic possibilities. The gut of the hematophagous insect (the invertebrate host of *T. cruzi*) has basically proteins and amino acids, and the parasite spends part of its life in this local using a subset of these amino acids (mainly proline, aspartate, arginine, and glutamate) to generate energy (Nowicki and Cazzulo, 2007). Some catabolic pathways involved in these and other processes are, at least partially, connected to energy metabolism, allowing the utilization of amino acids as a source of energy. Besides, amino acids are also involved in the differentiation process from the replicative to the non-replicative infective forms. Proline also participates in the intracellular differentiation cycle inside the mammalian host (Sylvester and Krassner, 1976; Silber et al., 2005). Together, these data suggest that enzymes participating in the amino acid metabolism may be good candidates for the development of new drugs against this parasite.

We identified, for example, arginine kinase (EC:2.7.3.3), which is present in the arginine and proline metabolism converting free arginine to phosphoarginine, a phosphagen that has a role as an energy source and is important during stress conditions (Silber et al., 2005). This function was already recognized by the consortium annotators and also by KEGG, showing that the methodology used in this study is able to correctly find the enzymatic functions present in this organism. This was the case for the majority of functions; only a few cases escaped detection (3 using e^{-20} as cut-off and 11 with e^{-80}).

Even using BLASTP with a conservative E-value (e^{-80}), our methodology was able to find additional enzymatic activities, not registered in KEGG nor identified by the consortium. For instance, we detected the presence of the enzyme ATP-L-glutamate 5-phosphotransferase (EC:2.7.2.11), present in the urea cycle and metabolism of amino groups. This enzyme is important in the transfer of a phosphate group from ATP to L-glutamate. The enzyme pyrroline-5-carboxylate synthetase (EC:1.2.1.41), the next metabolic step, was already present in KEGG for *T. cruzi*. Recognition of ATP-L-glutamate 5-phosphotransferase allowed a more accurate metabolic reconstruction, showing the importance of more sensitive methods for function attribution.

Other enzymatic functions were identified by annotators from the consortium (El-Sayed et al., 2005) but were absent from KEGG, such as EC:3.5.1.-, which is a hydrolase and participates in the convertion of N_2-acetyl L-lysine to L-lysine. EC:6.1.1.6 (lysine-tRNA ligase) is another example; it is also involved in lysine metabolism, and it is also absent from KEGG. Both functions were found with our methodology using a conservative cut-off (E-value = e^{-80}). Furthermore, we were able to identify enzymatic activities in the D-glutamine and D-glutamate metabolism with a rigorous cut-off (E-value = e^{-80}), which were not identified by KEGG nor by the consortium for the sequencing of the *T. cruzi* genome.

Other examples are glutathione reductase and thioredoxin reductase (EC:1.8.1.7), which are part of glutathione metabolism. *T. cruzi* lacks these enzymes but has an enzyme called trypanothione reductase. Trypanothione is synthesized from glutathione and spermidine in two consecutive steps. In the first reaction, Gsp is formed, which reacts with a second gluta-

thione molecule to form trypanothione. Although it forms the basis of the parasite thiol metabolism, trypanosomatids contain also significant levels of free glutathione. Glutathionylation of proteins is a protective mechanism against oxidative damage as well as a regulation mechanism of enzyme activities. So, the presence of this enzyme in this organism is essential for its survival which has already been experimentally proved (Melchers et al., 2007).

Several organisms apparently lack some enzymatic functions in their metabolic pathways. For a number of cases, this is certainly true, due to particularities with their life cycle and/or life style and also their evolution (Galperin and Koonin, 1999; Morett et al., 2003). Paramount to this point is the existence of alternative enzymes (or sets thereof) that can play the role of the apparently missing reactions. It is assumed that enzymes catalyzing the same reaction will typically have significant sequence and structural similarity. In fact, for the majority of enzymes, this is probably true (Fisher, 2005). However, a large number of enzymes are found where two or more forms with little or no demonstrable sequence similarity share the same function (Galperin et al., 1998). In this study, we identified several putative cases of analogy in the set of enzymatic activities involved in amino acid metabolism. We could observe that events of analogy were present in the majority of the amino acid metabolism maps, in accordance with preceding studies (Galperin et al., 1998; Morett et al., 2003).

