Ministério da Saúde

**FIOCRUZ**
**Fundação Oswaldo Cruz**

## INSTITUTO OSWALDO CRUZ

**Pós-Graduação em Biologia Celular e Molecular**

*ANA CAROLINA RAMOS GUIMARÃES*

# Identificação *in silico* de enzimas isofuncionais não-homólogas, um potencial reservatório de alvos terapêuticos.

Tese apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Doutor em Ciências.

**Orientadores:** Dr. Wim Maurits Degrave
Dr. Antonio Basílio de Miranda

Este trabalho foi realizado no Laboratório de Genômica Funcional e Bioinformática, do Instituto Oswaldo Cruz – Fundação Oswaldo Cruz (IOC/FIOCRUZ).

Rio de Janeiro
2010

Ministério da Saúde

FIOCRUZ
**Fundação Oswaldo Cruz**

**INSTITUTO OSWALDO CRUZ**

**Pós-Graduação em Biologia Celular e Molecular**

***ANA CAROLINA RAMOS GUIMARÃES***

Identificação *in silico* de enzimas isofuncionais não-homólogas, um potencial reservatório de alvos terapêuticos.

**ORIENTADORES: Dr. Wim Maurits Degrave**

**Dr. Antonio Basílio de Miranda**

**Aprovada em:**

**EXAMINADORES:**

Dr. Carlos Medicis Morel - **Presidente**

Dr. Ernesto Raúl Caffarena

Dr. Rodolpho Mattos Albano

Dr. Marcelo Alves-Ferreira (Revisor e suplente)

Dr. Fabio Faria da Mota

Rio de Janeiro, 16 de julho de 2010.

Aos meus pais.

**Agradecimentos:**

A preparação de uma tese é uma tarefa árdua e longa, requer a participação de muitas pessoas, seja de maneira profissional ou afetiva. Agradecer a essas pessoas é bastante complicado; seria mais fácil se houvesse um gráfico ou uma equação que transcrevesse em palavras todos os meus sentimentos...

Em primeiro lugar, devo agradecer ao "Cara lá de cima", que independente de credo ou religião, foi uma força propulsora, me iluminando e guiando em todos os momentos desta longa caminhada. "*Concedei-me, Senhor, a serenidade necessária para aceitar as coisas que não posso modificar, coragem para modificar as que eu posso e sabedoria para distinguir uma da outra."*

Aos meus orientadores, Wim e Antonio, que me ofereceram a oportunidade de desenvolver minhas idéias e acreditaram no meu potencial. E, além disso, tiveram visão crítica, fundamental para realização desse trabalho.

A mi tutor Fernando Alvarez, por ser mi puerto seguro en otro país y por toda la paciencia en mostrar los aspectos evolutivos de este trabajo.

Meus pais, Marina e Aristheu, pela compreensão, paciência e carinho mesmo nos momentos de "stress". Por me fazerem acreditar que tudo é possível e me ensinarem a enfrentar os momentos de angústia e sofrimento, tornando-me cada dia mais forte.

A minha "mãe preta", Maria, por se expressar apenas com um olhar e por sempre me querer bem. E claro, pelo feijão e mocotó...

A minha família brasileira, que muito amo, pelo carinho, paciência e incentivo, em especial a minha tia Lúcia, que tem um amor incondicional por todos os sobrinhos.

A mi familia uruguaya, Mabel, Chiquito, Tere, Mari, Bettina, Eduardo y Serrana, por hacerme sentir parte de ella. Gracias por todo!

A Mauricio, que me devolvio la sonrisa y me enseña a crecer todos los días, a quilômetros de distancia. Te quiero mucho!

Aos grandes e eternos amigos, CC, Érico, Nacho, Luana, Paulo e Maio. E as “Amigas da Carona”, Quel, Lili, Mari, Paulinha, Lalá e Rê que fizeram a travessia ponte Rio-Niterói se transformar em momentos de apoio e compreensão mútuos. Todos souberam compreender todas as minhas buscas – as existenciais e as teóricas – sendo atentos, ternos e fundamentais. A melhor parte da vida de uma pessoa está nas suas amizades (Abraham Lincoln).

Ao Marcelo, revisor, suplente, colaborador e amigo. Obrigado por sempre estar disponível para conversar, discutir trabalhos e tirar dúvidas.

Ao Marcos, por estar presente nos melhores e piores momentos, por cuidar de mim como um irmão mais velho e se tornar um grande amigo.

A Priscila e Laurrent por me mostrarem o fascinante mundo estrutural.

Ao Thomas, Fabinho e Nicolas, por estarem presentes *ao vivo* ou *in silico*. E a Monete, minha eterna discípula, que me estressa e faz milhões de perguntas, mas que me enche de orgulho. Valeu por seu companheirismo no trabalho e na vida.

A todos do Laboratório de Genômica Funcional e Bioinformática, que alegram o meu dia-dia.

E a todos aqueles que, embora não nomeados, se fizeram presentes em diversos momentos, o meu reconhecido e carinhoso muito obrigado!

Todos vocês são co-autores deste trabalho.

Por fim, fica o agradecimento a quem utilizar essa tese de qualquer forma, pois é isso que faz todo esse trabalho valer à pena.

“Feliz aquele que transfere o que sabe

e aprende o que ensina.”

Cora Coralina

**Lista de Tabelas**

**Tabela 1:** Grupos classificatórios do BATS com os respectivos parâmetros de pontuação.

**Tabela 2:** Classificação feita pela ferramenta Filtros, de acordo com a qualidade dos modelos gerados a partir dos parâmetros obtidos pelo BATS.

**Tabela 3:** Classificação enzimática baseada na reação catalisada e de acordo com a IUBMB.

**Capítulo I - Table 1:** Refinement of the initial Dataset (A) through the application of successive filters.

**Capítulo I - Table 2:** Examples of analogy found in the literature and the methods used.

**Capítulo II - Table 1:** Data obtained after clustering for enzymatic functions of amino acid metabolism.

**Capítulo II - 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.

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

**Capítulo III - Table 1:** Computational reconstruction of metabolic pathways in *Trypanosoma cruzi*: identification of enzymatic functions and analogy with human.

**Capítulo III - Table 2:** Computational reconstruction of the Krebs cycle and glycolysis-gluconeogenesis pathways in *Trypanosoma cruzi*: identification of enzymatic functions and analogy with human.

**Capítulo IV - Table 1:** *Trypanosoma cruzi* 3D protein models.

**Capítulo IV - Table 2:** Predicted proteins and enzymatic functions of *Trypanosoma cruzi* using diferent cut-offs and KEGG and Swiss-Prot databases.

**Capítulo IV - Table 3:** Comparison between *Homo sapiens* and *Trypanosoma cruzi* functions obtained from KEGG and Swiss-Prot databases.

**Capítulo IV - Table 4:** Enzyme Commission Numbers (EC) associated to modelled *Trypanosoma cruzi* proteins.

**Capítulo IV - Table 5:** Protein Models: AnEnPi and enzyme classifications, and model quality.

**Capítulo IV - Table 6:** List of modelled sequences classified by AnEnPi as analogous or specific of *Trypanosoma cruzi*, in relation to *Homo sapiens*.

**Capítulo IV - Table 7:** Classification according to the quality of the models built based on BLAST sequence identity and BATS coverage of the template in relation to the target.

**Lista de Figuras**

**Figura 1:** Esquema da simulação de modelo na biologia de sistemas.

**Figura 2:** Os três níveis representativos da anotação genômica: onde, o que e como?

**Figura 3:** Fluxograma representativo da visão geral do processo automático de anotação genômica.

**Figura 4:** Anotação multidimensional.

**Figura 5:** Representação esquemática das regiões funcionais típicas de uma proteína.

**Figura 6:** Exemplo de gráfico de Ramachandran apresentando correlações entre os ângulos phi (φ) e psi (ψ) de uma cadeia polipeptídica.

**Figura 7:** Ilustração representativa dos dois graus de liberdade presentes numa ligação peptídica, ângulos phi (φ) e psi (ψ).

**Figura 8:** Fluxograma representativo de todas as etapas executada pelo MHOLline.

**Figura 9:** Representação de estruturas análogas e homólogas nas asas de morcegos, aves e insetos.

**Figura 10:** Árvore evolutiva mostrando a formação de genes homólogos (ortólogos e parálogos) e análogos. O gene do ancestral comum 1 sofre um evento de especiação originando genes ortólogos, em espécies diferentes, assim como o evento de duplicação origina genes parálogos em uma mesma espécie. Por evolução convergente, genes de ancestrais distintos (1 e 2) podem codificar para proteínas com mesma função, denominadas análogas.

**Capítulo I - Figure 1:** Workflow of AnEnPi.

**Capítulo I - Figure 2:** Similarity matrix.

**Capítulo I - Figure 3:** Comparative analysis between *H. sapiens* and *L. major*.

**Capítulo II - Figure 1:** Workflow of methodology.

**Capítulo II - Figure 2:** A graph depicting the similarity scores of all pairwise comparisons between the sequences belonging to EC:1.1.1.42 (isocitrate dehydrogenase).

**Capítulo II - 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).

**Capítulo III - Figure 1:** Reconstruction of the glycolysis and gluconeogenesis pathways in *T. cruzi*.

**Capítulo III - Figure 2:** Reconstruction of the Krebs cycle pathway in *Trypanosoma cruzi*. The map was generated using the reference map by KEGG (http://www.genome.jp/kegg/pathway.html).

**Capítulo III - Figure 3:** Map of the oxidative phosphorylation metabolism in *Trypanosoma cruzi* with a representation of the multi-enzyme complexes and their corresponding enzymatic function.

**Capítulo IV - Figure 1:** Structural comparison between a medium to high quality model of 3-hydroxiacyl-CoA dehydrogenase from *Trypanosoma cruzi* and one homologous and one analogous structures from PDB (classifed according to the AnEnPi pipeline).

**Capítulo IV - Figure 2:** Structural and sequence comparison between 2,4-dienoyl CoA reductase (DECR) from *Trypanosoma cruzi* and *Homo sapiens*, analogous enzymes.

**Figura 11:** Diagrama de relacionamento preliminar do banco de dados AnEnDB.

**Figura 12:** Interface preliminar do banco de dados AnEnDB.

## Sumário


| | Página |
|---|---|
| **1. Introdução** | 2 |
| 1.1. Fundamentos Teóricos | 3 |
| 1.1.1. Bioinformática | 3 |
| 1.1.2. Biologia de Sistemas | 4 |
| 1.1.3. Genômica Funcional | 6 |
| 1.1.3.1. Anotação Funcional | 7 |
| 1.1.4. Genômica estrutural | 12 |
| 1.1.4.1. Função estrutural | 12 |
| 1.1.4.2. Modelagem Comparativa | 14 |
| 1.1.4.3. MHOLline | 17 |
| 1.1.4.4. Desenho Racional de Fármacos | 20 |
| 1.1.5. Enzimas e metabolismo | 22 |
| 1.1.5.1. Conceitos | 22 |
| 1.1.5.2. Classificação enzimática | 25 |
| 1.1.5.3. Origem e evolução de funções enzimáticas e vias metabólicas | 28 |
| 1.2. Modelos | 34 |
| **2. Objetivos** | 36 |
| 2.1. Gerais | 37 |
| 2.2. Específicos | 37 |
| **3. Resultados** | 38 |
| 3.1. Capítulo I: Implementação da metodologia | 39 |
| Artigo 1: AnEnPi: identification and annotation of analogous enzymes. | 40 |
| 3.2. Capítulo II: Reconstrução metabólica | 49 |
| Artigo 2: *In silico* reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi*. | 50 |
| 3.3. Capítulo III: Nova abordagem para busca de novos alvos moleculares | 61 |
| Artigo 3: A new approach for potential drug target discovery through *in silico* metabolic pathway analysis using *Trypanosoma cruzi* genome information. | 62 |
| 3.4. Capítulo IV: Análise comparativa e modelagem estrutural de proteínas análogas | 73 |
| Artigo 4: Structural Modelling and Comparative Analysis of Homologous, Analogous and Specific Proteins from Trypanosoma cruzi versus Homo sapiens: Putative Drug Targets for Chagas' Disease Treatment. | 74 |
| **4. Discussão** | 94 |
| **5. Conclusões** | 111 |
| **6. Referências Bibliográficas** | 114 |
| **7. Anexos** | 139 |

Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

# INSTITUTO OSWALDO CRUZ

## Pós-Graduação em Biologia Celular e Molecular

Identificação *in silico* de enzimas isofuncionais não-homólogas, um potencial reservatório de alvos terapêuticos.

**TESE DE DOUTORADO**

**RESUMO**

**Ana Carolina Ramos Guimarães**

O estudo da reconstrução metabólica em diversos organismos expõe a existência de compostos cruciais para a sua sobrevivência. Dentre estes compostos, estão as enzimas, responsáveis pela catálise das reações bioquímicas em vias metabólicas. Diferentemente das enzimas homólogas, as enzimas análogas (também conhecidas como enzimas isofuncionais não homólogas) são capazes de catalisar as mesmas reações, mas sem apresentar similaridade de sequência significativa no nível primário e, possivelmente, com diferentes estruturas tridimensionais. Um estudo detalhado destas enzimas pode desvendar novos mecanismos catalíticos, adicionar informações sobre a origem e evolução de vias bioquímicas e revelar alvos potenciais para o desenvolvimento de drogas. Para muitas enfermidades causadas por parasitas, as opções terapêuticas permanecem ineficientes ou inexistentes, exigindo a busca de novos alvos. Estes podem ser proteínas específicas do parasita ausentes no hospedeiro ou compostos presentes em ambos, mas com estrutura tridimensional substancialmente diferente, como as enzimas análogas. A ferramenta AnEnPi, capaz de identificar, anotar e comparar enzimas homólogas e análogas, foi desenvolvida e utilizada para reconstruir computacionalmente as vias metabólicas de alguns organismos modelo, como os tripanossomatídeos. Uma análise mais focada no metabolismo de aminoácidos de *Trypanosoma cruzi* identificou alvos promissores para o desenvolvimento de novas drogas. Além disso, uma revisão do metabolismo geral de *T. cruzi* foi realizada em outras vias metabólicas, levando em consideração esta nova abordagem de busca por potenciais alvos terapêuticos. Uma vez que a estrutura tridimensional é importante no estudo de analogia, a ferramenta MHOLline foi utilizada para a obtenção de modelos 3D a partir de homólogos, análogos e proteínas específicas de *T. cruzi* versus *Homo sapiens*. As estratégias utilizadas nesse trabalho apóiam o conceito de análise estrutural, juntamente com a análise funcional de proteínas, como uma interessante metodologia computacional para detectar potenciais alvos para o desenvolvimento de novas drogas.

Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

# INSTITUTO OSWALDO CRUZ

## Pós-Graduação em Biologia Celular e Molecular

Identificação *in silico* de enzimas isofuncionais não-homólogas, um potencial reservatório de alvos terapêuticos.

**TESE DE DOUTORADO**

## ABSTRACT

**Ana Carolina Ramos Guimarães**

The study of metabolic reconstruction in different organisms exposes the existence of crucial compounds for its survival. Examples of these compounds are the enzymes that are responsible for the catalysis of biochemical reactions in metabolic pathways. Unlike the homologous enzymes, the analogous enzymes (also known as non-homologous isofunctional enzymes) are able to catalyze the same reactions, but without significant sequence similarity at the primary level and possibly with different three-dimensional structures. A detailed study of these enzymes may exhibit new catalytic mechanisms, add information about the origin and evolution of biochemical pathways and reveal potential targets for drug development. For many diseases caused by parasites, therapeutic options remain inefficient or nonexistent, requiring the search for new drug targets. These targets may be specific proteins of the parasite (absent in the host) or compounds present in the both organisms but with different three-dimensional structure, like analogous enzymes. The tool AnEnPi approach was able to identify, annotate and compare homologous and analogous enzymes. It was developed and used to reconstruct computationally the metabolic pathways of some model organisms such as trypanosomes. A more focused analysis on the amino acids metabolism of *Trypanosoma cruzi* identified promising targets for the development of new drugs. Furthermore, a review of the general metabolism of *T. cruzi* was carried out in other metabolic pathways, taking into account this new approach in the search for potential therapeutic targets. Since the three-dimensional structure is important in the study of analogy, the tool MHOLline was used to obtain 3D models for homologous, analogous and specific proteins of *T. cruzi* versus *Homo sapiens*. The strategies used in this study support the concept that structural analysis together with protein functional analysis could be an interesting computational methodology to detect potential targets for structure-based rational drug design.

# *Introdução*

## 1. Introdução

Este capítulo de introdução se inicia com uma breve descrição da Bioinformática, área de estudo no qual este trabalho está inserido.

Posteriormente, são introduzidos alguns conceitos e definições que fornecem uma sustentação teórica a esse estudo. Primeiramente, a Biologia de Sistemas, mostrando a importância da interdisciplinaridade das tecnologias de alto desempenho na geração e análise de dados. Em seguida, uma pequena exposição de Genômica Funcional, dando ênfase à Anotação, e Genômica Estrutural, principalmente abordando a modelagem comparativa (incluindo o *pipeline* MHOLline), essenciais para a realização de etapas importantes nas análises voltadas para a determinação de analogia.

Em seguida, é apresentada uma resumida revisão sobre enzimas e metabolismo, expondo a classificação enzimática adotada neste estudo, assim como características relacionadas com a origem e evolução das funções enzimáticas e vias metabólicas.

Por fim, são apresentados os organismos modelo utilizados neste estudo, os tripanossomatídeos patogênicos *Leishmania major*, *Trypanosoma brucei* e *Trypanosoma cruzi*, escolhidos por sua grande relevância socio-econômica e pela disponibilização recente das suas sequências genômicas. Além disso, dados do genoma de *Homo sapiens* foram utilizados com a finalidade de encontrar análogos funcionais como potenciais alvos para o desenvolvimento de novas drogas.

## 1.1. Fundamentos Teóricos

### 1.1.1. Bioinformática

> *"A Computação está para a Biologia da mesma forma que a matemática está para física."* (Harold Morowitz[1]).

O que é a Bioinformática? Para muitos é o desenvolvimento de ferramentas e métodos computacionais para análise, manipulação, construção, edição e gerenciamento de dados biológicos. A análise destes dados em laboratório é difícil e custosa; portanto, técnicas computacionais são essenciais (Craven & Shavlik, 1994).

Durante os últimos anos, a bioinformática tornou-se um dos mais visíveis domínios da ciência moderna. No entanto, esse "novo campo" tem uma história longa, juntamente com os triunfos da genética, biologia molecular e celular, e ao progresso na tecnologia computacional do século passado (Ouzounis & Valencia, 2003). Seu início ocorreu paralelamente aos avanços das ciências biológicas e computacionais. Com o desenvolvimento do primeiro computador moderno digital, ou seja, onde os dados são armazenados em um alfabeto binário (de zeros e uns) e da descoberta da dupla hélice de DNA (Watson & Crick, 1953), percebeu-se que a informação genética também poderia ser escrita com um alfabeto, só que quaternário e não binário, pois são usadas 4 letras, as famosas A, C, G, e T. Mais tarde descobriu-se que a forma dos genes operarem também poderia ser, até certo ponto, digital: os genes podem ser "ligados" ou "desligados" (Jacob & Monod, 1961). Apenas esta observação já seria suficiente para prever, na década de 50, que um dia a informática e a biologia molecular iriam juntas fazer nascer uma nova área do conhecimento (Setúbal, 2003).

As pesquisas atuais que envolvem a bioinformática estão relacionadas com diversas aplicações que compreendem o desenvolvimento de técnicas analíticas e quantitativas à modelagem de sistemas biológicos, incluindo tanto uma análise

[1] Biofísico norte-americano que se dedicou com especial interesse à termodinâmica dos sistemas vivos. Suas pesquisas atuais articulam os fundamentos da biologia com as ciências da informação. Publicou também sobre a origem da vida, tópico do qual é considerado um dos maiores conhecedores mundiais.

funcional como também o armazenamento e distribuição desses dados (Thornton, 1998; Skrabanek *et al.*, 2008).

### 1.1.2. Biologia de Sistemas

A determinação das estruturas primárias de macromoléculas biológicas de um organismo (genômica) já não é um fator limitante para o seu entendimento funcional. Da mesma forma, a compreensão da regulação da transcrição gênica (transcriptômica) tem sido bastante fortalecida pela disponibilidade de diversas tecnologias, como por exemplo, os microarranjos. O passo adiante é a compreensão das funções bioquímicas de todos os produtos gênicos (proteômica), o relacionamento entre estes produtos e destes com os substratos e moléculas reguladoras (metabolômica), e a totalidade de suas interações (interatômica). Métodos bioquímicos tradicionais que visam o entendimento de tudo isso requerem uma demanda grande de tempo, e por isso, entram em cena métodos mais eficientes para responder a estas perguntas, ou pelo menos para sugerir possíveis soluções na era pós-genômica (Kandpal *et al*., 2009).