The computational approach described in this study for the identification of enzymatic functions and analogy, implemented in AnEnPi and used here with *T. cruzi* proteins as a model, may be used for other organisms as well. The identification of more genes will help obtain a more accurate view of metabolism. Also, the detection of analogy is interesting not only from an evolutionary point of view but also from a practical one, since analogous enzymes may be prime candidates for drug development, and are therefore worthy of recognition and annotation.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, et al. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215: 403-410.

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, et al. (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 309: 409-415.

Fisher HP (2005). Towards quantitative biology: integration of biological information to elucidate disease pathways and to guide drug discovery. *Biotechnol. Annu. Rev.* 11: 1-68.

Fitch WM (1970). Distinguishing homologous from analogous proteins. *Syst. Zool.* 19: 99-113.

Galperin MY and Koonin EV (1999). Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. *Genetica* 106: 159-170.

Galperin MY, Walker DR and Koonin EV (1998). Analogous enzymes: independent inventions in enzyme evolution. *Genome Res.* 8: 779-790.

Goto S, Okuno Y, Hattori M, Nishioka T, et al. (2002). LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* 30: 402-404.

Hieter P and Boguski M (1997). Functional genomics: it's all how you read it. *Science* 278: 601-602.

Kanehisa M (2003). Post-genome Informatics. Oxford University Press, Oxford.

Kanehisa M and Bork P (2003). Bioinformatics in the post-sequence era. *Nat. Genet.* 33 (Suppl.): 305-310.

Kanehisa M, Goto S, Hattori M, Oki-Kinoshita KF, et al. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res.* 34: D354-D357.

Kirchhoff LV, Paredes P, Lomeli-Guerrero A, Paredes-Espinoza M, et al. (2006). Transfusion-associated Chagas disease (American trypanosomiasis) in Mexico: implications for transfusion medicine in the United States. *Transfusion* 46: 298-304.

Kotera M, Okuno Y, Hattori M, Goto S, et al. (2004). Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. *J. Am. Chem. Soc.* 126: 16487-16498.

Ma H and Zeng AP (2003). Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics* 19: 270-277.

Melchers J, Dirdjaja N, Ruppert T and Krauth-Siegel RL (2007). Glutathionylation of trypanosomal thiol redox proteins. *J. Biol. Chem.* 282: 8678-8694.

Morett E, Korbel JO, Rajan E, Saab-Rincon G, et al. (2003). Systematic discovery of analogous enzymes in thiamin biosynthesis. *Nat. Biotechnol.* 21: 790-795.

Nowicki C and Cazzulo JJ (2007). Aromatic amino acid catabolism in trypanosomatids. *Comp. Biochem. Physiol. A Mol. Integr. Physiol*. Mar 15 [Epub ahead of print; PMID: 17433885].

Overbeek R, Larsen N, Pusch GD, D'Souza M, et al. (2000). WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res.* 28: 123-125.

Popescu L and Yona G (2005). Automation of gene assignments to metabolic pathways using high-throughput expression data. *BMC Bioinformatics* 6: 217.

Selkov E Jr, Grechkin Y, Mikhailova N and Selkov E (1998). MPW: the metabolic pathways database. *Nucleic Acids Res.* 26: 43-45.

Silber AM, Colli W, Ulrich H, Alves MJ, et al. (2005). Amino acid metabolic routes in *Trypanosoma cruzi*: possible therapeutic targets against Chagas' disease. *Curr. Drug Targets Infect. Disord.* 5: 53-64.

Sylvester D and Krassner SM (1976). Proline metabolism in *Trypanosoma cruzi* epimastigotes. *Comp. Biochem. Physiol. B* 55: 443-447.