A evolução das tecnologias de alto desempenho para a análise e representação de moléculas biológicas criou uma mudança de paradigma na investigação das ciências da vida. Uma abordagem integrativa, usando dados provenientes de todas as “ômicas”, leva ao conceito de Biologia de Sistemas, que investiga os sistemas como um todo e não como componentes individuais de forma isolada (Snyder & Gallagher, 2009). A Biologia de Sistemas é um campo da ciência que descreve quantitativamente a interação entre os componentes individuais da célula. O objetivo final de tal abordagem é desenvolver modelos computacionais de sistemas complexos de modo que a resposta do sistema biológico para qualquer tipo de perturbação (como por exemplo, ambiental, mutacional, etc) possa ser predita. Do ponto de vista da engenharia, a biologia de sistemas pode ser considerada como um conjunto de processos físico-químicos e mecânicos interligados e que ocorrem em paralelo e em série em um organismo (Hatzimanikatis, 1999).

Na visão da biologia de sistemas, o conhecimento biológico prévio juntamente com os dados experimentais é usado para definir as relações entre os vários

componentes do sistema e extrair parâmetros biológicos. Estes parâmetros serão usados na construção de modelos bioquímicos ou matemáticos na geração de hipóteses que possam explicar um determinado fenômeno biológico. A comparação entre os dados experimentais oriundos da análise do comportamento do sistema, juntamente com os dados do modelo gerado pode ser usada para examinar este modelo ou o conjunto de parâmetros que melhor descreve estes dados, e assim, tais dados podem ser aperfeiçoados com o auxílio de novos desenhos experimentais até chegar a um modelo refinado (Aggarwal & Lee, 2003). Este esquema é demonstrado na Figura 1.

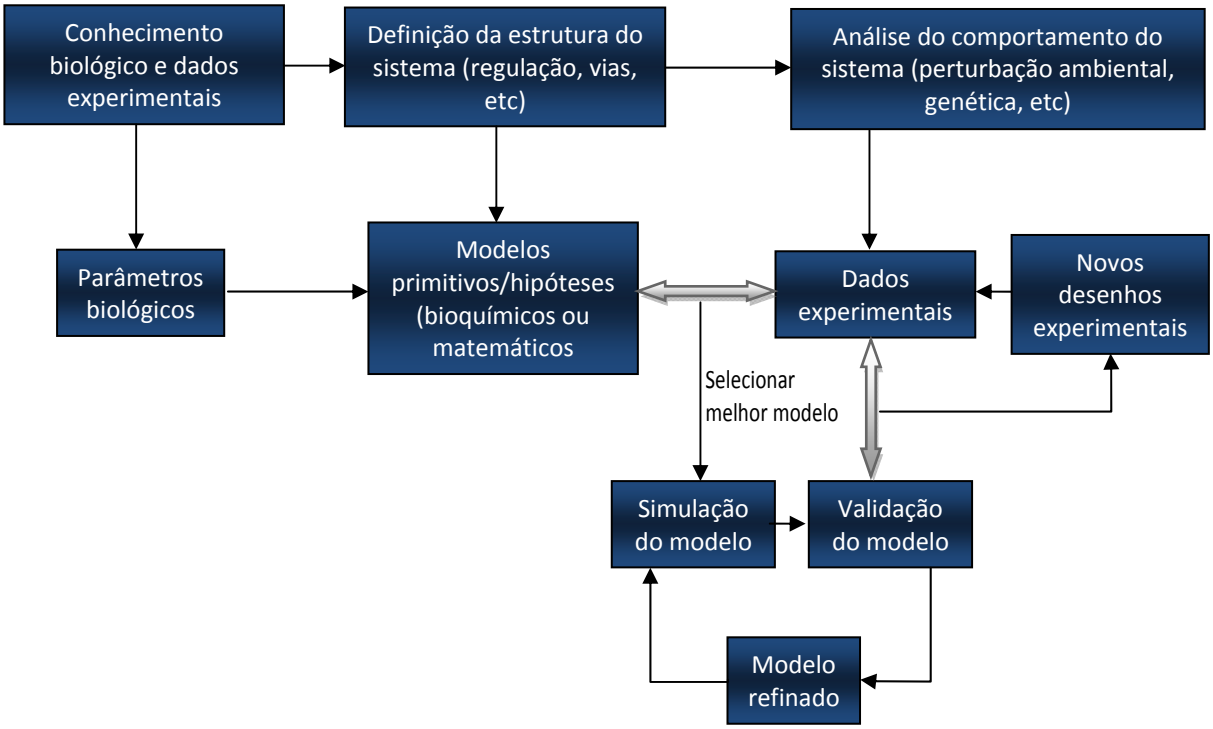



**Figura 1:** Esquema da simulação de modelo na biologia de sistemas. Adapatado de Aggarwal & Lee, 2003.

O entendimento da função de processos individuais e suas interações têm uma grande importância no avanço das pesquisas em diversas áreas, como por exemplo, na biotecnologia industrial, agrícola e farmacêutica. Esta última vem ganhando espaço, principalmente no desenvolvimento de novas drogas (Hatzimanikatis, 1999). Nos últimos anos, a metabolômica desempenha papel cada vez mais importante na pesquisa farmacêutica e de desenvolvimento. O entendimento do perfil metabólico de fluidos biológicos e de tecidos pode fornecer uma visão panorâmica das mudanças de metabólitos endógenos em resposta a

perturbações celulares ocasionadas por algumas doenças. Além disso, pode complementar os dados de transcriptômica e proteômica na identificação de novos *loci* de características quantitativas nos estudos de associação em um genoma (Xu *et al*., 2009).

### 1.1.3. Genômica Funcional

A conclusão bem sucedida de vários projetos genoma levou a uma nova etapa nas ciências biológicas, a era pós-genômica. Um dos mais importantes desafios nesta nova era é a elucidação de funções celulares, que podem ser vistas como um comportamento particular de um sistema complexo de interações entre várias proteínas (Ouzounis & Valencia, 2003).

As sequências genômicas fornecem informações que permitem a investigação de processos biológicos através de abordagens abrangentes. A análise de genomas também teve impacto na medicina, uma vez que se tornou possível a identificação de genes envolvidos em várias doenças, assim como a presença de mutações em tais genes (Fields *et al*., 1998). A análise sistemática da função de um gene é muito mais complexa do que o sequenciamento em larga escala das sequências de um genoma (Raamsdonk *et al*., 2001). As sequências provenientes dos projetos genoma são utilizadas para predizer o primeiro conjunto de proteínas possíveis de um dado organismo, comparando estas sequências com todas as outras conhecidas em bancos de dados. Para muitas destas proteínas, é possível assinalar uma provável função quando existe uma correspondência razoável com uma sequência protéica a partir do genoma de outro organismo e com uma função pré-estabelecida, uma vez que é presumida que a similaridade na sequência geralmente reflete uma similaridade de função. O conhecimento do papel biológico de proteínas comuns entre organismos certamente pode, muitas vezes, fornecer forte inferência ao seu papel em outros organismos (Fields *et al*., 1998).

A Genômica Funcional congrega análises de dados provenientes de transcritos de genes, proteínas e metabólitos de modo a responder à pergunta colocada por todos os projetos de sequenciamento de genoma: qual é a função biológica de cada gene e seus produtos codificados? Assim, a genômica funcional

oferece a perspectiva de obter uma visão verdadeiramente holística da vida (Colebatch *et al*., 2003).

A análise em grande escala na genômica funcional só se tornou viável devido ao desenvolvimento de sistemas de armazenamento e consulta a esses dados. A utilização de reconhecimento de padrões aprimorou significativamente o desenvolvimento de repositórios de dados públicos, melhorando assim, o compartilhamento desses dados (Jones *et al*., 2004).

#### 1.1.3.1. Anotação Funcional

> "*Anotação de sequências é um processo múltiplo, pelo qual uma ou mais sequências brutas de DNA ou de aminoácido são analisadas com a finalidade de atribuir características biológicas para o entendimento do contexto biológico em que estas se inserem, ou seja, sua função.*" (Stein, 2001).

Uma vez obtidos os dados do sequenciamento das moléculas de DNA, é preciso saber o que representa cada fragmento ou região. O processo de anotação consiste em identificar, analisar e interpretar sinais nas sequências de DNA, extraindo o seu significado biológico. Uma sequência per se tem pouco ou nenhum valor, se não for anotada e estas anotações podem representar um volume de informações muito maior que as próprias sequências (Valencia, 2005).

Em projetos genoma, este processo normalmente é realizado em três etapas: (i) anotação de sequências de nucleotídeos; (ii) de sequências protéicas; e (iii) de processos biológicos, como exemplificado na figura 2. Na primeira etapa, procura-se identificar a natureza de uma determinada sequência, tentando desvendar a sua localização no genoma, se faz parte de uma região codificante ou regulatória, se será transcrita ou não, se faz parte de uma região repetitiva ou se apresenta algum marcador genético conhecido em seu interior. A finalidade desta etapa de anotação é situar os genes num contexto genômico, ou seja, mapear no genoma do organismo cada um dos possíveis genes e caracterizar as regiões não-gênicas, sejam estas regulatórias ou não. Assim, a anotação busca caracterizar regiões

codificadoras para proteínas, regiões intergências e reguladoras, fronteiras íntron/éxon, sítios promotores e de iniciação e término de transcrição, regiões codificadoras para RNA´s estruturais e funcionais, sequências repetitivas, elementos centroméricos e teloméricos onde aplicável, e outros elementos estruturais no genoma. Existem diversos algoritmos de predição gênica que são utilizados nesta etapa; os mais utilizados buscam por elementos como o códon de iniciação de proteínas (a trinca de nucleotídeos ATG) e códons de terminação na mesma fase de leitura, seguido de abordagens estatísticas para estimar a probabilidade da fase aberta de leitura ser de fato codificadora para uma proteína. A etapa seguinte consiste na análise das características protéicas a partir das sequências de aminoácidos preditas na fase anterior, no processo de anotação das sequências protéicas. Existem diversas maneiras de inferir função a uma proteína; um fluxograma representativo desta etapa é mostrado na figura 3, onde a anotação pode ser feita através da transferência de uma anotação previamente descrita por homologia na sequência. Esta homologia funcional pode ser inferida através de busca por similaridade em bancos de dados diversos, tanto de sequências quanto de motivos, domínios, estruturas, etc. A anotação automática requer uma validação manual, juntamente com uma mineração nos bancos de dados para corrigir possíveis erros gerados na automatização da predição de funções protéicas.

A etapa de anotação seguinte é bastante desafiadora, pois relaciona a informação genômica com os processos biológicos, gerando um mapa funcional do organismo estudado, sendo plausível identificar/propor quais vias bioquímicas estão completas ou incompletas no organismo e quais vias alternativas este possui. Aqui é possível explanar como o metabolismo do organismo pode influenciar seu modo de vida e seu comportamento (Stein, 2001). A disponibilização recente de dados de redes de interação de proteínas dentro das células em diversas espécies vem aumentando o desenvolvimento de métodos computacionais para interpretar estes dados de modo a auxiliar na elucidação de funções protéicas (Sharan *et al*., 2007).

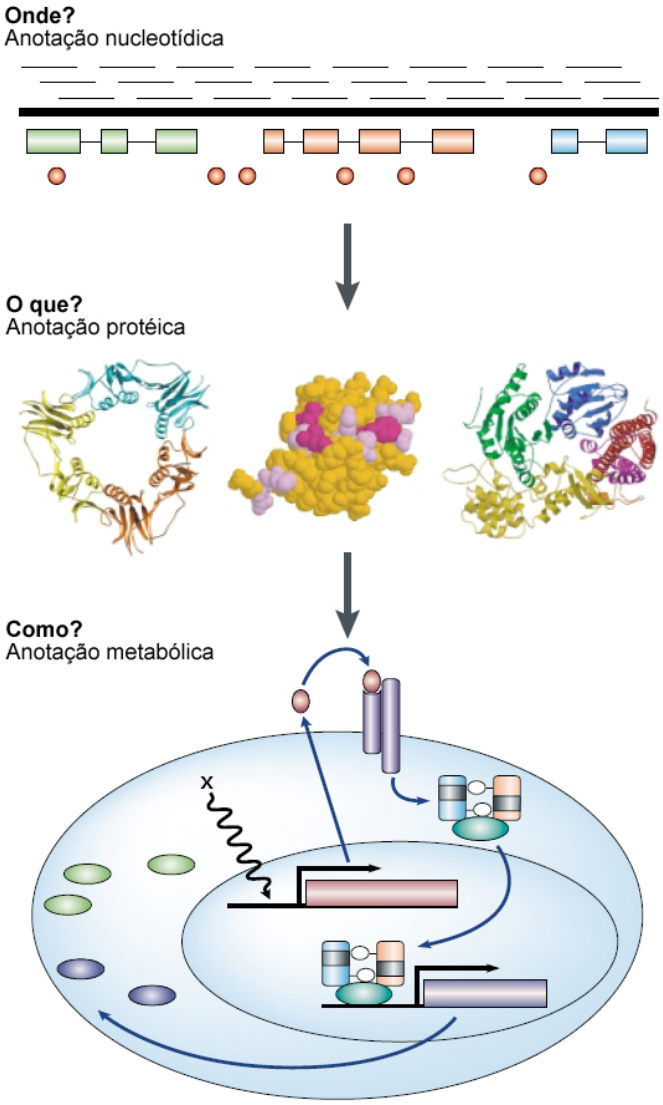



**Figura 2:** Os três níveis representativos da anotação genômica: onde, o que e como? Adaptado de Stein, 2001.


**Figura 3:** Fluxograma representativo da visão geral do processo automático de anotação genômica. Adaptado de Frishman, 2007.

Outra maneira de visualizar a informação sobre a função de componentes celulares, suas interações, localização espacial e alterações ocorridas durante a evolução é através do exame de diferentes níveis hierárquicos de anotação do genoma (figura 4). A anotação unidimensional envolve a identificação de genes no genoma e a inferência de função conhecida ou predita para identificar os produtos gênicos. A segunda dimensão especifica os componentes celulares e suas interações. As interações físicas e químicas entre estes componentes levam a reconstruções de redes que simulam as duas dimensões. O conhecimento sobre a organização dos componentes dentro da célula representa a terceira dimensão e está relacionada com papéis importantes na função celular. A última dimensão está

relacionada com as modificações nas sequências genômicas que podem ocorrer durante a evolução e tem grande importância na plasticidade genômica, assim como na identificação do surgimento e extinção de algumas funções dentro de uma célula (Reed *et al*., 2006).

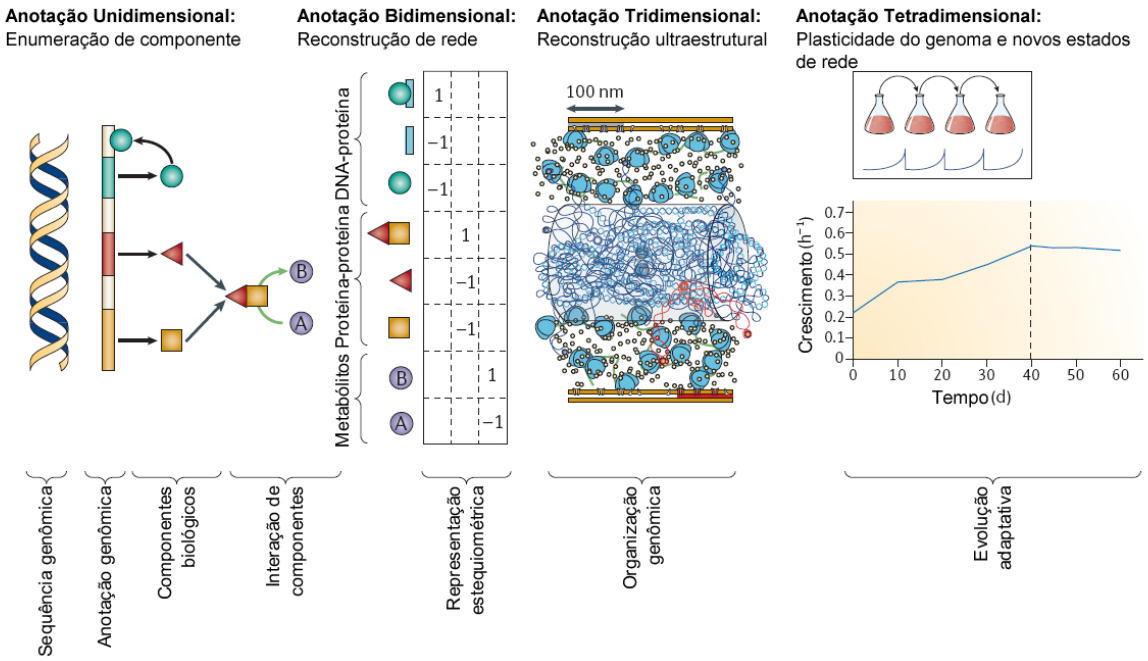



**Figura 4:** Anotação multidimensional. A anotação unidimensional representa uma lista de componentes da rede. A segunda dimensão pode ser representada pela interação entre estes componentes, onde uma matriz de coeficiente estequiométrico é usada para representar as interações dos componentes. A organização estrutural do genoma também pode ser representada espacialmente numa anotação tridimensional. Modificações na sequência genômica podem ser caracterizadas na quarta dimensão da anotação. Adaptado de Reed *et al*., 2006.

Embora os sistemas de anotação atuais tentem minimizar os erros, ainda existem problemas comuns relevantes à automatização desse processo. Entre estes problemas, podemos citar a presença de erros de anotação em bancos de dados públicos que leva a propagação de tais erros em novas inferências. Estes erros podem ser introduzidos durante os desenhos experimentais e na obtenção e interpretação dos resultados, sendo de difícil detecção. Quando a informação experimental errônea é estocada nos bancos de dados, os erros são expandidos pelos processos automatizados de anotação funcional (Valencia, 2005). Uma validação manual utilizando informações curadas tende a diminuir a propagação desses erros. Entretanto, é extremamente difícil reproduzir computacionalmente o

complexo processo de decisão de um curador humano, que leva em consideração diversos parâmetros, como a análise de dados disponíveis na literatura, a proposição de novos desenhos experimentais, e resultado de outras análises computacionais (Bork & Bairoch, 1996; Galperin & Koonin, 1998).

É importante ressaltar que os dados armazenados nesses repositórios devem ser anotados de uma maneira consistente, levando em consideração termos funcionalmente equivalentes, como proposto pelo *Gene Ontology Consortium* (http://www.geneontology.org). O objetivo do *Gene Ontology* (GO) é produzir um vocabulário dinâmico controlado, que pode ser aplicado a todos os organismos no que diz respeito a processos biológicos, funções moleculares e componentes celulares (Ashburner *et al*., 2000). A classificação feita pelo GO é baseada em dados da literatura científica e nas relações evolutivas das entidades anotadas, para conferir uma declaração formal sobre a função de um gene. As anotações são feitas usando diferentes provas, que podem ser utilizadas para estimar sua confiabilidade (Thomas *et al*.,2007).

### 1.1.4 Genômica estrutural

#### 1.1.4.1. Função estrutural

Um exemplo significativo da importância que o conhecimento da estrutura tridimensional de macromoléculas fornece para o entendimento de processos biológicos é o próprio modelo de estrutura em dupla-hélice do ácido desoxirribonucléico (DNA) proposto por Watson e Crick (1953). Mais de cinquenta anos depois, inúmeras estruturas de diferentes macromoléculas já foram propostas (Ring & Cohe, 1993; Berman *et al*., 2000)

A estrutura tridimensional de uma proteína é muito mais informativa do que sua sequência de aminoácidos por si só. O conhecimento da estrutura nos permite explicar a bioquímica do mecanismo pelo qual a proteína implementa sua funcionalidade (Watson *et al.*, 2005). Se a função é desconhecida, uma similaridade baseada na estrutura tridimensional (3D) entre duas sequências pode revelar mais detalhes sobre sua função. Se a estrutura 3D possui uma conformação com uma

determinada função conhecida, pode-se inferir que a proteína possui uma função similar ou idêntica à de outras proteínas com essa conformação. Uma possível explicação para esse fato é que, provavelmente, descendem de um ancestral comum. Em raríssimos casos, a função de proteínas de um ou de ambos pode se alterar durante a evolução, enquanto suas conformações permanecem inalteradas e, portanto, nestes casos, a mesma conformação pode dar origem a duas funções distintas (Sánchez *et al*., 2000; Cherkasov *et al*., 2004). Existem conformações presentes em diversas proteínas que desempenham funções diferentes; porém, a maior parte das conformações atualmente conhecidas está associada a uma única função (Friedberg, 2006; Todd *et al.*, 1996). Além disso, a estrutura tridimensional protéica por sua vez, é mais bem preservada do que a estrutura primária em si, tanto que proteínas com pouca ou nenhuma similaridade de sequência podem manter uma semelhança estrutural (Rost, 1997). Os sítios ativos das proteínas fazem parte das regiões funcionalmente importantes da estrutura, tendo uma maior probabilidade de terem sido fortemente conservados ao longo do tempo evolutivo como visto na figura 5 (Sjolander, 2004; La *et al.*, 2005; Marsden *et al*., 2006).

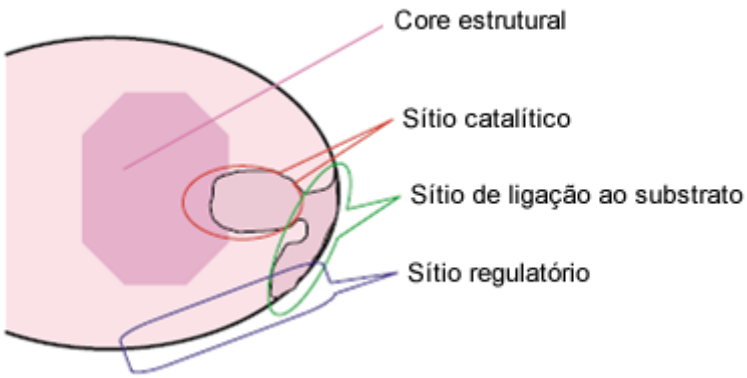



**Figura 5:** Representação esquemática das regiões funcionais típicas de uma proteína. O modelo geral de evolução sugere que a maior conservação está relacionada com o sítio catalítico, seguida pel cerne (*core*) estrutural, especificidade do substrato e regiões regulatórias.

O sucesso na genômica estrutural pode ser visto no aumento do número de estruturas depositadas no maior banco especializado em estruturas tridimensionais, o *Protein Data Bank* (PDB) (Berman *et al*., 2000). Este sucesso se deve ao desenvolvimento de tecnologias experimentais e computacionais para a predição de estrutura 3D. Entretanto, houve um aumento na elucidação de estruturas que não possuem uma função definida, denominadas apenas como hipotéticas. Embora

forneçam uma valiosa contribuição ao conhecimento da estrutura das proteínas de forma geral, o seu valor pode ser significativamente reforçado pelo conhecimento das funções biológicas que desempenham (Watson *et al.*, 2005).

### 1.1.4.2. Modelagem Comparativa

O número de estruturas terciárias resolvidas, através de cristalografia, Resonância Magnética Núclear ou outros métodos, e armazenadas em bancos de dados ainda é bem menor do que a quantidade de registros contendo informações sobre a estrutura primária de muitas proteínas biologicamente relevantes. Para minimizar essa situação, foram desenvolvidos métodos de predição de estrutura secundária e terciária. Uma sequência aminoacídica pode ser utilizada na predição de elementos de estrutura secundária, como trechos de alfa-hélice ou folha-beta, que agrupados auxiliam na determinação da estrutura terciária (Garnier, 1990). O padrão de enovelamento de proteínas permite que aminoácidos dispostos em regiões distantes possam interagir, formando distintos padrões conformacionais. Alinhamentos estruturais entre proteínas pertencentes a uma mesma família podem identificar regiões similares envolvidas com uma determinada função, e assim, ser um método importante de inferência de função (Laskowski *et al.*, 2003; Marsden *et al.*, 2006). Computacionalmente são aplicadas duas estratégias para a predição de estrutura tridimensional: (i) métodos físicos, que se baseiam nas interações entre os átomos, incluindo dinâmica molecular e minimização de energia; e (ii) métodos empíricos, que dependem de estruturas que já foram determinadas experimentalmente.

Um dos métodos que contribui para a predição da estrutura tridimensional de proteínas é a modelagem comparativa, um procedimento computacional que prediz a estrutura da proteína o mais próximo da estrutura cristalográfica utilizando uma estrutura molde resolvida tridimensionalmente, evolutivamente relacionada, cuja estrutura primária possui similaridade com a sequência da proteína que se quer modelar (Goldsmith-Fischman & Honig, 2003). Tem como objetivo descrever e prever estruturas moleculares, propriedades do estado de transição e equilíbrio de reações, propriedades termodinâmicas, entre outras. Os métodos utilizados em modelagem molecular envolvem estudos de minimização de energia, análise

conformacional, simulações de dinâmica molecular, entre outros; e são aplicáveis desde átomos isolados a macromoléculas (Williams & Lemke, 2002; Sánchez *et al*., 2000).

Na modelagem comparativa, quatro etapas são fundamentais: i) identificação do molde de referência, que é a seleção de sequências e estruturas que possuem similaridade com a sequência alvo; ii) alinhamento entre as sequências de referência e alvo, para determinar o melhor molde a ser utilizado nas etapas seguintes; iii) construção do modelo, que utiliza as informações contidas no alinhamento da etapa anterior em principalmente duas técnicas de construção, corpos rígidos[2] ou satisfação de restrições espaciais[3]; iv) validação do modelo, através da identificação de possíveis erros relacionados aos métodos empregados, à escolha das referências e ao alinhamento das sequências alvo e de referência. Nesta fase o modelo escolhido é aceito ou rejeitado, sendo que neste último caso todo o protocolo deve ser revisto no intuito de melhorar a escolha das referências, o alinhamento ou a utilização de outros métodos (Garnier, 1990). A qualidade do modelo gerado é consequência das ligações entre os átomos envolvidos, os ângulos conformacionais gerados, as ligações peptídicas entre os aminoácidos que compõem a proteína, a planaridade dos anéis e os ângulos de torção nas cadeias principais e laterais. Uma maneira de quantificar tais parâmetros é através do grau de similaridade entre as estruturas tridimensionais do modelo gerado com o de sua proteína molde. Este grau de similaridade é calculado pelo desvio médio quadrático (*Root Mean Square Deviation - RMSD*), que mede as distâncias interatômicas entre as duas estruturas (Valas *et al.*, 2009). O gráfico de Ramachandran (Ramakrishnan, 2001) exemplificado na figura 6 também identifica resíduos alocados em regiões favoráveis e desfavoráveis através da correlação entre os ângulos *phi* e *psi* (figura 7) entre os resíduos presentes na estrutura do modelo gerado, levando a um

[2] Construção de modelos por partes, baseado no fato de que proteínas homólogas possuem regiões estruturalmente conservadas, como alfa-hélice e folhas beta. Estas estruturas secundárias são alinhadas, considerando a média da posição de carbonos-alfa das estruturas das regiões estruturalmente conservadas nas sequências de referência, enquanto que as regiões que não satisfazem as exigências são normalmente as de alças que conectam as regiões conservadas.

[3] Construção de modelos a partir do alinhamento entre as sequências referência e de estudo, onde se obtém informações das restrições espaciais (distâncias e ângulos) do molde que são transferidas para o modelo. O tamanho das ligações e os ângulos preferenciais são obtidos de campos de força enquanto que os de átomos não ligados são obtidos por análise estatística de um grupo representativo de estruturas conhecidas.

melhoramento deste modelo (Marti-Renom *et al*., 2002). Métodos de simulação de dinâmica molecular também têm sido utilizados para melhorar os modelos gerados (Fan & Mark, 2004).

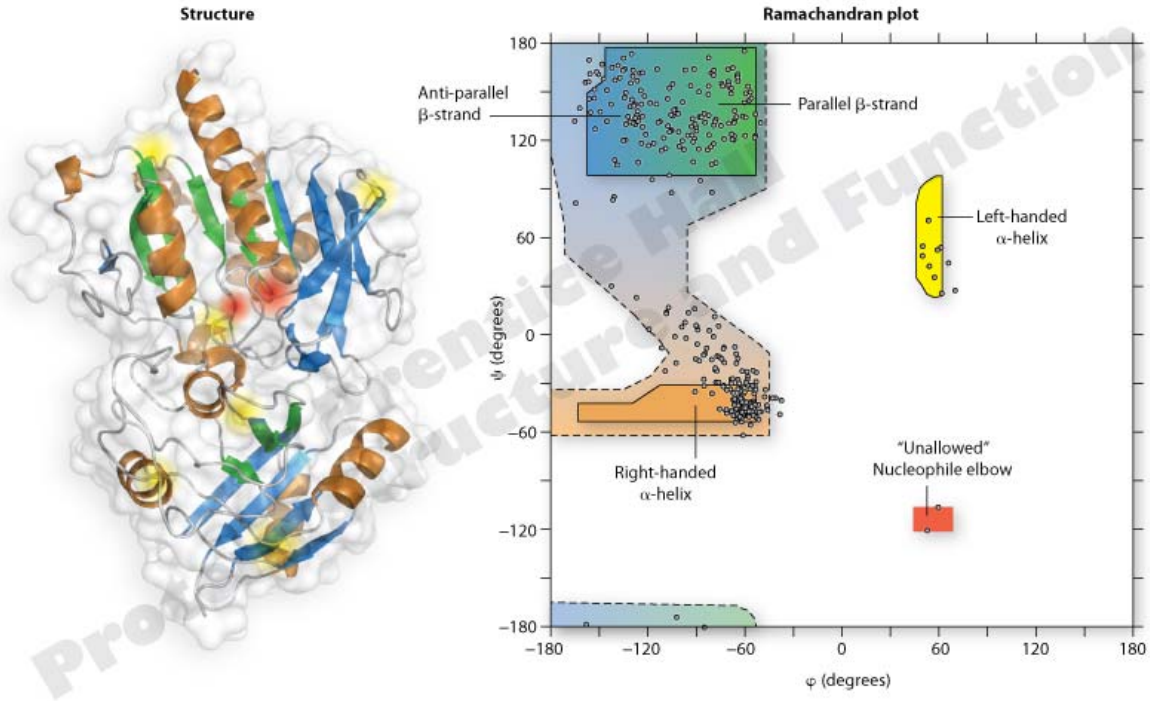



**Figura 6:** Exemplo de gráfico de Ramachandran apresentando correlações entre os ângulos *phi* (φ) e *psi* (ψ) de uma cadeia polipeptídica. No gráfico, cada ponto representa um resíduo aminoacídico da proteína e quais destes se encontram nas regiões energicamente mais favoráveis (coloridas) e desfavoráveis (branca), com exceção da prolina, e orienta a avaliação da qualidade estereoquímicas de modelos teóricos ou experimentais de proteínas. Fonte: http://adam.steinbergs.us/images/books/Proteins/ ramachandran-plot1.jpg

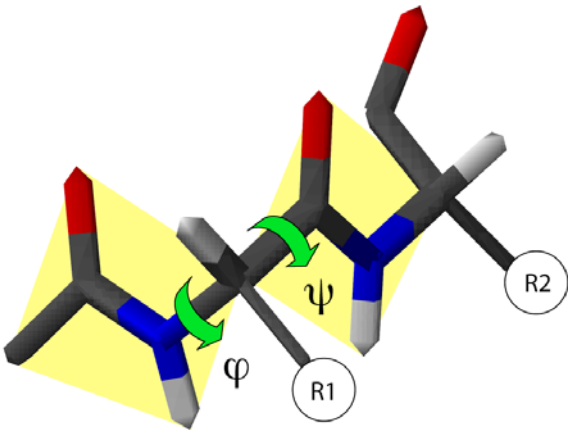



**Figura 7:** Ilustração representativa dos dois graus de liberdade presentes numa ligação peptídica, ângulos *phi* (φ) e *psi* (ψ). Fonte: http://en.wikipedia.org/wiki/File: Dihedrals.png

### 1.1.4.3. MHOLline

Um dos *workflows* biológicos utilizados para a predição automatizada de estruturas tridimensionais de proteínas, e que foi usado nesse estudo, é o MHOLline (http://www.mholline.lncc.br). Este reúne um conjunto específico de programas de análise de estrutura protéica, incluindo métodos de modelagem comparativa e de identificação de regiões transmembranares (Cavalcanti *et al.*, 2005).

O MHOLline inicia com um arquivo de entrada no formato FASTA contendo uma ou mais sequências de proteínas. Para a identificação do molde a ser utilizado na construção do modelo, o MHOLline usa o algoritmo BLAST - *Basic Local Alignment Search Too*l (Altschul *et al*., 1997) para a busca de sequências com estrutura tridimensional resolvida depositadas no PDB. Uma etapa de refinamento do modelo é feita com a utilização de um programa desenvolvido especialmente para este *pipeline* denominado BATS (*Blast Automatic Targeting for Structures*), que identifica as sequências para que a técnica de modelagem comparativa possa ser aplicada. A escolha de tais sequências é feita a partir do arquivo de saída do BLAST, usando os parâmetros de pontuação resultantes dos alinhamentos e segue os critérios definidos na tabela 1. As sequências alocadas no grupo 0 (G0) ou no grupo 1 (G1) retornam para o usuário em novos arquivos FASTA, que podem ser novamente submetidas ao MHOLline. As sequências que se encontram no grupo 3 (G3) não podem ter sua estrutura tridimensional predita por técnicas de modelagem comparativa, também retornando para o usuário em arquivo FASTA. Já as sequências reunidas no grupo 2 (G2) continuam no fluxo do MHOLline. Para cada sequência de entrada neste grupo é escolhida apenas uma estrutura PDB como molde para a etapa de modelagem comparativa, realizada através do programa Modeller (Eswar *et al*., 2006). A escolha é feita selecionando a maior pontuação do BATS para cada sequência de entrada e, em caso de empate, a melhor resolução do PDB é utilizada como critério de decisão. O cálculo da pontuação BATS é realizado pela equação:

$$S_f = W_iP_i + W_sP_s + W_nP_n + W_lP_l + W_gP_g$$

onde $W_i$, $W_s$, $W_n$, $W_l$, $W_g$ e $P_i$, $P_s$, $P_n$, $P_l$, $P_g$ são, respectivamente, o peso e os valores atribuídos a identidade e similaridade do BLAST, o número de aminoácidos alinhados, o *Length Variation Index* (LVI) e o *Gap Relative Strength Index* (GRSI). O LVI e o GRSI são calculados segundo as equações:

$$LVI = \frac{Q_L - N_{Ali}}{Q_L}$$

$$GRSI = \frac{wG_a + G_g}{Q_L}$$

onde $Q_L$ é o comprimento da sequência de consulta, $N_{Ali}$ é o número de aminoácidos alinhados entre sequências *query* e *subject*, *w* é a penalidade para a abertura de uma lacuna (*gap*) no alinhamento, *Ga* representa o número de aberturas de novas lacunas e *Gg* o número de lacunas subsequentes abertas.

**Tabela 1:** Grupos classificatórios do BATS com os respectivos parâmetros de pontuação.

| Grupos | Parâmetros de pontuação |
|---|---|
| G0 | Sequência sem alinhamento |
| G1 | E-value > $10e^{-5}$ ou Identidade < 0,15 |
| G2 | E-value ≤ $10e^{-5}$ e Identidade ≥ 0,25 e *LVI* ≤ 0,7 |
| G3 | E-value ≤ $10e^{-5}$ e (0,15 ≤ Identidade < 0,25 ou *LVI* > 0,7) |

Uma ferramenta denominada Filtros classifica as sequências selecionadas pelo BATS em sete grupos distintos de acordo com a qualidade do modelo a ser gerado, conforme a tabela 2.

**Tabela 2:** Classificação feita pela ferramenta Filtros, de acordo com a qualidade dos modelos gerados a partir dos parâmetros obtidos pelo BATS.

| Qualidade | Identidade | *LVI* |
|---|---|---|
| 1. Muito Alta | ≥ 75% | ≤ 0,1 |
| 2. Alta | ≥ 50% e < 75% | ≤ 0,1 |
| 3. Boa | ≥ 50% | > 0,1 e ≤ 0,3 |
| 4. Média para Boa | ≥ 35% e < 50% | ≤ 0,3 |
| 5. Média para Baixa | ≥ 25% e < 35% | ≤ 0,3 |
| 6. Baixa | ≤ 25% | > 0,3 e ≤ 0,5 |
| 7. Muito Baixa | ≤ 25% | > 0,5 e ≤ 0,7 |

Outra ferramenta, a ECNGet, atribui um número classificatório quando a sequência é uma enzima. Essa classificação é baseada na nomenclatura de enzimas estipulada pela IUBMB (*International Union of Biochemistry and Molecular Biology*). Somente as sequências alocadas no G2 do BATS passam por esta ferramenta, que utiliza o código de identificação do PDB para associar um *EC number* à sequência

Após a identificação do molde, um alinhamento global é gerado e a construção do modelo tridimensional é executada. Os modelos gerados passam por uma validação de sua qualidade estereoquímica através do programa PROCHECK (Laskowski *et al*., 1993). Todas as sequências FASTA passam pelo programa HMMTOP (Tusnády & Simon, 1998) a fim de identificar as regiões transmembrana. Todas as etapas realizadas pelo MHOLline se encontram representadas na figura 8.

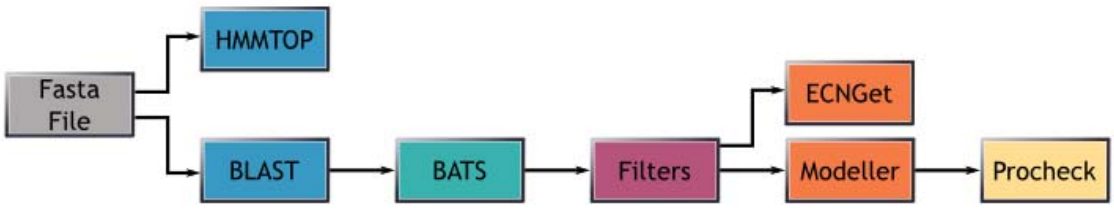



**Figura 8:** Fluxograma representativo de todas as etapas executada pelo MHOLline. Fonte: http://www.mholline.lncc.br/index.php?pag=2.

Para cada programa executado é gerado um arquivo final e por fim um relatório sintetizando todas as etapas realizadas.

### 1.1.4.4. Desenho Racional de Fármacos

O extensivo crescimento no campo da biologia molecular nas últimas décadas tem levado ao desenvolvimento de novas e poderosas ferramentas experimentais e computacionais que permitem a análise de respostas biológicas complexas em resposta à exposição a substâncias químicas, como por exemplo, os fármacos (Harrill & Husyn, 2008). Estes são essenciais para a prevenção e tratamento de doenças. O procedimento de desenvolvimento de tais fármacos é um trabalho custoso e requer um tempo considerável. Para entender os desafios na busca do fármaco ideal, diferentes métodos são empregados, sejam experimentais ou computacionais, levando em conta abordagens multidisciplinares que constituem a base do desenho racional de fármacos (Mandal *et al*., 2009). Recentemente, a descoberta de novos fármacos tornou-se possível devido à disponibilidade de estruturas tridimensionais de biomoléculas resolvidas por ressonância magnética nuclear e difração de raio-X, além do desenvolvimento de metodologias de *docking*[4] e dinâmica molecular[5] (Müller, 2009).

O desenho racional de fármacos pode ser dividido em duas categorias: (i) desenvolvimento de pequenas moléculas com as propriedades desejadas, cujo papel funcional nos processos celulares e a informação estrutural são conhecidos; e (ii) desenvolvimento de pequenas moléculas com propriedades pré-definidas, com função e estrutura tridimensional não necessariamente conhecidas. Estas são normalmente selecionadas através de análises em larga escala de genomas e proteomas. O desenvolvimento dessas moléculas, após a identificação de um possível alvo, requer uma análise sob vários aspectos, tais como: avaliação da ligação entre receptor e ligante; lipofilicidade e hidrofilicidade; absorção, distribuição, metabolismo e excreção; biodegradação, toxicidade, entre outros (Mandal *et al*., 2009).

[4] Método que prevê a orientação preferencial de uma molécula quando vinculada a(s) outra(s) para formar um complexo estável.

[5] Tipo de simulação computacional no qual os átomos e moléculas podem interagir durante um período de tempo por aproximações físicas conhecidas, dando uma visão do movimento das partículas.

Tradicionalmente, a maioria dos protótipos de fármacos disponíveis atualmente utiliza uma triagem cega na busca de alvos, através da varredura de um grande número de compostos sintéticos ou produtos naturais com atividade contra patógenos íntegros, em ensaios *in vitro* e em modelos animais. Esta abordagem é bastante trabalhosa e com elevado custo financeiro, levando ao aumento da utilização de técnicas baseadas em um exame mais focado, buscando compostos contra alvos moleculares específicos (Rockey & Elcock, 2002). A utilização de dados estruturais permite a análise de propriedades específicas de um composto que podem influenciar na interação com seu receptor. Além disso, alinhamentos estruturais entre diferentes moléculas fornecem um índice de similaridade que pode ser correlacionado com a atividade farmacológica (Cherkasov *et al.,* 2004).

O “gargalo” anteriormente existente para o desenho racional de fármacos, a pequena porcentagem de genes conhecidos, foi rapidamente solucionado com o sequenciamento de diversos genomas, possibilitando o desenvolvimento de novas abordagens computacionais para a identificação de genes candidatos ao desenvolvimento de novas drogas e seus respectivos desenhos experimentais. A informação genômica, combinada com estratégias de mineração de dados levou à reconstrução de vias bioquímicas completas, onde é possível identificar moléculas que participam de passos essenciais e que possam ser eleitas como novos alvos para potenciais intervenções terapêuticas (Jorgensen, 2004). Assim, a genômica estrutural tem sido útil na identificação de alvos moleculares, utilizando métodos de alto desempenho na determinação da estrutura tridimensional e proporcionando uma ferramenta importante para triagem inicial de drogas-alvo e sua otimização (Congreve *et al.*, 2005). Uma abordagem de genômica funcional tem sido usada para preencher a lacuna existente na inferência de função das estruturas primárias para a identificação de possíveis alvos de drogas utilizando técnicas de bioquímica, biologia celular e molecular, e bioinformática. Esta abordagem permite uma melhor compreensão do papel desempenhado pelas moléculas nas vias metabólicas envolvidas em uma variedade de doenças (Lindsay, 2003).

### 1.1.5. Enzimas e metabolismo

#### 1.1.5.1. Conceitos

As enzimas têm um papel crucial em muitos aspectos da vida. Diversas civilizações já as usavam há milhares de anos sem entender o seu funcionamento. A existência das enzimas já é conhecida há mais de um século. Um dos primeiros estudos foi realizado pelo químico sueco Jakob Berzelius Jon, que em 1835 identificou sua ação catalítica. Mas somente em 1926, James B. Summer conseguiu obter uma forma pura desta molécula, isolando e cristalizando urease de feijão (Bennet & Frieden, 1969). Ao longo do tempo, a ciência tem desvendado o mistério acerca das enzimas e aplicado esse conhecimento para fazer melhor uso em um número crescente de aplicações, como, por exemplo, na produção de alimentos, tecidos, combustível, entre outras (Ke & Klibanov, 2000). A utilização de enzimas no diagnóstico de doenças é um dos importantes benefícios derivados da intensa pesquisa em bioquímica desde a década de 1940. No entanto, foi somente num passado recente que o interesse em enzimologia para diagnóstico se multiplicou (Bohager, 2006).

Enzimas são proteínas altamente especializadas capazes de catalisar as reações que ocorrem nos sistemas biológicos; são produzidas por todos os organismos vivos e essenciais para todos os processos metabólicos. Apenas um pequeno número de moléculas de ácido ribonucléico (RNA) é capaz de exercer alguma propriedade catalítica (Nelson & Cox, 2000). A catálise é uma maneira de fazer com que uma reação química ocorra de forma mais rápida, reduzindo a quantidade de energia necessária para o processo. Na ausência das enzimas, estas reações ocorreriam de forma muito lenta impossibilitando um metabolismo adequado. Como controlam a maior parte das reações que ocorrem no metabolismo dos organismos, a correta identificação, caracterização funcional e localização no genoma se tornam imprescindíveis (Bohager, 2006).

Como toda proteína, as enzimas são formadas por aminoácidos unidos através de ligações peptídicas, e muitas delas requerem a presença de outros compostos (cofatores) para que sua atividade catalítica possa ser exercida. Este

complexo ativo é referido como holoenzima, ou seja, a apoenzima (porção proteíca) juntamente com o cofator (Holum, 1968). Os cofatores podem ser de diferentes naturezas: coenzima (substância orgânica não protéica), grupo prostético (substância orgânica protéica), íon metal. Uma das propriedades das enzimas que as tornam tão importantes é a especificidade que apresentam em relação às reações que catalisam. Existem poucas enzimas que apresentam uma especificidade absoluta. Em geral, existem quatro tipos distintos de especificidade: (i) especificidade absoluta, onde a enzima catalisa apenas uma reação; (ii) especificidade de grupo, onde a enzima vai agir somente sobre as moléculas que têm características funcionais específicas; (iii) especificidade *Linkage*, em que a enzima age em um determinado tipo de ligação química, independentemente do restante da estrutura molecular; e (iv), especificidade estereoquímica, no qual a enzima age em um isômero estéreo ou ótico particular (Martinek, 1969).

A reação básica desempenhada pelas enzimas pode ser representada por:

$$E + S \rightarrow ES \rightarrow P + E$$

onde E representa a enzima, S o substrato, ES o complexo formado entre a enzima e o substrato, e P o produto gerado na reação. Vários fatores influenciam as reações enzimáticas, como temperatura, pH, concentração da enzima e do substrato, presença de inibidor ou ativador (Harrow & Mazur, 1958).

Como apresentam um alto grau de especificidade por seus substratos e são fundamentais para qualquer processo bioquímico, as enzimas, normalmente, agem em sequência organizada, catalisando reações sucessivas em vias metabólicas que garantem a manutenção da vida nos diversos organismos (Nelson & Cox, 2000). Elucidar as funções celulares como um complexo sistema de interações é a principal meta na metabolômica. Com o avanço das tecnologias experimentais, principalmente nas décadas de 60 e 70, foi possível clarificar o entendimento do papel das enzimas no metabolismo intermediário em diversos organismos, e assim, propor modelos metabólicos bastante elaborados (Hieter & Boguski, 1997).

As enzimas metabólicas são de grande importância, pois participam de reações essenciais à sobrevivência e manutenção de um organismo vivo, incluindo os processos de crescimento celular, reprodução, resposta ao meio ambiente, mecanismos de sobrevivência, manutenção da estrutura e integridade celular. Estas reações ocorrem dentro das células gerando alterações químicas e físicas, no que chamamos de metabolismo. A complexidade de uma célula está intimamente relacionada com a função de seus procedimentos subjacentes. Desta maneira, torna-se imprescindível a identificação e anotação correta dos componentes das vias metabólicas nos organismos (Selkov *et al*., 1998; Karp *et al*., 1999).

A genômica fornece informações quase completas sobre genes e produtos gênicos que participam do metabolismo celular para um número crescente de organismos (Schilling *et al*., 2000). A disponibilidade de genomas sequenciados para uma ampla variedade de organismos tem conduzido o desenvolvimento de modelos de vias metabólicas. Métodos de genômica comparativa permitiram a identificação e a anotação funcional dos genes que codificam para proteínas envolvidas em muitas vias metabólicas. Além disso, métodos computacionais foram desenvolvidos para auxiliar na determinação de modelos metabólicos de dados provenientes da estrutura primária dos genomas de muitos organismos. Um modelo metabólico visa capturar a estrutura de conectividade entre as moléculas que participam de um determinado metabolismo. Assim, uma rede metabólica é um modelo gráfico que descreve a presença de metabólitos, de reações catalisadas por enzimas, e as relações entre substrato e produto nestas reações (Pitkänen *et al*., 2010). O entendimento do metabolismo de diversos organismos está relacionado com uma série de implicações, como a engenharia metabólica, o estudo do metabolismo envolvido em doenças, a decifrar o potencial patogênico de parasitas, a identificação de alvos para desenvolvimento de novas drogas, entre outros (Sigurdsson *et al.*, 2009).

A reconstrução metabólica pode ser feita pela transferência funcional de modelos metabólicos propostos para organismos de espécies evolutivamente relacionadas (Francke *et al*., 2005). Centenas de vias metabólicas foram automaticamente reconstruídas utilizando diferentes métodos e bases de dados. Um exemplo é dado pelo banco de dados do KEGG (*Kyoto Encyclopedia of Genes and*

*Genomes*), onde redes metabólicas complexas são representadas por diagramas de reações enzimáticas, ligando as proteínas e outros produtos com os genes responsáveis por vários processos celulares, tal como o metabolismo. Uma parte do KEGG é suplementada por um conjunto de tabelas de grupos de ortólogos. Cada via metabólica de referência pode ser vista como uma rede de enzimas ou atividades enzimáticas. Uma vez que os genes codificantes para enzimas sejam identificados no genoma (normalmente por similaridade de sequência e/ou correlação posicional), e suas atividades enzimáticas sejam adequadamente atribuídas, as vias metabólicas específicas de um organismo podem ser reconstruídas computacionalmente, através da correlação entre os genes no genoma com os produtos gênicos em vias metabólicas de referência (determinadas com base em dados experimentais). Pressupondo uma boa conservação das vias metabólicas na maioria dos organismos (de bactérias a vertebrados), torna-se possível construir um mapa contendo uma via de referência e então gerar vias específicas para um determinado organismo através de uma abordagem computacional (Ma & Zeng, 2003; Kanehisa & Goto, 2000; Kanehisa *et. al*, 2004).

Os métodos automáticos atualmente utilizados na reconstrução metabólica geram apenas modelos que devem ser manualmente curados. O número de vias metabólicas reconstruídas com curadoria de alta qualidade permanece baixo, mesmo com a grande quantidade de genomas sequenciados (Seo & Lewin, 2009; Pitkänen *et al*., 2010). Além disso, existe uma fração importante de enzimas com funções específicas envolvidas no metabolismo secundário e que ainda não tiveram modelos metabólicos propostos (Kotera *et al.*, 2004). Portanto, o conhecimento sobre vias bioquímicas para uma análise global ainda é insuficiente.

#### 1.1.5.2. Classificação enzimática

Existem diversas formas de classificação enzimática. Normalmente, essa classificação é baseada na reação por elas catalisada. Dentro desta classificação, existem 6 tipos principais de reação, representados na tabela 3. Exceto por algumas enzimas, como pepsina e tripsina, a maioria é denominada pela adição do sufixo “ase”. A IUBMB, numa tentativa de padronizar a nomenclatura de enzimas,

recomenda que os nomes indiquem tanto o substrato quanto a reação catalisada (http://www.chem.qmul.ac.uk/iubmb/enzyme/history.html).

No final dos anos 50, tornou-se evidente que a nomenclatura adotada até então estava desatualizada devido ao crescimento do número de enzimas identificadas. Muitas enzimas eram conhecidas por diferentes nomes e o mesmo nome também era dado para enzimas distintas. Diversas tentativas de nomenclatura foram propostas por diferentes grupos. Assim, durante um Congresso de bioquímica realizado em Bruxelas, em 1955, foi criada uma Comissão Internacional de Enzimas, com consentimento da IUPAC (*International Union of Pure and Applied Chemistry*). O propósito desta comissão era criar um código de regras sistemáticas que serviria como um guia para a nomeação consistente de novas enzimas. A partir daí, adotou-se a classificação proposta por esta comissão, onde as enzimas são classificadas com números específicos de acordo com a reação que catalisam, mas que também contém informação de substrato e cofatores. Esses números classificatórios passaram a ser denominados de *EC numbers* (*Enzyme Comission Numbers*) e são baseados em dados experimentais publicados para enzimas individuais, avaliados pela *Joint Commission on Biochemical Nomenclature* (JCBN), da IUBMB (Kanehisa & Bork, 2003; Kotera *et al.,* 2004; http://www.chem.qmul.ac.uk/iubmb/enzyme/history.html).

**Tabela 3:** Classificação enzimática baseada na reação catalisada e de acordo com a IUBMB.

| Grupo de enzimas | Reação catalisada | Exemplos |
|---|---|---|
| 1. Oxidoredutase | Transferência de átomos ou moléculas de hidrogênio e oxigênio de um substrato para outro ou associadas à reações de oxidação | Dehidrogenases Oxidases |
| 2. Transferase | Transferência de um grupo específico (fosfato, metil, etc) de um substrato para outro | Transaminases Cinases |
| 3. Hidrolase | Hidrólise de um substrato | Enzimas digestivas |
| 4. Liase | Remoção não hidrolítica de um grupo ou adição de um grupo ao substrato | Aldolases |
| 5. Isomerase | Mudança na forma molecular de um substrato | Fumarase |
| 6. Ligase | Junção de duas moléculas pela formação de novas ligações | Sintetases |

A classificação por EC possui quatro níveis hierárquicos. O primeiro deles está relacionado com a classe da enzima, descrevendo a reação global que a enzima catalisa. Os níveis seguintes possuem significados diferentes dependendo da classe enzimática que pertencem. O quarto nível descreve a especificidade da reação enzimática, definindo a relação com o substrato específico para gerar o produto, além dos cofatores utilizados. O mecanismo pelo qual a reação é executada é consequência dos resíduos funcionais que formam o sítio ativo da proteína (Gherardini *et al.,* 2007). Existem algumas reações conhecidas que estão presentes em diversas vias metabólicas, mas que não possuem EC associado porque a atividade enzimática ainda não tem uma caracterização completa com os dados experimentais disponíveis. Na verdade, a maioria das enzimas que participam do metabolismo secundário não são susceptíveis de serem classificadas por um EC uma vez que cada etapa da reação provavelmente não será plenamente caracterizada de uma maneira tradicional (Barret *et al.*, 1992; Kotera *et al*., 2004). Vale lembrar que a classificação baseada nos ECs não é uma classificação molecular de enzimas *per se*. Em princípio, diferentes moléculas com diferentes sequências aminoacídicas e estruturas tridimensionais podem ser classificadas com o mesmo número de EC porque catalisam a mesma reação enzimática (Kotera *et al*., 2004).

Para minimizar os problemas existentes na classificação por EC, outro sistema de classificação tem sido proposto, o *RC number* (*Reaction Classification*), que atribui um número usando o alinhamento entre a estrutura química das enzimas (Hattori *et al.*, 2003). O sistema de classificação por EC se baseia em um acúmulo de conhecimento humano, entretanto o sistema por RC é derivado de comparações computacionais da estrutura química. Desta maneira, espera-se extrair características comuns em reações diferentes que não foram observadas anteriormente e com isso, melhorar o entendimento das relações dos mecanismos de reação entre as famílias de enzimas. Uma grande vantagem é a classificação de enzimas que são difíceis de caracterizar experimentalmente ou que participam de vias secundárias ou alternativas. Entretanto, a utilização desta classificação ainda não é usada pela IUBMB (Kotera *et al.,* 2004).

Um desenvolvimento relativamente recente de classificação e nomenclatura tem sido o *Gene Ontology* (GO) que oferece um vocabulário controlado e, mais importante, é uma ontologia legível por máquina, tornando-se muito utilizado em bioinformática (Ashburner *et al.,* 2000). Uma ontologia é uma maneira formal de representação do conhecimento em que os conceitos são descritos tanto pelo seu significado quanto pela relação com outras entidades (Bard & Rhee, 2004).

### 1.1.5.3. Origem e evolução de funções enzimáticas e vias metabólicas

> "Nada na Biologia faz sentido exceto à luz da Evolução" (Dobzhansky[6], 1973).

Muitos dos problemas enfrentados na elucidação da anotação funcional de proteínas se devem às limitações no conhecimento da evolução destas funções, mesmo com o avanço nas análises de suas sequências primárias e estruturais (Valencia, 2005). Um dos aspectos fascinantes da genômica moderna é a mudança radical trazida para a biologia evolutiva. Durante a evolução da vida, vários eventos evolutivos ocorreram para permitir a diversificação rápida de reações enzimaticamente catalisadas, resultando em um aumento no tamanho dos genomas. A comparação das sequências de todos os genes dos genomas de todos os organismos tornaria possível a reconstrução da história evolutiva de cada gene em sua totalidade (Gogarten & Olendzenski, 1999).

A evolução pode ser considerada como o processo de alterações na sequência nucleotídica que ocorrem durante a história das espécies diferenciando-as de seus ancestrais. A maioria dos evolucionistas acredita que toda a vida na Terra descende de um ancestral comum, habitualmente chamado de LUCA *(Last Universal Common Ancestor* - Último Ancestral Universal Comum*)*. Esta proposição baseia-se em características comuns, como o código genético universal, compartilhadas pelos organismos atuais. Uma das teorias aceitas para o surgimento da vida em nosso planeta inicia a partir de uma sopa primordial, contendo um mínimo de nutrientes necessários a manutenção da vida (Miller, 1953). Uma

[6] Notável geneticista e biólogo evolutivo ucraniano. Leal defensor da Evolução Darwiniana, publicou o famoso ensaio anti-criacionista *Nothing in Biology Makes Sense Except in the Light of Evolution*. No Brasil, teve grande importância no desenvolvimento da genética.

hipótese aceita é que os organismos primordiais evoluíram até o LUCA, entidade que representa a divergência do ponto de partida de todas as formas de organismos existentes atualmente (Woese, 1998).

A história evolutiva de cada gene pode ser compreendida através da análise comparativa entre as sequências de todos os genes em todos os genomas disponíveis. Assim, é possível recompor o cenário evolutivo com a ocorrência de diversos eventos que levaram à evolução destes genes, como: (i) descendência vertical com modificação, conhecida como especiação; (ii) duplicação gênica; (iii) perda de genes; (iv) transferência horizontal de genes; e (v) fusão, fissão e outros arranjos gênicos. Estes eventos estão relacionados com alguns conceitos evolutivos (Homologia e Analogia).

Numa visão morfológica, estruturas com uma mesma origem embrionária em diferentes organismos são consideradas homólogas, exercendo ou não a mesma função. Estas estruturas sugerem uma ancestralidade comum. Em contrapartida, estruturas análogas possuem semelhança morfológica, condizente com a execução de funções similares, mas com distinta origem embrionária. Um exemplo clássico de estruturas análogas são as asas dos insetos e das aves, com origens embrionárias distintas, mas adaptadas à execução da mesma função: o vôo. As estruturas análogas não refletem por si sós qualquer grau de parentesco. Elas fornecem indícios da adaptação de estruturas de diferentes organismos a uma mesma variável ecológica através de evolução convergente (Sattler, 1984; Fitch, 2000). A Figura 9 exemplifica as estruturas análogas e homólogas nas asas de morcegos, aves e insetos.

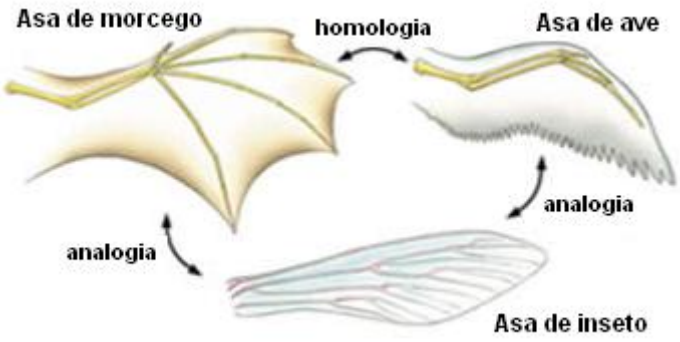



**Figura 9:** Representação de estruturas análogas e homólogas nas asas de morcegos, aves e insetos. Fonte: http://www.mundoeducacao.com.br/upload/conteudo_legenda/42732948a1e7ffa8f78a8696d01567d9.jpg

Em relação a genes e seus produtos, a homologia é definida como estruturas ou sequências que possuem um relacionamento de ancestralidade comum. Genes homólogos podem ser classificados em ortólogos e parálogos quando oriundos de transferência vertical de um único gene ancestral (através de eventos de especiação), e por duplicação gênica, respectivamente. Uma distinção clara entre ortólogos e parálogos é crítica para a construção de uma classificação evolutiva robusta de genes e para uma anotação funcional confiável de genomas recentemente sequenciados (Koonin, 2005). Na maioria das vezes, proteínas homólogas possuem estrutura tridimensional semelhante. Em contrapartida, existe o termo analogia, que se refere a uma mesma função protéica proveniente de origem evolutiva independente, ou seja, descendem com convergência a partir de genes ancestrais distintos. Essas proteínas normalmente não possuem similaridade estrutural (Fitch, 1970). Um esquema da árvore evolutiva da formação de genes homólogos (ortólogos e parálogos) é representado na Figura 10.

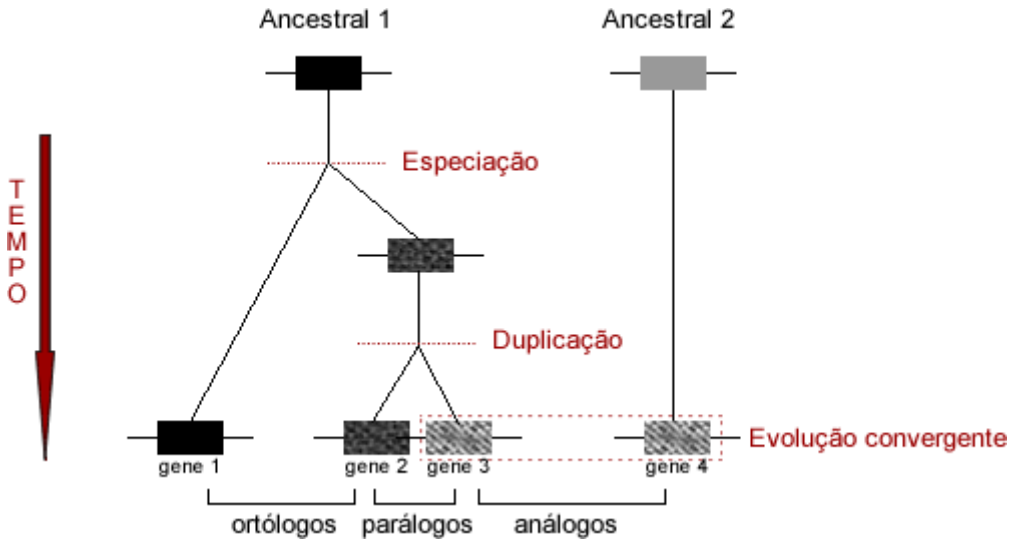



**Figura 10:** Árvore evolutiva mostrando a formação de genes homólogos (ortólogos e parálogos) e análogos. O gene do ancestral comum 1 sofre um evento de especiação originando genes ortólogos (genes 1 e 2), em espécies diferentes, assim como o evento de duplicação origina genes parálogos em uma mesma espécie (genes 2 e 3). Por evolução convergente, genes de ancestrais distintos (Ancestral 1 e Ancestral 2) podem codificar para proteínas com mesma função, denominadas análogas (genes 3 e 4).

A maioria dos estudos de evolução molecular está focada em genes e proteínas individuais. No entanto, a compreensão da evolução de redes moleculares requer uma perspectiva global do sistema, ligando as informações de evolução molecular de genes a um sistema de vias metabólicas celulares (Vitkup *et al*., 2006). A origem e a evolução das vias metabólicas básicas representaram um passo crucial na evolução molecular e celular. As primeiras células possuíam limitado número de enzimas, o que leva à proposição de que estas enzimas possuíam uma ampla especificidade de substrato, permitindo a catálise de várias reações químicas diferentes. Assim, a rede ancestral metabólica hipotética foi, provavelmente, composta por um número limitado de nós (enzimas) que eram altamente interligados, participando de distintos processos biológicos, diferente dos modelos de redes metabólicas atuais com estruturas extremamente complexas (Fani & Fondi, 2009).

A análise comparativa de genes e genomas de organismos pertencentes à Archaea, Bacteria e Eukarya revelou que, durante evolução, diferentes forças e mecanismos moleculares podem ter sido responsáveis pela expansão de genomas e habilidades metabólicas. Um dos mais importantes é a duplicação gênica, permitindo

a formação de novos genes a partir de genes pré-existentes. Outros mecanismos também podem ter influenciado na taxa de evolução metabólica, como a montagem modular de novas proteínas por eventos de fusão gênica e transferência horizontal, sendo que este último poderia permitir a transferência de toda ou parte de uma via metabólica. Esta idéia é suportada por experimentos de evolução dirigida que mostraram que novas especificidades de substrato podem aparecer em algumas semanas a partir de enzimas existentes por eventos de recombinação dentro de um gene (Hall & Zuzel, 1980; Ohta, 2000). Provavelmente, os genes ancestrais que deram origem às vias metabólicas atuais codificavam para proteínas com sequências simples, e que evolutivamente (por eventos de fusão gênica e duplicação de motivos internos) foram ganhando outros domínios até se tornarem mais complexas (Liò *et al.*, 2007).

Existem diversas hipóteses sobre a origem e evolução das vias metabólicas. Todas elas baseiam-se em eventos de duplicação gênica. A primeira tentativa de explicar em detalhe a origem das vias metabólicas foi feita por Horowitz, e é conhecida como a hipótese retrógrada (Horowitz , 1945). Ele sugeriu que as enzimas biossintéticas foram adquiridas por duplicação gênica na ordem inversa encontrada nas vias atuais. Se um composto fosse essencial para a sobrevivência de células primordiais, após o esgotamento na sopa primordial, uma pressão seletiva serviria para permitir a sobrevivência e reprodução de células que fossem capazes de realizar a transformação de um composto quimicamente relacionado. Assim, eventos de duplicação levariam a rotas cada vez mais complexas, até que uma via inteira fosse criada, começando com a síntese do produto final, depois uma via intermediária, e assim por diante até o precursor inicial. A descoberta de operons ajudou a reafirmar esse modelo, onde genes pertencentes ao mesmo operon e/ou para a mesma via metabólica deveriam formar uma família de genes parálogos (Horowitz, 1965). Esta hipótese possui muitas controvérsias, uma delas está no fato de que muitas vias anabólicas necessitam de muitos intermediários instáveis, além do desenvolvimento de rotas biossintéticas envolvendo reações diferentes.

Uma alternativa menos conhecida foi proposta por Granick, onde a biossíntese de alguns produtos finais pode ser explicada pela evolução para frente a partir de precursores relativamente simples que antecederam o aparecimento de

compostos mais complexos e, por conseguinte, as enzimas catalisando etapas anteriores de uma via metabólica são mais antigas do que as posteriores (Granick, 1957). Esta hipótese é talvez improvável, pois seria necessário que cada um dos intermediários fosse decorrente do desenvolvimento de múltiplos genes simultaneamente.

A hipótese “colcha de retalhos” (*Patchwork*) sugere que as vias metabólicas podem ter sido montadas através do recrutamento de enzimas primitivas que poderiam reagir com uma grande variedade de substratos quimicamente relacionados. Essas enzimas não-específicas podem ter permitido que células com pequenos genomas pudessem superar as suas capacidades limitadas de codificação (Ycas, 1974; Jensen, 1976). Esta hipótese também é consistente com a possibilidade de que uma via metabólica ancestral pode ter tido uma enzima primitiva catalisando duas ou mais reações semelhantes em substratos relacionados dentro da mesma via e cuja especificidade do substrato foi aperfeiçoada como resultado de eventos posteriores de duplicação. A hipótese de *patchwork* é suportada por várias linhas de evidência, como a ampla especificidade de substrato de algumas enzimas que podem catalisar uma série de reações químicas diferentes (Gerlt & Babbitt, 1998).

Outra hipótese mais recente foi postulada para a origem de algumas vias metabólicas, onde enzimas surgem por reações não enzimáticas seguidas pela aquisição da função enzimática. Esta hipótese é corroborada pelo conhecimento de que algumas reações químicas podem ocorrer de forma espontânea (Lazcano & Miller, 1996).

Com os dados atuais disponíveis, foi possível reconstruir as vias metabólicas para diversos organismos. Através da comparação entre estas vias, percebeu-se que algumas estavam ausentes ou incompletas em alguns organismos (Galperin & Koonin, 1998; Morett *et al.*, 2003). Em outros casos, algumas atividades enzimáticas essenciais haviam sido substituídas por outras atividades, com a mesma função enzimática, mas com pouca ou nenhuma similaridade em suas estruturas primárias e terciária (Galperin *et al.*, 1998; Gherardini *et al.*, 2007; Omelchenko *et al.*, 2010).

Tais formas alternativas são denominadas enzimas análogas e são resultado de acontecimentos evolutivos independentes.

A evolução das enzimas análogas se deu, provavelmente, por um mecanismo de recrutamento de enzimas existentes, as quais passaram a adquirir novas funções em virtude da mudança de especificidade ao substrato ou pela modificação de mecanismos catalíticos, como proposto na hipótese da colcha de retalhos. Esta hipótese é suportada pela presença das formas análogas de uma dada atividade enzimática em um número limitado de espécies (Galperin *et al.*, 1998). Algumas enzimas análogas executam funções relacionadas com a adaptação a novos ambientes e para estilos de vida diferentes, e podem estar associadas a diferentes linhagens filogenéticas e/ou possuir distintos mecanismos de catálise (Gherardini *et al.*, 2007; Omelchenko *et al.*, 2010).

Um estudo detalhado de enzimas análogas pode revelar novos mecanismos catalíticos, adicionar informação sobre a origem e evolução das vias bioquímicas e expor potenciais alvos para o desenvolvimento de novas drogas (Doolittle, 1994). Somente a comparação na estrutura primária não é capaz de provar que duas sequências estão evolutivamente relacionadas ou não. Uma origem comum só pode ser inferida a partir da conservação da estrutura da proteína (Murzin, 1996).

Estudos prévios já indicavam a existência de atividades enzimáticas decorrentes de eventos de origem independente em uma fração substancial, cerca de 25% (Galperin *et al.*, 1998; Hegyi & Gerstein, 1999; Morett *et al.*, 2003; George *et al.*, 2004; Gherardini *et al.*, 2007; Omelchenko *et al.*, 2010). No entanto, o conhecimento global sobre tais eventos ainda não tinha sido amplamente explorado.

## 1.2. Modelos

Para a realização desse trabalho, alguns modelos foram escolhidos para exemplificar a utilização de análogos. Entre eles, estão os tripanossomatídeos, *Trypanosoma cruzi*, *Trypanosoma brucei* e *Leishmania major.* Estas espécies pertencem à ordem Kinetoplastidae, cujos membros apresentam um ou dois flagelos, que tipicamente emergem de uma bolsa flagelar e apresentam uma única

mitocôndria que se ramifica por todo o corpo do protozoário. Uma característica importante desta ordem é a presença de uma grande quantidade de DNA organizado na forma de minicírculos e maxicírculos localizados numa estrutura intramitocondrial, chamada de cinetoplasto (DeSouza *et al*., 2000).

Estes organismos são parasitas obrigatórios, alternando entre hospedeiros vertebrados e invertebrados, sendo os agentes etiológicos de doenças humanas graves, como a doença do sono (*T. brucei*), a doença de Chagas (*T. cruzi*), e as leishmanioses (*Leishmania* sp.) (Opperdoes, 1994). A doença de Chagas afeta cerca de 8 milhões de pessoas na América Latina, enquanto que a doença do sono está presente principalmente no continente africano. As leishmanioses, entretanto, possuem uma distribuição mundial, atingindo principalmente as regiões tropicais e subtropicais. Tais doenças já foram caracterizadas há bastante tempo, mas ainda não existe uma terapia adequada para controlá-las, pois normalmente estão associadas com baixa eficácia, toxicidade dos medicamentos, dificuldades com a administração e alto custo (El-on, 2009; Radwanska, 2010; Rassi *et al*., 2010).

A análise dos genomas fornece uma vista global do potencial metabólico desses parasitas. Em termos gerais, *T. brucei* tem o repertório metabólico bem restrito, seguido pelo *T. cruzi*. O *T. brucei* não possui uma etapa intracelular no ciclo de vida, tendo, portanto maior acesso a nutrientes no plasma, o que se reflete em seu metabolismo. Os tripanossomatídeos são organismos que desenvolveram soluções alternativas a fim de administrar suas vias metabólicas, as quais já vêm sendo descritas na literatura ao longo dos anos. Uma ampla revisão do metabolismo desses organismos, juntamente com análises visando a busca de alvos moleculares para desenvolvimento de drogas contra estes parasitas é apresentada na seção de resultados.

Além destes parasitas, dados oriundos do projeto genoma de *Homo sapiens* também foram utilizados neste estudo para comparações com os genomas dos parasitas, a fim de buscar proteínas diferenciadas que pudessem ser utilizadas também como alvos para o desenvolvimento de drogas.

# *Objetivos*

**Geral**:

Identificação, anotação e reconstrução metabólica de enzimas isofuncionais não-homólogas em genomas procarióticos e eucarióticos e sua utilização na busca de novos alvos terapêuticos.

**Específicos:**

1. Implementação de uma metodologia para busca e identificação de enzimas análogas e reconstrução de vias metabólicas;

2. Anotação de genomas, através da utilização de ontologias para a descrição de genes e seus produtos, utilizando a metodologia de busca por análogos;

3. Reconstrução metabólica de organismos, mapeando a existência de enzimas análogas e específicas (quando comparadas entre parasita e hospedeiro) como potenciais novos alvos terapêuticos;

4. Obtenção de listas organismo-específicas de enzimas candidatas a se tornarem novos alvos terapêuticos, utilizando uma abordagem baseada na presença de atividades enzimáticas compartilhadas pelo parasita e pelo hospedeiro, mas com estruturas tridimensionais diferenciadas devido a origem evolutiva distinta entre estas atividades;

5. Obtenção de listas organismo-específicas de enzimas candidatas a se tornarem novos alvos terapêuticos, utilizando uma abordagem baseada na ausência de atividades enzimáticas no hospedeiro, mas que se encontram presentes no parasita;

6. Obtenção de modelos 3D para os potenciais análogos funcionais identificados utilizando a ferramenta MHOLline (http://www.mholline.lncc.br/);

7. Construção de um banco de dados com as análises obtidas neste estudo.

# *Resultados*

# Capítulo I – Implementação da metodologia

A primeira etapa deste trabalho constituiu na construção e implementação da metodologia utilizada para identificar, anotar e comparar enzimas análogas, combinando algoritmos de bioinformática. O AnEnPi (Analogous Enzymes Pipeline) foi desenvolvido para buscar genes análogos utilizando como entrada, seqüências de proteínas armazenadas no banco de dados do KEGG. Acredita-se que proteínas análogas são o resultado de eventos evolutivos independentes, ao invés de terem se originado de uma proteína ancestral comum. Essas proteínas, como mencionado anteriormente, possuem a mesma função, mas com diferentes estruturas primárias, secundárias e terciárias.

O AnEnPi é capaz de realizar as seguintes tarefas: i) agrupamento de sequências protéicas gerando grupos usando um arquivo multifasta como entrada para busca de enzimas análogas (onde as enzimas/proteínas em diferentes grupos de uma mesma atividade enzimática são consideradas análogas); ii) Análise e interpretação de todos os grupos gerados a partir de todas as enzimas provenientes do KEGG; iii) detecção de analogia intragenômica nos dados provenientes do KEGG (existência de formas distintas de uma mesma atividade enzimática em um mesmo organismo); iv) detecção de analogia intergenômica entre dois organismos pertencentes ao banco de dados do KEGG (formas distintas de uma mesma atividade enzimática em dois organismos distintos); v) anotação da proteína (utilizando BLAST ou HMMER); vi) geração de mapas, usando as ferramentas fornecidas pelo KEGG.

A interface é estruturada de acordo com as tarefas executadas pelo AnEnPi, consequentemente, o usuário pode analisar os resultados após cada etapa, ou refazer apenas uma etapa em particular. Os resultados obtidos com o AnEnPi podem ser recebidos pelo usuário por e-mail ou através da interface web.

A ferramenta AnEnPi está disponível para acesso público em http://bioinfo.pdtis.fiocruz.br/AnEnPi/. Esta encontra-se descrita no artigo “AnEnPi: identification and annotation of analogous enzymes.”, publicado em *BMC Bioinformatics* 2008; 9:544.

BMC Bioinformatics

Software



# AnEnPi: identification and annotation of analogous enzymes

Thomas D Otto*[1,2], Ana Carolina R Guimarães[1], Wim M Degrave[1] and Antonio B de Miranda[1]

Address: [1]Laboratory for Functional Genomics and Bioinformatics, Oswaldo Cruz Institute, Fiocruz, Rio de Janeiro, Brazil and [2]Ataulpho de Paiva Foundation, Rio de Janeiro, Brazil

Email: Thomas D Otto* - otto@fiocruz.br; Ana Carolina R Guimarães - carolg@fiocruz.br; Wim M Degrave - wdegrave@fiocruz.br; Antonio B de Miranda - antonio@fiocruz.br

* Corresponding author

Published: 17 December 2008

*BMC Bioinformatics* 2008, **9**:544 doi:10.1186/1471-2105-9-544

Received: 12 May 2008
Accepted: 17 December 2008

This article is available from: http://www.biomedcentral.com/1471-2105/9/544



## Abstract

**Background:** Enzymes are responsible for the catalysis of the biochemical reactions in metabolic pathways. Analogous enzymes are able to catalyze the same reactions, but they present no significant sequence similarity at the primary level, and possibly different tertiary structures as well. They are thought to have arisen as the result of independent evolutionary events. A detailed study of analogous enzymes may reveal new catalytic mechanisms, add information about the origin and evolution of biochemical pathways and disclose potential targets for drug development.

**Results:** In this work, we have constructed and implemented a new approach, AnEnPi (the Analogous Enzyme Pipeline), using a combination of bioinformatics tools like BLAST, HMMer, and in-house scripts, to assist in the identification, annotation, comparison and study of analogous and homologous enzymes. The algorithm for the detection of analogy is based i) on the construction of groups of homologous enzymes and ii) on the identification of cases where a given enzymatic activity is performed by two or more proteins without significant similarity between their primary structures. We applied this approach to a dataset obtained from KEGG Comprising all annotated enzymes, which resulted in the identification of 986 EC classes where putative analogy was detected (40.5% of all EC classes). AnEnPi is of considerable value in the construction of initial datasets that can be further curated, particularly in gene and genome annotation, in studies involving molecular evolution and metabolism and in the identification of new potential drug targets.

**Conclusion:** AnEnPi is an efficient tool for detection and annotation of analogous enzymes and other enzymes in whole genomes. It is available for academic use at: http://bioinfo.pdtis.fiocruz.br/AnEnPi/

## Background

Enzymes catalyze biochemical reactions and are classified according to the recommendations of the Nomenclature Committee of the International Union of Biochemistry [1]. Each enzymatic activity has a recommended name and an Enzyme Commission (EC) number assigned, depending on the reaction that it catalyzes [2]. For a better understanding of the metabolism of a given species it is of utmost importance to locate, identify and annotate the genes encoding such enzymatic activities. Most

approaches to perform these tasks are based on sequence similarity searches, using computational tools like BLAST [3] or Hidden Markov Models (HMMer [4]) and curated databases.

However, comparisons between computational reconstructions of metabolic pathways from different organisms revealed the existence of gaps [5]. An organism can truly lack a part of a pathway, use an alternative one, or the function is present but unannotated for different reasons (for instance, genome assembly problems). Another explanation is that some of these apparent gaps might involve alternative enzymes, also known as functional analogs [6]. Such enzymes are generally believed to be the result of independent evolutionary events [7]. Some properties of analogous enzymes include its association with different phylogenetic origins, possession of distinct catalytic mechanisms and also different foldings [8]. Automated annotation approaches, normally used for preliminary gene identification and characterization, usually employ methods based on sequence similarity criteria. These may not be able to detect analogs, as these enzymes exhibit virtually no significant sequence similarity between their primary structures [9]. In some cases it is possible to use other types of data, such as the genomic context or the experimental detection of a particular enzymatic activity, to identify the genes coding for the missing activities [10].

However, most often such genes are not characterized as analogous in the accompanying annotation, for Example in public databases such as KEGG [11].

Previous work performed by other groups suggest that the fraction of enzymatic activities where multiple events of independent origin have occurred may be substantial, in the order of 25% [9]. However, to our knowledge a global survey of these events, which also has the potential to shed light on the evolution of biochemical pathways and genome organization, has not been done.

Analogous enzymes may also constitute a huge and largely untapped resource for the identification of drug targets. Strategies to find candidate genes as potential targets for drug development usually focus on parasite-specific genes and even complete biochemical pathways [12], or for structural differences between homologues. Unfortunately, due to technical limitations, the number of available 3D structures represents only a fraction of all proteins identified so far, limiting direct structural comparisons and inducing researchers to rely on the comparison of annotation data. Since analogous enzymes, which may have Substantially different foldings – a desired prerequisite for drug development – are not annotated as such, they may be overlooked as possible candidates for drug development.

To help in the process of identification and annotation of analogous enzymes, we implemented a web based Tool named AnEnPi. It analyses and compares genomic datasets for analogous enzymes, by clustering the primary structures of enzymes with the same described activity and using a Blastp similarity raw score of 120 as cut-off [7]. This resulted in a list of clusters that reflect substantial structural differences between enzymes with the same activity but with possibly different evolutionary origins.

## Methods

AnEnPi was programmed in Perl using the CGI-interface. All clusters as well as their HMMer-models are available for download on the web page.

An overview of AnEnPi is shown in Figure 1. For clustering we used the similarity score with a cut-off Value 120 of BLASTp pair wise comparisons between all proteins included in a specified dataset, based on the experimental work of Galperin [7]. In the work described here, groups are composed of proteins sharing the same enzymatic activity (EC classes). Within a group, protein sequences are clustered. Enzymes within a given cluster are considered homologous, while enzymes in different clusters (of the same group/function) are considered analogous. These clusters are stored in a flat file database, which can be used to annotate or re-annotate a set of proteins. To improve visualization, metabolic maps can be generated automatically.

### *Dataset*

We have applied AnEnPi to cluster a dataset composed of 311 reference metabolic pathways and 1,871,732 protein sequences of 36 eukaryotes, 398 eubacteria and 31 archaeabacteria obtained from the KEGG database [11,13]. In total, 326,013 sequences had a corresponding EC number assigned describing their enzymatic activity, belonging to 2,433 different EC classes. This result forms the main dataset of clusters used by AnEnPi.

### *Clustering*

The clustering algorithm was implemented similarly to the method proposed by Galperin [7]. First, sequences with less than 100 amino acids were excluded from the dataset. For each enzymatic activity, an all-against-all BLASTp [3] (using a maximum e-value of 0.01 and standard parameters) was executed and results were transformed in a graph where each node represents an enzyme [14]. Two nodes are connected by an unweighted and undirected edge if they belong to the same EC class and have a similarity score higher or equal to 120 (an e-value close to $e^{-6}$) [7]. This parameter (and others) can be mod-

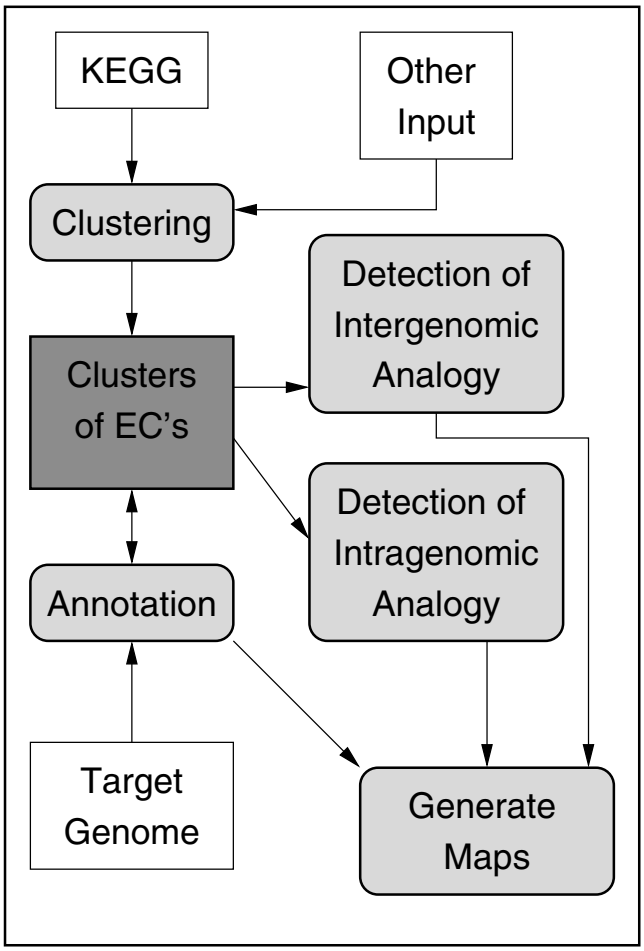



**Figure 1**
**Work flow of AnEnPi**. Databases are represented as rectangles. Darker gray rectangles represent the five datasets of clusters. Light gray rectangles are the modular functions of AnEnPi, described in the text.

ified by the user. All sequences connected in the graph were assigned to the same cluster and are considered homologous. Sequences not connected by a path in the graph are considered analogous. Therefore, the number of disconnected sub graphs would, in principle, represent the number of times that the enzymatic activity in question is thought to have appeared during evolution within the current dataset. As a representation of the graph, an adjacency-matrix [15] was implemented (Figure 2). Each cluster is finally stored in a flat file database.

#### *Filters for the datasets*

Dataset a, the less conservative, is composed by all clusters formed after the initial clustering step. This dataset was further refined, using more stringent criteria. Filters were applied in four successive steps: Firstly, all clusters with only one sequence (singlets) were excluded (dataset B). Secondly, all enzymatic activities not defined up to the fourth level of the EC classification were also excluded (dataset C). Thirdly, all clusters of a determined function with proteins annotated as subunits of this function and Belonging to the same species were joined (dataset D). Finally, all clusters displaying putative intragenomic analogy (here defined as the identification of analogy between two enzymes in the same genome) were also joined (dataset E).

#### *Metabolic map reconstruction*

Each result can be visualized as a metabolic map by using an external resource (a KEGG tool – [16]). Further, EC classes with potential cases of analogy or without representative sequences are highlighted. Color codes are used to discriminate the significance of the results, as well as the presence of analogy.

#### *Detection of analogy*

In this work, we define a potential case of analogy if the sequences from a given enzymatic activity present in the genome of a single organism or between two organisms are placed in different clusters after grouping (intragenomic and inter-genomic analogy, respectively). Orthologs or recently duplicated paralogs would be placed in the same cluster. Therefore AnEnPi compares, within a single species or between two species, the presence of a given function in each cluster, for all species currently represented in KEGG database. In the metabolic reconstruction step, the presence of a given enzymatic activity in a genome, the presence of analogy and the degree of significance of similarity searches can be highlighted. The result is an interactive list (in HTML or text format) with links to the EC classes and the metabolic maps.

#### *Annotation*

For the purpose of annotation and identification, the user can perform similarity searches either by BLASTp or HMMSearch. In the first case, the database is composed of all proteins present in the clusters; in the second case, the database is composed of the probabilistic models constructed for each cluster, if the cluster has more than one element. For the construction of the latter, a multiple alignment was executed with ClustalW [17] and then transformed into a HMMer model with the functionalities available in the HMMer package. This type of annotation is based on the quality of previously annotated databases.

Therefore, we have introduced filters allowing the construction of different datasets, minimizing the number of wrongly identified cases of analogy and wrongly attributed functions. To our knowledge, AnEnPi is the only tool that provides annotation functionalities with emphasis on analogous enzymes.

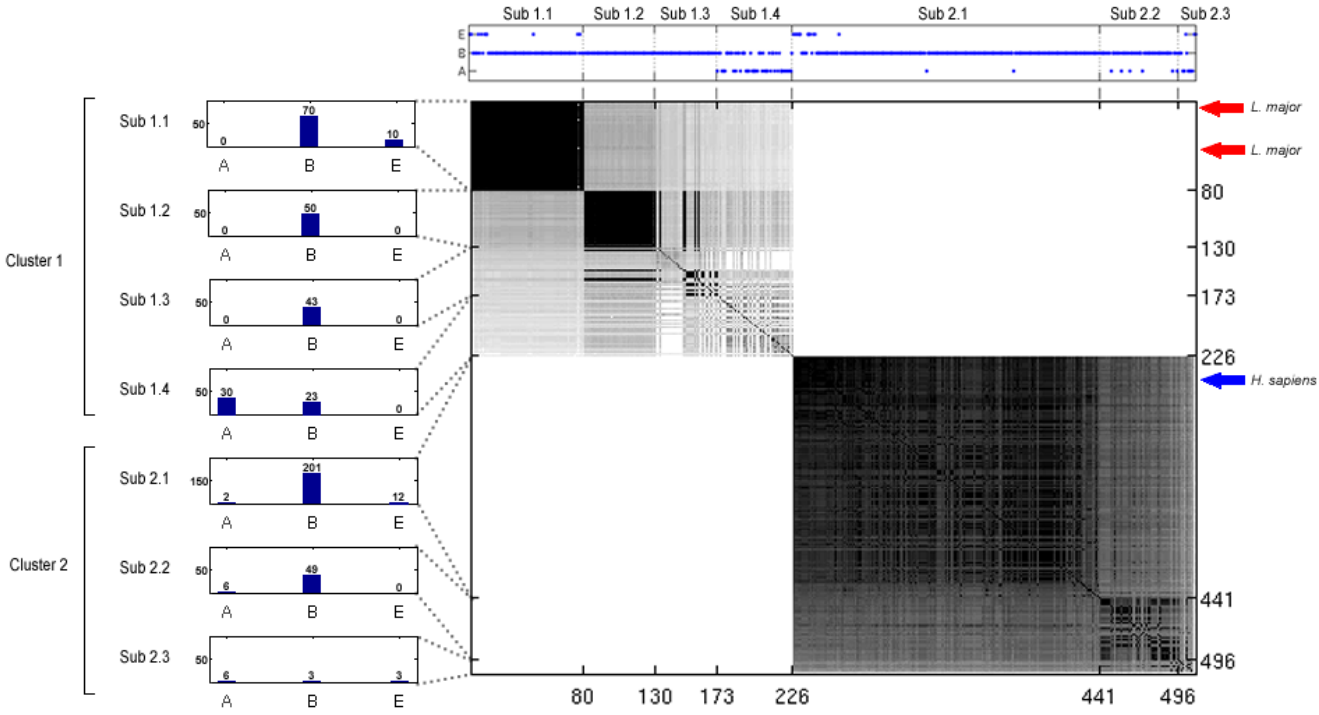



**Figure 2**

**Similarity matrix**. Similarity matrix (central figure) of EC 4.2.1.2 (fumarate hydratase). Each point of the 508 × 508 matrix represents the blastp similarity score of two enzymes. All scores above 1500 are reset to 1500. Higher similarity scores yield darker points, (white represents a score below 120). Sequences were sorted by the similarity score, using the longest enzymes as reference to the other enzymes. The three arrows on the right site indicate the positions of enzymes of *L. major* (red arrows) and *H. sapiens* (blue arrow). Histograms on the left display the distribution of organisms represented in each cluster, for the three kingdoms: archaebacteria (A), bacteria (B) and eukaryotes (E). In each of the two main (analogous) clusters, sub-clusters can be observed. The graph at the top of the matrix displays the kingdom of the organism for every enzyme in the matrix.

### *Front-end*

All components described so far (Figure 1) were included in a user-friendly web-based interface named AnEnPi. All main functionalities are independent processes and may be used in different contexts, for instance in the identification of analogous enzymes, in sequence annotation, clustering or metabolic reconstruction. Also, sequences entered by the user can be clustered and converted into an annotation database for similarity searches. Results are displayed in a web page.

## Results

We implemented a web based interface http://bioinfo.pdtis.fiocruz.br/AnEnPi/ called AnEnPi (Analogous Enzyme Pipeline) that can be used for the annotation and visualization of metabolic pathways, and for the detection of events of analogy. As an example, we applied AnEnPi to identify possible cases of analogy between enzymes present in the genome of *L. major* [18] and the human genome, using a clusters From the KEGG database to form our reference datasets.

### *Work flow, user interface, sequences and organisms*

Clustering of the dataset obtained from KEGG produced 6,701 clusters, with 986 enzymatic activities (from the 2,433 represented in the KEGG dataset) having more than one cluster (approximately 40.5%) (Table 1). 2,199 sequences formed singlets, while 328 EC classes had more than three clusters. Table 1 shows the number of enzymatic activities with putative analogy before and after the four steps of data filtering. After the third filtering step (dataset D), the number of functions with more than 5 clusters drops to 46. Still, even after the application of these 3 filters, 19% of the enzymatic activities contain putative analogous sequences.

### *Comparison with literature data*

To validate our results, we have searched the literature for known cases of analogy, predicted or confirmed through diverse approaches, such as computational and/or experimental methods. We compared our results with those of [7], where 108 cases of analogy were described. Only for three enzymatic activities the number of clusters produced

**Table 1: Refinement of The initial Dataset (A) through the application of successive filters.**

| Datasets | # Clusters | | | | Max. Clusters | % Analogous |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | > 3 | | |
| A | 1447 | 459 | 199 | 328 | 131 | 40.5 |
| B | 1600 | 345 | 113 | 180 | 78 | 26.2 |
| C | 1560 | 316 | 91 | 97 | 46 | 20.7 |
| D | 1619 | 302 | 73 | 70 | 23 | 19.4 |
| E | 1897 | 142 | 23 | 1 | 5 | 8.1 |

Table 1: A, dataset obtained after clustering; B, dataset obtained after the exclusion of singlets (clusters With only one sequence); C, dataset obtained after the exclusion of EC's which are not defined up to the Fourth level (incomplete EC' s); D, dataset obtained after the joining of clusters where some sequences Were annotated as 'subunits'; E, dataset obtained after the joining of clusters with putative intragenomic Analogy. Max. Cluster, the maximum number of clusters found for one specific enzymatic activity; % analogous, fraction of enzymatic activities where analogy was detected. # Clusters: number of functions with, respectively, 1, 2, 3 or more than 3 clusters.

by AnEnPi was smaller, in each case due to dataset differences. For all other cases we found at least the same number of clusters, demonstrating the consistency of the results. Table 2 illustrates some cases of analogy found in the literature. All functions listed also display potential cases of analogous sequences in our results, provided that the enzymatic activity in question is included in KEGG.

### *Adjacency matrices and kingdom line*

Figure 2 shows an adjacency-matrix for EC 4.2.1.2.(fumarate hydratase), a representation of an all-against-all BLASTp of all proteins belonging to this enzymatic activity. Each point in the matrix indicates the similarity score between two enzymes. Two main clusters can be seen, where sequences from one cluster have no detectable similarity with sequences from the other cluster. Some subclusters can be seen inside each main cluster, representing groups of more similar sequences, particularly inside cluster 1.

The histograms on the left and the 'kingdom line' above the matrix show the distribution of the organisms represented in the matrix in terms of kingdoms. Sub-cluster 1.1 has sequences derived from eukaryotes and eubacteria, while sub-clusters 1.2 and 1.3 from eubacteria only. The remaining sequences of cluster 1 do not form a well-defined sub-cluster, but archaeabacterial sequences are present only in this structure.

However, further identification of subclusters will be studied in another work, as a refinement of the present methodology. As an example of a potential case of analogy, for this EC function the location of the corresponding enzymes of *L. major* (two genes) and *H. sapiens* are displayed.

### *Analogy identification*

When applying AnEnPi to find analogy between *H. sapiens* and *L. major*, thirty-five potential cases of analogy were found using dataset D (Table 1). In twelve cases (EC 1.1.1.2, EC 1.3.1.34, EC 1.3.3.4, EC 2.3.1.48, EC 2.7.1.2, EC 2.7.4.2, EC 3.5.1.14, EC 3.6.1.23, EC 4.2.1.1, EC 4.2.1.2 (Figure 2), EC 5.3.1.6 and EC 5.3.3.2), intergenomic analogy was found. The smallest cluster found comprised 8 individual protein sequences. The great majority (80 well as intra-genomic analogy. For example, in EC 2.1.1.17 (phosphatidylethanolamine N-methyltransferase), enzymes of *L. major* and *H. sapiens* share the cluster #4, but enzymes of *L. major* were also found in cluster #3. Therefore, *L. major* enzymes from cluster #3 are analogous to *H. sapiens* sequences present in cluster #4, and the sequences from both organisms present in cluster #4 are homologous. This function can also be used to display any other differences between two species in the web frontend (Figure 3).

### *Intragenomic analogy*

As described above, EC 2.1.1.17 is also a case of intragenomic analogy in *L. major*. With AnEnPi we detected a total

**Table 2: Examples of analogy Found in The literature and the Methods used.**

| EC | Enzyme | Organism | Ref. | Method |
|---|---|---|---|---|
| 1.1.1.42 | Isocitrate dehydrogenase | *Escherichia coli/Azotobacter vinelandii* | [7] | a |
| 2.7.1.4 | Fructokinase | *Homo sapiens/Streptococcus mutans* | [7] | a |
| 3.2.1.86 | 6-phospho-beta-glucosidase | *E. coli/H. sapiens* | [7] | a |
| 3.4.21.72 | Immunoglobulin A (IgA) proteases | *Streptococcus sanguis/Neisseria gonorhoeae* | [27] | a |
| 2.1.1.- | N-methyltransferase 1 | *Schizosaccharomyces pombe/Chlamydia pneumoniae/Archaeoglobus fulgidus* | [9] | a, b |
| 2.7.7.- | Adenylyltransferase | *Bacillus subtilis/Saccharomyces cerevisiae* | [9] | a, b |
| 3.1.3.11 | Fructose-1,6-bisphosphatase | *Prochlorococcus marinus/E. coli* | [28] | c |
| 2.3.1.- | Enoyl thioester reductase | *E. coli*/Yeast/Rat | [29] | d |
| 1.13.11.2 | Catechol 2,3-dioxygenase | *Pseudomonas sp.*/TOL plasmid (pww0) | [30] | e |
| 2.7.2.3 | Glyceric acid 3-phosphate kinase | *Pisum sativum* | [31] | f |
| 5.3.1.1 | Triose phosphate isomerase | *Pisum sativum* | [31] | f |

Table 2: a – computational, b – genetic complementation, c – genome sequencing, d – stereochemical assay, e – biochemical assay and f – isoelectric focusing

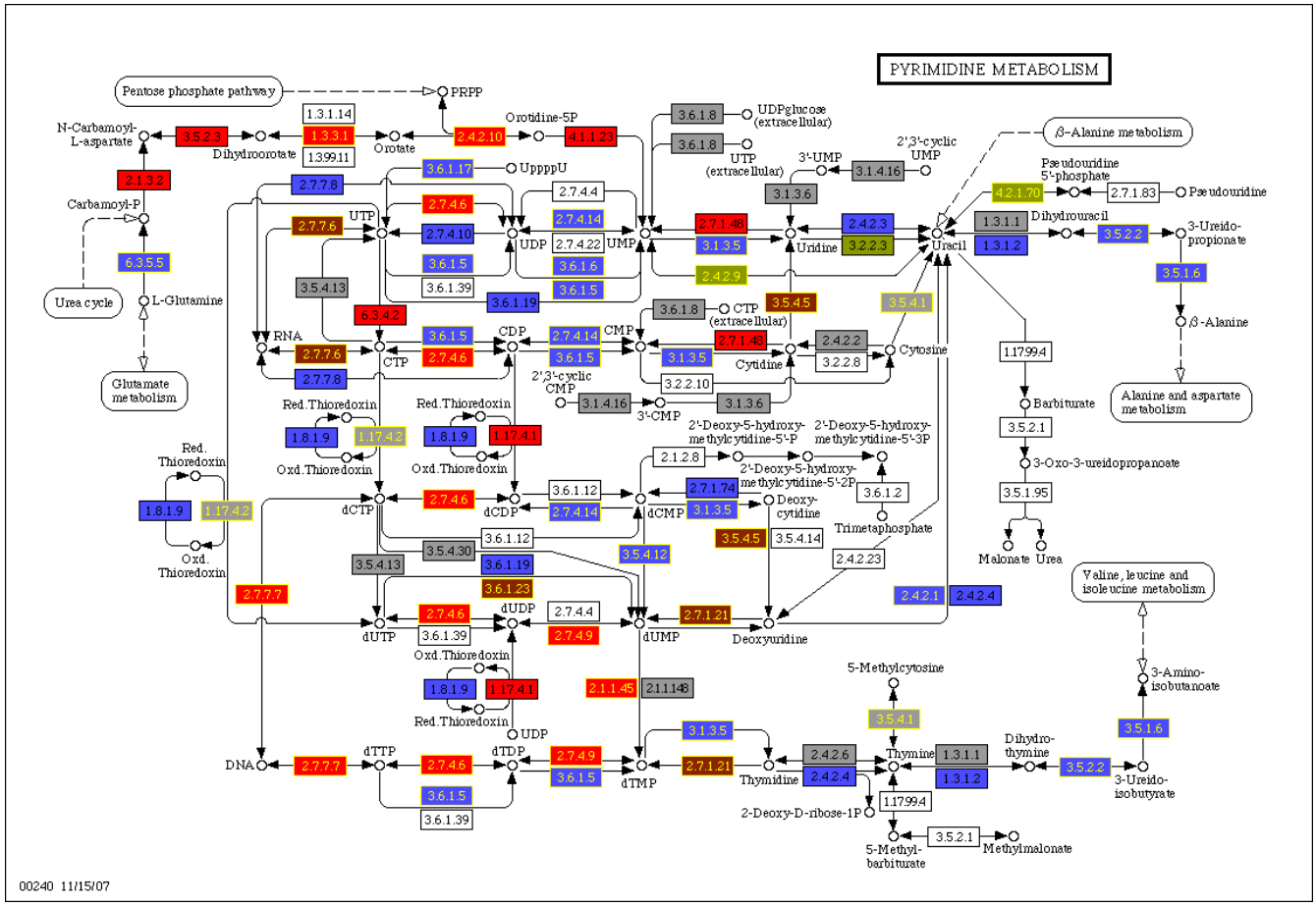



**Figure 3**

**Comparative analysis between H. sapiens and L. major**. Brown: analogy between two genes with the same function; red: function present in both species; blue: Function present only in *H. sapiens*; green: function present only in *L. major*; gray: function not found in both species and white: no representative enzymes in KEGG.

of 12 cases of intragenomic analogy in *L. major* using dataset D. Application of AnEnPi to Datasets A, B and C returned 34, 34 and 23 cases, respectively. No intragenomic analogy is detected when using dataset e, because all clusters (from a particular enzymatic activity) with sequences from the same species are joined. These cases of intragenomic analogy were not related to a particular metabolic pathway.

One example of intragenomic analogy can be seen in the fructose and mannose metabolism (KEGG map 00051), where two unrelated sequences of phosphomannomutase (EC 5.4.2.8) were found in the *L. major* genome. A bibliographic search revealed almost no data besides the identification of these twelve enzymatic Activities in *L. major*, neither were we able to find systematic studies of intragenomic analogy in general.

## Discussion

We described in this work AnEnPi, a tool that can be used for the annotation and detection of analogous enzymes [19], improving the understanding of the biochemical pathways of the species under analysis. It offers functionalities for clustering, annotation, or pairwise comparisons between different species, intended for the identification and improvement of annotation of putative analogous enzymes.

Other tools like KAAS [20] or KOBAS [21] also perform whole genome annotation of enzymes, but AnEnPi is unique in the detection, comparison and visualization of events of analogy. In the advanced parameter settings, each threshold can be modified, such as for clustering, which should be of use for a large group of users.

The identification of structurally unrelated enzymes sharing the same enzymatic activity may reveal new catalytic mechanisms, lead to studies on the origin and evolution of biochemical systems and pathways, and also provide new candidates for drug design and development [22]. AnEnPi is an implementation of a methodology designed to help in the identification and annotation of putative

events of independent origin of enzymatic activities through the clustering of their primary sequences [23]. AnEnPi also provides information for a more detailed reconstruction of metabolic pathways, including the significance of similarity scores and the presence/absence of alternative forms of a given enzymatic activity.

It is not a simple task to determine if two different proteins are derived from the same ancestor. Two homologous proteins may lack major sequence similarity and yet share a common origin, for example after many years of evolution [8]. The cut-off used in this work, a similarity score of 120, is based on the observation that there is a lack of similarity between the tertiary structures of proteins below this value [7].

Still, it is possible that two enzymes assigned by AnEnPi as analogous are in fact derived from the same ancestor but have diverged up to a point where their primary sequences no longer share recognizable similarity. Molecular modeling techniques, together with appropriate evolutionary methods, could be used to ascertain that the tri-dimensional structures and sequences of the enzymes assigned as analogous are indeed different, suggesting their independent origin.

To overcome some of these difficulties, methods to deduce functional information from a certain gene in the absence of sequence data have been recently proposed [24]. Needless to say, most approaches rely on high-quality annotation. As a matter of fact, problems with the data structure of some databases may create undesirable biases in our analyses. For instance, we have observed that annotation for a specific enzymatic activity for one particular subunit of a multimeric enzyme is commonly 'inherited' by all other subunits composing that enzyme. If these subunits are encoded by unrelated genes and do not have the same function, false cases of analogy will be computed. False cases of analogy will also appear for enzymatic activities that are dependent on the simultaneous presence of more than one type of subunit to form the catalytic site. In other words, if a hetero-multimeric enzyme is composed of subunits with different origins, AnEnPi may interpret the lack of similarity between said subunits as another case of analogy.

Although we have so far no automatic way to further refine our dataset, the distribution pattern of species over the clusters of a given enzymatic activity may indicate the presence of false positives and therefore serve as a criterion for their identification: the presence of representatives (proteins) from the same organism in several clusters would mean that that organism has several unrelated enzymes able to fulfill the same metabolic step. While this may be real, it is likely that a substantial part of these events are indeed annotation artifacts. In general, our results were congruent with the available literature on the subject (Table 2).

It is thus important to discriminate between i) two (or more) subunits of a given heteromultimeric enzyme encoded by unrelated genes and ii) two (or more) enzymes actually sharing the same function, also encoded by unrelated genes. Table 1 displays the results found when applying these criteria to improve the dataset.

Most likely, the majority of the clusters with only one representative sequence are possibly cases derived from wrong annotations or cases of very divergent sequences, which are not included in other clusters due to the cut-off used. As an example, analysis of *T. brucei* data produced 14 singlets. The annotation of the metabolic pathways in this organism was done manually [25], and results entered in the KEGG database.

The user should choose the best dataset for his purposes. To be conservative, we have employed in most of our analyses the dataset D, minimizing the number of false positives (and consequently probably loosing other real cases of analogy). Using the dataset E, though very restrictive and probably an underestimation, we obtained a set of analogies with a higher probability of being true cases, without possible errors due to multimeric proteins; in fact, even after applying all these criteria for data filtering, still 8.6% of all enzyme classes have potential cases of analogy. No doubt, a better handling of inconsistencies generated during the annotation of multimeric enzymes would improve the identification and provide a better estimation of the frequency and distribution of the cases of intragenomic analogy.

The ability to identify potential cases of analogy between genes from two different species (Figure 3), as well as differences in assigned functions, can be used to indicate the possibility of alternative pathways or disclose candidates for drug development. One example analyzed in more detail is fumarate hydratase (EC 4.2.1.2.) from *H. sapiens* and *L. major*, whose sequences were assigned to distinct clusters. AnEnPi can help by producing a list of shared enzymatic activities between the two organisms without detectable similarity at their primary level, reflecting substantial differences between their folding patterns. Also, the overall pattern of similarity scores shown in Figure 2 suggests that fumarate hydratase is evolving in distinct ways, depending on the group of organisms in question. More detailed studies are underway to investigate these points.

Currently, we are developing a database with all putative analogy events stored in a comprehensive way, linked to

information from drug databases. AnEnPi is also being updated, with the inclusion of information from hundreds of new organisms.

## Conclusion

AnEnPi is a versatile tool designed to assist the user in the identification, clustering and annotation of analogous enzymes. Its modular structure allows its utilization in other contexts. Addition of color codes to represent biological attributes allows for a better visualization of metabolic pathways, with more meaningful biological information, facilitating the interpretation of the results.

## 1 Availability and requirements

AnEnPi is freely accessible at http://bioinfo.pdtis.fiocruz.br/AnEnPi/.

• Project name: AnEnPi – Analogous Enzyme Pipeline (Webserver)

• Project home page: http://bioinfo.pdtis.fiocruz.br/AnEnPi/

• Operating system: Linux

• Programming language: Perl and HTML

• Licence: AnEnPi is accessible under a GPL license

## 2 Abbreviations

EC: enzyme commission; AnEnPi: Analogous Enzyme Pipeline

## Authors' contributions

The tool was developed and implemented by TO. He and AG designed the web interface and carried out the experiments. All authors analyzed together the data, discussed the results, wrote the manuscript and approved the final version.

## Acknowledgements

We thank CNPq, CAPES, FAPERJ, the World Community Grid, the Bioinformatics Platform PDTIS/Fiocruz [26] and Fiocruz/PAPES for financial support. We are very grateful to Marcos Catanho, Nicolas Carels, Fernando a Álvarez Valin and Hector Romero for valuable discussions and advice.

## References

1. **Nomenclature Committee of the International Union of Biochemistry** [http://www.chem.qmul.ac.uk/iubmb/]
2. Kotera M, Okuno Y, Hattori M, Goto S, Kanehisa M: **Computational Assignment of the EC Numbers for Genomic-Scale Analysis of Enzymatic Reactions.** *Journal of the American Chemical Society* 2004, **126(50):**16487-16498.
3. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25:**3389-3402.
4. Durbin R, Eddy SR, Krogh A, Mitcison G: *Biological sequence analysis: probabilistic models of protein and nucleic acids* Cambridge Uiversity Press; 1998.
5. Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO: **Towards quantitative biology: Integration of biological information to elucidate disease pathways and to guide drug discovery.** *Biotechnology Annual Review* 2003, **11:**1-68.
6. Fitch WM: **Distinguishing homologous from analogous proteins.** *Systematic Zoology* 1970, **19(2):**99-113.
7. Galperin MY, Walker DR, Koonin EV: **Analogous enzymes: independent inventions in enzyme evolution.** *Genome Res* 1998, **8:**779-790.
8. Galperin MY, Koonin EV: **Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes.** *Genetica* 1999, **106(1–2):**159-170.
9. Morett E, Korbel JO, Rajan E, Saab-Rincon G, Olvera L, Olvera M, Schmidt S, Snel B, Bork P: **Systematic discovery of analogous enzymes in thiamin biosynthesis.** *Nat Biotechnol* 2003, **21(7):**790-795.
10. Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, B? U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UCM, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth TJ, Churcher C, Clark LN, Corton CH, Cronin A, Aavies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, Macleod A, Mooney PJ, Moule S, Martin DMA, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream MA, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CMR, Tait A, Tivey AR, Aken SV, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM: **The Genome of the African Trypanosome *Trypanosoma brucei*.** *Science* 2005, **309(5733):**416-422.
11. Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG.** *Nucleic Acids Res* 2006, **34:**D354-D357.
12. Karp PD: **Integrated pathway/genome databases and their role in drug discovery.** *Trends in Biotechnology* 1999, **17(7):**275-281.
13. **Kegg ftp download** [ftp://ftp.genome.ad.jp/pub/kegg/]
14. Aho AV, Hopcroft JE, Ullman JD: *The Design and Analysis of Computer Algorithms. Addison-Wesley* 1974.
15. Cormen TT, Leiserson CE, Rivest RL: *Introduction to Algorithms* 2nd edition. Mit Press; 2001.
16. **Color Objects in KEGG Pathways** [http://www.genome.jp/kegg/tool/color_pathway.html]
17. Thompson JD, Higgins DG, Gibson TJ: **Clustal W: improving the sensitivity of progressive Multiple sequence alignment through sequence weighting, position-specific gap penalties and Weight matrix choice.** *Nucleic Acids Res* 1994, **22(22):**4673-4680.
18. Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E, Rajandream MA, Adlem E, Aert R, Anupama A, Apostolou Z, Attipoe P, Bason N, Bauser C, Beck A, Beverley SM, Bianchettin G, Borzym K, Bothe G, Bruschi CV, Collins M, Cadag E, Ciarloni L, Clayton C, Coulson RMR, Cronin A, Cruz AK, Davies RM, Gaudenzi JD, Dobson DE, Duesterhoeft A, Fazelina G, Fosker N, Frasch AC, Fraser A, Fuchs M, Gabel C, Goble A, Boffeau A, Harris D, Hertz-Fowler C, Hilbert H, Horn D, Huang Y, Klages S, Knights A, Kube M, Larke N, Litvin L, Lord A, Louie T, Marra M, Masuy D, Matthews K, Michaeli S, Mottram JC, Muller-Auer S, Munden H, Nelson S, Norbertczak H, Oliver K, O'neil S, Pentony M, Pohl TM, Price C, Purnelle B, Quail MA, Rabbinowitsch E, Reinhardt R, Rieger M, Rinta J, Robben J, Robertson L, Ruiz JC, Rutter S, Saunders D, Schr M, Schein J, Schwartz DC, Seeger K, Seyler A, Sharp S, Shin H, Sivam D, Squares R, Squares S, Tosato V, Vogt C, Volckaert G, Wambutt R, Warren T, Wedler H, Woodward J, Zhou S, Zimmermann W, Smith DF, Blackwell JM, Stuart KD, Barrell B, Myler PJ: **The genome of the Kinetoplastid parasite, *Leishmania major*.** *Science* 2005, **309(5733):**436-342.

19. Karp PD: **Call for an enzyme genomics initiative.** *Genome Biol* 2004, **5(8):**401.
20. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: **KAAS: an automatic genome annotation and Pathway reconstruction server.** *Nucleic Acids Res* 2007, **35:**W182-W185.
21. Wu J, Mao X, Cai T, Luo J, Wei L: **KOBAS server: a web-based platform for automated annotation And pathway identification.** *Nucleic Acids Res* 2006, **1:**720-724.
22. Fischer HP: **Towards quantitative biology: Integration of biological information to elucidate Disease pathways and to guide drug discovery.** *Biotechnology Annual Review* 2005, **11:**1-68.
23. Stein L: **Genome annotation: from sequence to biology.** *Nat Rev Genet* 2001, **2(7):**493-503.
24. Morett E, Korbel JO, Rajan E, Saab-Rincon G, Olvera L, Olvera M, Schmidt S, Snel B, Bork P: **Discovery of Analogous Enzymes in Thiamin Biosynthesis by Anticorrelation.** *Genome Informatics* 2002, **13:**365-366.
25. El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, Renauld H, Worthey EA, Hertz-Fowler C, Ghedin E, Peacock C, Bartholomeu DC, Haas BJ, Tran AN, Wortman JR, Alsmark UCM, Angiuoli S, Anupama A, Badger J, Bringaud F, Cadag E, Carlton JM, Cerqueira GC, Creasy T, Delcher AL, Djikeng A, Embley TM, Hauser C, Ivens AC, Kummerfeld SK, Pereira-Leal JB, Nilsson D, Peterson J, Salzberg SL, Shallom J, Silva JC, Sundaram J, Westenberger S, White O, Melville SE, Donelson JE, Andersson B, Stuart KD, Hall N: **Comparative Genomics of Trypanosomatid Parasitic Protozoa.** *Science* 2005, **309(5733):**404-409.
26. Otto TD, Catanho M, Degrave W, de Miranda AB: **The PDTIS bioinformatics platform: from sequence to function.** *RECIIs* 2007, **1(2):**286-294.
27. Gilbert JV, Plaut AG, Wright A: **Analysis of the immunoglobulin A protease gene of Streptococcus sanguis.** *Infect Immun* 1991, **59(1):**7-17.
28. Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, Barbe V, Duprat S, Galperin MY, Koonin EV, Gall FL, Makarova KS, Ostrowski M, Oztas S, Robert C, Rogozin IB, Scanlan DJ, de Marsac NT, Weissenbach J, Wincker P, Wolf YI, Hess WR: **Genome sequence of the *cyanobacterium Prochlorococcus marinus* ss120, a nearly minimal oxyphototrophic genome.** *PNAS* 2003, **100(17):**7-17.
29. Reynolds KA, Holland KA: **The mechanistic and evolutionary basis of stereospecificity for hydrogen Transfers in enzyme-catalysed processes.** *Chemical Society Reviews* 1997, **26:**337-343.
30. Kukor JJ, Olsen RH: **Catechol 2,3-dioxygenases functional in oxygen-limited (hypoxic) environments.** *Appl Environ Microbiol* 1996, **62(5):**583-585.
31. Anderson LE, Pacold I: **Chloroplast and Cytoplasmic Enzymes: Three Distinct Isoenzymes Associated with the Reductive Pentose Phosphate Cycle.** *Plant Physiology* 1970, **45:**583-585.

# Capítulo II – Reconstrução metabólica

A reconstrução das vias metabólicas representada neste capítulo é exemplificada pelo metabolismo de aminoácidos em *Trypanosoma cruzi*. Este parasita é o agente etiológico da doença de Chagas, afeta a população na maior parte da América Central e do Sul, e constitui um significativo problema de saúde e sócio-econômico. O parasita tem um metabolismo em grande parte baseado no consumo de aminoácidos, principalmente nas formas epimastigotas, que participam de uma diversidade de vias metabólicas, gerando muitos compostos cruciais para a sobrevivência do parasita. O estudo de suas enzimas tem potencial para revelar novos alvos terapêuticos e fomentar o desenvolvimento de novas drogas. Utilizando a ferramenta AnEnPi, foi possível reconstruir computacionalmente (*in silico*) as vias do metabolismo de aminoácidos de *T. cruzi* com o objetivo de vincular a informação genômica com a informação funcional.

Estes dados fazem parte do artigo "*In silico* reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi*.", publicado na revista *Genetics and Molecular Research* (GMR) 2008; 7(3):872.

Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

# *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

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

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.

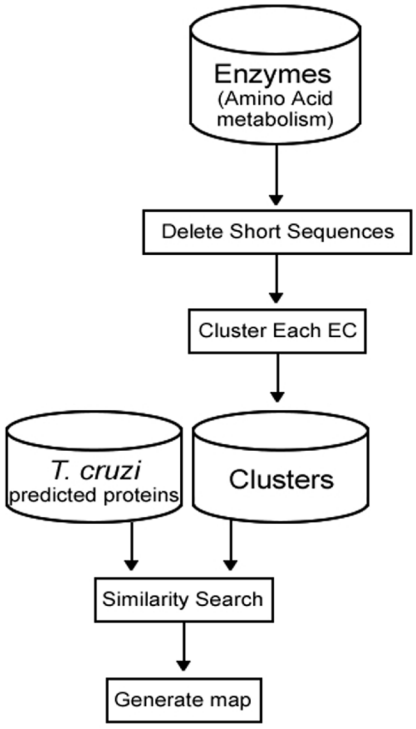



**Figure 1.** Workflow of methodology. EC = Enzyme Commission number.

## *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

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.

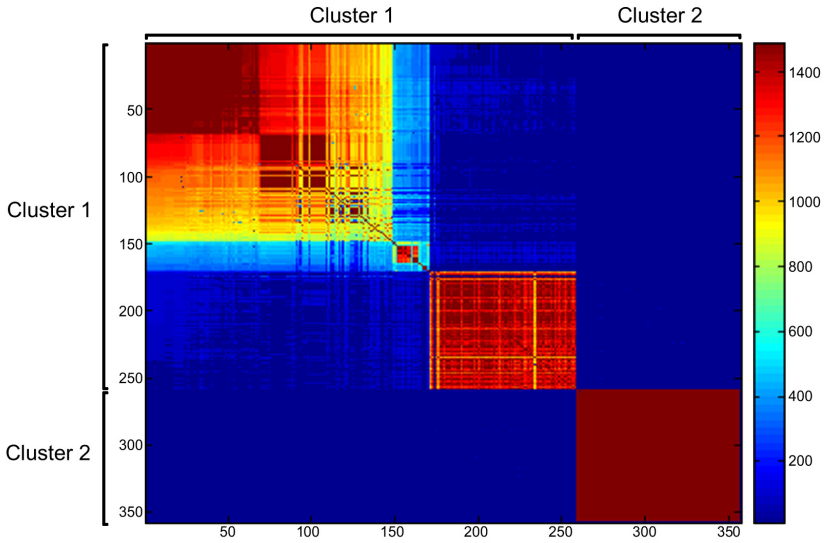



**Figure 2.** A graph depicting the similarity scores of all pairwise comparisons between the sequences belonging to EC:1.1.1.42 (isocitrate dehydrogenase). The order of the sequences along the axis is arbitrary. The scale of the similarity score is presented as a vertical bar on the right of the figure.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

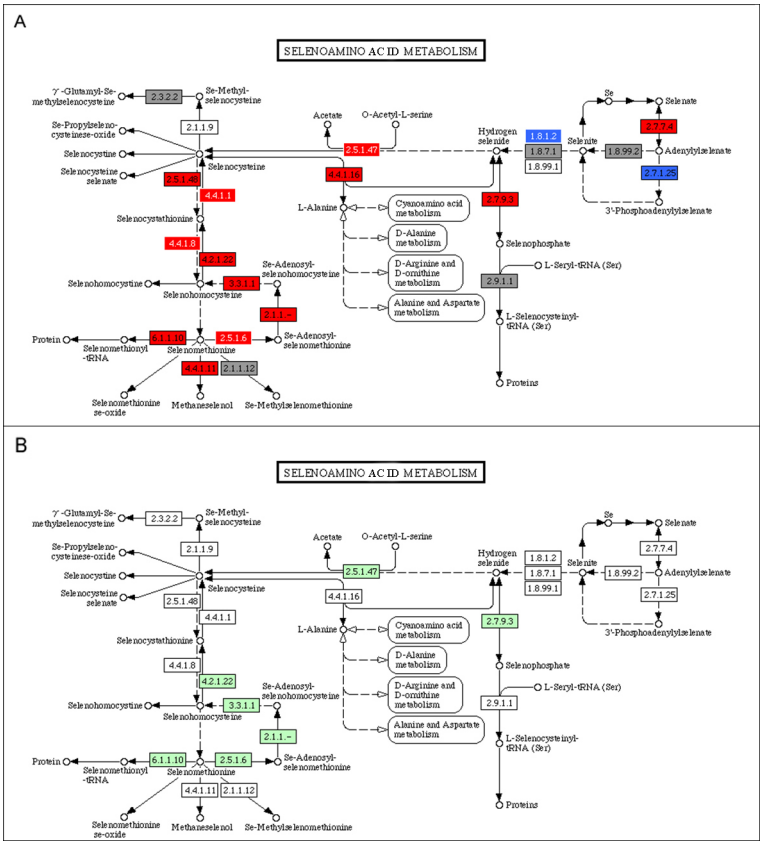



**Figure 3.** Maps of selenoamino acid metabolism showing the enzymatic functions of *Trypanosoma cruzi* identified by AnEnPi (A) and annotated for *T. cruzi* by KEGG (B). **A.** The background colors of the boxes specifying EC classes represent attributes of the enzymatic activities, respectively: white - 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.

# Capítulo III – Nova abordagem para busca de novos alvos moleculares

Este capítulo engloba uma revisão do metabolismo de *Trypanosoma cruzi* e a discussão de uma nova abordagem para a busca de potenciais alvos moleculares para o desenvolvimento de drogas contra este parasita.

A escolha de *T. cruzi* se deu porque as opções de fármacos disponíveis atualmente para o tratamento da doença de Chagas na fase crônica não são satisfatórias. Além disso, a utilização de dados dos genomas do parasita e de humanos pode possibilitar a identificação de novos alvos para medicamentos e o desenvolvimento de tratamentos adequados para as fases aguda e crônica da doença de Chagas. Usando um estudo comparativo *in silico* entre enzimas dos dois organismos através da anotação e busca de enzimas análogas e específicas, foi possível prever um número considerável de funções enzimáticas em *T. cruzi* que podem vir a serem usadas como alvos moleculares, não afetando o hospedeiro humano.

A revisão, juntamente com os dados analisados, estão no artigo “A new approach for potential drug target discovery through *in silico* metabolic pathway analysis using *Trypanosoma cruzi* genome information.”, publicado nas *Memórias do Instituto Oswaldo Cruz* 2009; 104(8):1100.

# A new approach for potential drug target discovery through *in silico* metabolic pathway analysis using *Trypanosoma cruzi* genome information

**Marcelo Alves-Ferreira[1], Ana Carolina Ramos Guimarães[1], Priscila Vanessa da Silva Zabala Capriles[2], Laurent E Dardenne[2], Wim M Degrave[1]/+**

[1]Laboratório de Genômica Funcional e Bioinformática, Instituto Oswaldo Cruz-Fiocruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil [2]Grupo de Modelagem Molecular de Sistemas Biológicos, Laboratório Nacional de Computação Científica, Ministério da Ciência e Tecnologia, Petrópolis, RJ, Brasil

*The current drug options for the treatment of chronic Chagas disease have not been sufficient and high hopes have been placed on the use of genomic data from the human parasite* Trypanosoma cruzi *to identify new drug targets and develop appropriate treatments for both acute and chronic Chagas disease. However, the lack of a complete assembly of the genomic sequence and the presence of many predicted proteins with unknown or unsure functions has hampered our complete view of the parasite's metabolic pathways. Moreover, pinpointing new drug targets has proven to be more complex than anticipated and has revealed large holes in our understanding of metabolic pathways and their integrated regulation, not only for this parasite, but for many other similar pathogens. Using an* in silico *comparative study on pathway annotation and searching for analogous and specific enzymes, we have been able to predict a considerable number of additional enzymatic functions in* T. cruzi. *Here we focus on the energetic pathways, such as glycolysis, the pentose phosphate shunt, the Krebs cycle and lipid metabolism. We point out many enzymes that are analogous to those of the human host, which could be potential new therapeutic targets.*

Key words: *Trypanosoma cruzi* - metabolism - metabolic pathways - drug target - analogous enzyme

Chagas disease, which is caused by an infection with the kinetoplastid protozoan parasite *Trypanosoma cruzi*, remains a serious public health problem in most Latin American countries despite successes in controlling transmission through vector control and blood donor screening in the affected regions. It is estimated that about 12 million people are chronically infected with Chagas disease (Dias 2006). In addition, due to migration, Chagas disease has become a threat in several additional countries (Schmunis 2007).

The acute phase, which appears shortly after infection, is often hard to notice and after a silent period of a several years or decades a chronic phase develops in about one-third of the infected individuals. Due to this clinical evolution, Chagas disease is often considered a "silent killer," impairing early specific diagnosis and treatment (Tarleton 2007, Bilate & Cunha-Neto 2008). The main clinical manifestations of Chagas disease are mostly cardiac, but also digestive alterations and the pathogenesis is attributed to host immune system disturbances associated with a very low and often undetectable parasite presence (Rocha et al. 2007, Marin-Neto et al. 2008). Until now, there has been no immediate prospect for the development of a vaccine and current clinical therapy based on heterocyclic nitro compounds (nifurtimox and benznidazol) is quite unsatisfactory for chronic patients, thus calling attention to the search for new therapeutic approaches (Coura & Castro 2002, Dias 2006).

The different developmental stages and changing biochemical interactions between the parasite and the vertebrate host during the life cycle of *T. cruzi* render the development of new drugs more difficult (de Souza 2002).

The availability of genomic (El-Sayed et al. 2005) and proteomic data (Atwood et al. 2005) from the parasite has set high hopes for the identification of new drug targets and a large amount of data is currently accessible (http://tcruzidb.org/tcruzidb/, http://www.genedb.org/, http://tritrypdb.org/tritrypdb/ and http://eupathdb.org/eupathdb/). However, difficulties with whole genome assembly due to the highly repetitive nature of *T. cruzi* sequence content, together with the heterozygosity of the analysed strain and the presence of many predicted proteins with unknown or unsure functions has hampered a full view of the metabolic pathways in this parasite. A lack of insight into the regulation and interplay of those pathways, the limited tools for functional genetic analysis to prove the essentiality of certain enzymes and little knowledge on functional differences with corresponding human enzymes are causes for the slow development of new drug candidates. Moreover, drug development for neglected diseases is not widely met with the necessary attention and determination that the many patients and people at risk deserve, which has been largely discussed on different occasions.

Financial support: MS-DECIT (41.0544/2006-0, 410401/2006-4), CNPq (PQ 304771/2008-2), FAPERJ to WMD, CNPq (486019/2007-1) to MAF, CAPES, Fiocruz PAPES, Fiocruz PDTIS Bioinformatics Platform, PIDC, WorldCommunityGrid
+ Corresponding author: wdegrave@fiocruz.br
Received 10 June 2009
Accepted 28 October 2009

In eukaryotes, the metabolism can be organised in pathways that are well conserved between different phyla; however, a considerable amount of variability can be noted in enzymatic characteristics, pathways, cellular compartments and regulatory mechanisms when species are compared. Protozoa, like the Kinetoplastida, present a number of features that are distinct from those seen in other organisms, including several organelles that are absent in mammalian cells (de Souza 2005, de Souza et al. 2008). One such organelle, the acidocalcisome, was discovered more recently in trypanosomatids and in the Apicomplexa (Docampo et al. 2005). The organelle has an important role in the regulation of calcium, phosphor and other elements though the presentation of several families of transporters on the membrane. The glycosome, which is a special type of peroxysome, is typical of trypanosomatids and contains seven enzymes of the glycolytic pathway; enzymes involved in carbohydrate metabolism, lipid metabolism (beta oxidation of fatty acids and their elongation, synthesis of lipid ethers and some steps of isoprenoid synthesis) and metabolism of purines and pyrimidines; and is an important generator of redox potential in the form of NADPH (Michels et al. 2000, Moyersoen et al. 2004). The mitochondria, vital organelles for the respiratory process and ATP production, participate in several metabolic cycles and are related to the process of cell death. Recent studies have indicated that the mitochondrial membrane of trypanosomatids has a peculiar lipid composition. The specific replication and editing processes, associated with the already well studied unusual DNA structure of this organelle, represent potential targets for new drugs.

The trypanosomatids have important and fundamental differences in their processes to generate energy, although some pathways are common (Tielens & Van Hellemond 1998). The parasitic forms of trypanosomatids also have metabolic differences during their life cycle, such as the Krebs cycle in *Trypanosoma brucei*, which is nearly inactive in bloodstream forms found in the vertebrate host, but is active in the procyclic form in the invertebrate host (Hannaert et al. 2003).

Even so, some of the glycosome enzymes are involved in amino acid metabolism, others in fatty acid synthesis and two enzymes participate in gluconeogenesis (van Weelden et al. 2005). All these particularities constitute potential new drug targets.

Over the last 100 years, different methods and drugs have been used against *T. cruzi* infection (Coura & Castro 2002). Since the 60s, nitrofurans, such as nifurtimox (Lampit™) and later imidazole compounds, such as benznidazole (Rochagan™), have been used due to their oxidising action, through the formation of reactive radicals and their ability to block enzymes and lipids. Newer chemicals were developed based on the nitrofuran and nitroimidazole backbone in an attempt to augment activity while diminishing the many negative side effects of the drugs. These drugs, which include ketoconazole, fluconazole, itraconazole and posaconazole, showed higher efficiency against the parasite and were also shown to inhibit enzymes in steroid biosynthesis (Urbina & Docampo 2003, Liñares et al. 2006). These triazole (antifungal) agents are known to block the synthesis of ergosterol through the inhibition of the enzyme cytochrome P450 lanosterol 14a-demethylase, leading to the accumulation of methylated sterol precursors. Inhibitors of nucleotide metabolism, such as the xantine oxidase inhibitor alopurinol, were also evaluated (Maya et al. 2007). The trypanosomatid specific enzyme tripanothione reductase has been extensively analysed as a potential drug target and both chemical compounds (Maya et al. 2007, Perez-Pineiro et al. 2009) and natural products (Galarreta et al. 2008) were found to inhibit the enzyme. Cruzipain, a cysteine proteinase involved in cellular processes and host-parasite interactions, was pointed out as another potential target and in vivo assays with different chemical inhibitors were quite promising (Doyle et al. 2007, Brak et al. 2008, Fricker et al. 2008). Several other enzymes have also been targeted, such as gliceraldehyde-3-phosphate dehydrogenase (Freitas et al. 2009), pteridine reductase (Cavazzuti et al. 2008), dihydrofolate reductase/thymidilate synthase, farnesyl-pyrophosphate synthase and DNA topoisomerase (Paulino et al. 2005), while a large number of natural products are also being screened. These studies seem quite promising and a decisive breakthrough is eagerly awaited from the scientific community.

A potential new drug target should be a molecule that is essential to the parasite, for which high affinity inhibitors can be identified and which have little or no effect on the eventual human homologue. Specificity, low toxicity, bioavailability, stability and ease of administration are additional requirements for potential drug candidates. Strategies usually aim to identify parasite specific enzymes, complete biochemical pathways (Karp et al. 1999) or structural differences between human and parasite homologues. The latter approach is hampered by the limited availability of 3D structures. With this in mind, we developed a methodology to identify cases of enzyme analogy, where two different proteins share the same function or activity while lacking similarity between their primary sequences (Galperin et al. 1998). In these cases, the enzymes are thought to have originated through independent evolution. Therefore, we designed a computational workflow and tool, AnEnPi (Otto et al. 2008), to detect them from genomic data. The absence of (detectable) primary sequence similarity allows for the expectation of a substantially different folding of the protein. Besides the importance of identifying analogous enzymes in parasites and humans as potential drug targets, the identification of structurally unrelated enzymes sharing the same enzymatic activity may reveal new catalytic mechanisms and shed light on the origin and evolution of biochemical systems and pathways. Previously, we analysed the amino acid metabolic pathways of *T. cruzi* (Guimarães et al. 2008).

Here we further explore the metabolic pathways of *T. cruzi* to assess the extent of analogy in the parasite, while suggesting important new candidates for drug targeting.

## MATERIALS AND METHODS

AnEnPi analysis of the metabolic pathways of *T. cruzi* was essentially done as previously described (Otto et al. 2008). Briefly, the AnEnPi tool was used to cluster

a dataset composed of 311 reference metabolic pathways and 1,871,732 protein sequences of 36 eukaryotes, 398 eubacteria and 31 archaeabacteria obtained from the KEGG database (Kanehisa et al. 2006). In total, 326,013 sequences had a corresponding EC number assigned to them describing their enzymatic activity, which belong to 2,433 different EC classes. This result forms the main dataset for clustering used by AnEnPi. Sequences with less than 100 amino acids were excluded and an all-against-all BLASTp was performed for each enzymatic activity. Enzymes were considered to be homologous if they belonged to the same EC class and had a similarity score greater than or equal to 120 (corresponding to an e-value close to $10^{-6}$) (Galperin et al. 1998). Enzymes in the same EC class with very low or undetectable similarity clustered in separate groups and were considered analogous. The resulting dataset was further refined by excluding all clusters with only one sequence (singlets) or where the enzymatic activity was not defined up to the fourth level of EC classification. Furthermore, all clusters of a determined function, with proteins annotated as subunits of this function and belonging to the same species, were joined. The results can be visualised as a metabolic map through the use of an external (KEGG) resource. For the purpose of this analysis, a potential case of analogy was identified if the sequences from a given enzymatic activity, which were present in the genome of a single organism or between two organisms, were placed in different clusters after grouping (intra-genomic and inter-genomic analogy, respectively). Using this method, orthologs and paralogs would end up in the same cluster. AnEnPi maps the presence of a given function, within a single species or between two species, in one or more clusters. For metabolic analysis, we used the *T. cruzi* genomic TSK-TSC v5.0 dataset of 19,607 predicted proteins (http://tcruzidb.org). Tentative function was annotated using BLASTp with e-value cut-offs of $10^{-20}$, $10^{-40}$ or $10^{-80}$ and subsequently assigned to clusters of enzyme activities derived from the KEGG database as explained above. In the current study, analogous enzymes in the parasite were identified in comparison with enzymes in the corresponding human metabolic pathways as exemplified in Figs 1-3.

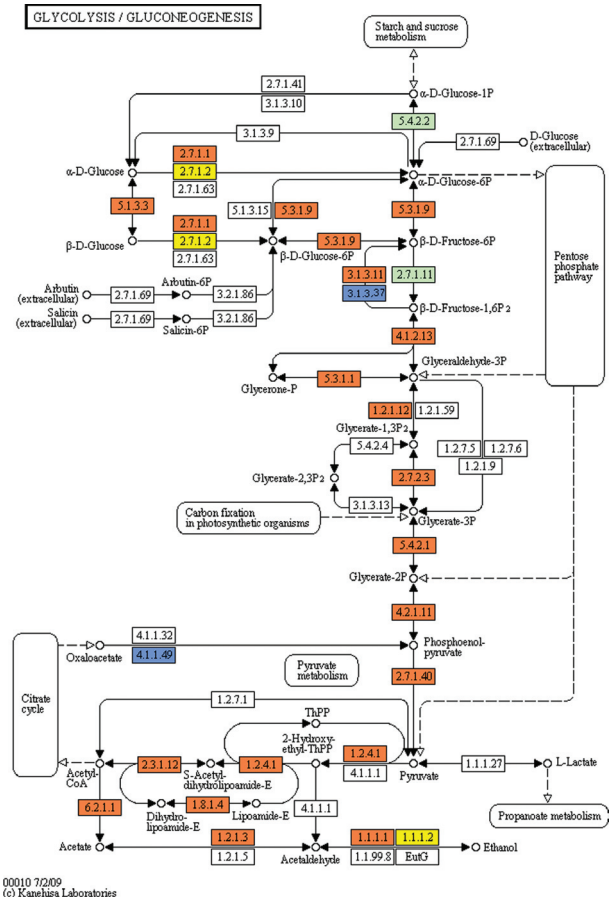



Fig. 1: reconstruction of the glycolysis and gluconeogenesis pathways in *T. cruzi*. Colored boxes symbolize enzymatic activities annotated by AnEnPi methodology. The map was generated using the reference map by KEGG (http://www.genome.jp/kegg/pathway.html). Green boxes: enzymatic activities without putative cases of analogy (for all species in KEGG database); orange boxes: enzymatic activities with putative cases of analogy (also for all species in KEGG database); yellow boxes: enzymatic activities with probable cases of analogy between *Trypanosoma cruzi* and *Homo sapiens*; blue boxes: enzymatic activities specific for *T. cruzi* and absent in *H. sapiens*.

## RESULTS AND DISCUSSION

Many aspects of *T. cruzi* biochemistry have been investigated and described over the last few decades and several remarkable features can be noted in the major metabolic pathways of the parasite, which may constitute valuable focal points for further research and drug development. Using AnEnPi, we performed a detailed analysis of the enzymes predicted from the *T. cruzi* CL-Brener genome sequence (as available from http://tcruzidb.org). Using this method we were able to tentatively annotate a considerable amount of additional enzymatic functions for the parasite, compared to the KEGG prediction, as shown in Table I. Using e-value cut-offs of $10^{-20}$, $10^{-40}$ and $10^{-80}$ we found 165, 111 and 65 additional protein sequences, respectively, with probable enzymatic functions, compared to the 396 functions assigned for *T. cruzi* using KEGG. For the subset of activities present in the parasite, 74% (295) had cases of analogy in KEGG displaying more than one cluster, as opposed to 40.5% when analysing all pathways from all species in the database (Otto et al. 2008). For protein functional annotation using an e-value of $10^{-20}$, AnEnPi pointed out 127 additional cases of analogy. When comparing the corresponding human enzymes, AnEnPi could find 33 analogous enzymes in *T. cruzi* and 124 enzymatic functions that are not present in humans. Table II shows the results of this analysis for the glycolysis and gluconeogenesis pathways and for the Krebs cycle. The corresponding metabolic maps are shown in Figs 1 and 2. Lipid and energy metabolism of the parasite have received considerable attention as potential drug targets in *T. cruzi* and are briefly reviewed below.

### Lipid metabolism

Lipids have essential roles in biological membranes and are sources of stored energy. They also perform sev-

TABLE I

Computational reconstruction of metabolic pathways in *Trypanosoma cruzi*: identification of enzymatic functions and analogy with human

| | KEGG annotation | AnEnPi analysis | | |
|---|---|---|---|---|
| | | $e^{-20}$ | $e^{-40}$ | $e^{-80}$ |
| Enzymatic functions for *T. cruzi* | 396 | 561 | 507 | 461 |
| Enzymatic activities with analogy | 295 | 422 | 376 | 337 |
| Analogy human - *T. cruzi* | NA | 33 | 33 | 29 |
| Functions specific for *T. cruzi* | NA | 124 | 106 | 81 |

number of those activities for which cases of analogy have been detected (considering all species in the KEGG database). Number of analogous enzymes between *Homo sapiens* and *T. cruzi* and number of enzymatic functions encountered that are specific for *T. cruzi* as a result of AnEnPi analysis. NA: not applicable to the KEGG annotation.

eral other functions, such as being cofactors for enzymes, acting as hormones, being intra and extra-cellular signal messengers and involvement in protein anchorage to membranes and transporters. In trypanosomes, several molecules, such as miltefosine and azasterol, have been evaluated as structural analogues of lipids, phospholipids and steroids. They function as inhibitors of key pathways in the synthesis of lipids and steroids, respectively (Croft et al. 2003, Roberts et al. 2003). The lipid metabolism of trypanosomatids has been studied since the 60s (Korn & Greenblatt 1963, Korn et al. 1969, Dixon et al. 1971, 1972) but the identification of key differences between the host and the parasite, as well as the study of the enzymes themselves, has not been fully achieved. The sterol biosynthesis pathway has been better characterised, especially because of its similarity with the corresponding pathways in fungi and yeasts where the main sterol is ergosterol (Roberts et al. 2003). Most of the experimental work is being done in *T. brucei* because this parasite can be subjected to gene silencing through RNAi (Lee et al. 2006, van Hellemond & Tielens 2006, Stephens et al. 2007). Several enzymes of the pathway have been cloned from *T. cruzi* and *Leishmania* and experimental drugs that affect these enzymes are being evaluated (Urbina 1997, Magaraci et al. 2003, Pourshafie et al. 2004, Lorente et al. 2005).

The first step in the synthesis of isoprenoid, steroids and other lipid components, such as farnesyl and dolichol, involves the synthesis of isopentenyl-pyrophosphate. There are two pathways for the synthesis of this compound: the mevalonate pathway from the acetyl-CoA precursor and the pathway via DOXP/MEP (1-deoxy-D-xylulose 5-phosphate/2-*C*-methyl-D-erythritol 4-phosphate), which includes glyceraldehyde-3-phosphate and pyruvate as precursors. The latter is present in algae, some bacteria and in protozoa apicomplexa and is targeted by antiparasitic drugs, such as fosmidomycin and derivates (Wiesner & Jomaa 2007, Haemers et al. 2008). On the other hand, higher eukaryotes and trypanosomatids synthesise isoprenoids through the mevalonate pathway (Urbina 1997, Buhaescu & Izzedine 2007).

Careful analysis of each enzyme in the pathway is under way, including the HMG-CoA reductase in several trypanosomatids (Pena-Diaz et al. 1997, Montalvetti et al. 2000, Hurtado-Guerrrero et al. 2002) and the mevalonate kinase in *Leishmania major* (Sgraja et al. 2007).

We are studying two enzymes from this pathway: isopentenyl diphosphate isomerase and phophomevalonate kinase, both of which are analogous to the corresponding human proteins (M Alves-Ferreira, unpublished observations).

The steroid metabolism in trypanosomatids was initially studied through the characterisation of intermediate metabolites using isotopic labelling and inhibitors for several yeast enzymes (Roberts et al. 2003). In epimastigotes of *T. cruzi,* ergosterol and ergosta-5,7,-dien-3β-ol represent about 40% of the steroids, while estigmata-5,7-dien-3β-ol and ergosta-5,7,22-trien-3β-ol represent about 30% (Beach et al. 1986). The amastigote form of the parasite does not seem to express the $\Delta^5$ and $\Delta^{22}$ desaturases and 80% of total sterol was cholesterol, which was probably present through incorporation from the host cells (Liendo et al. 1999). The presence of variants with substitutions on carbon-24 and the therapeutic effects of azasterols on the parasites clearly show the importance of the esterol-24-metyltranferase enzyme in the biosynthesis, besides the overall importance of the ergosteroids in the biology of the parasites. The enzyme sterol-24-metyltransferase has been described in several species of *Leishmania*, in *T. cruzi and* in *T. brucei* (Magaraci et al. 2003, Pourshafie et al. 2004, Jiménez-Jiménez et al. 2006). The exact intracellular localisation of the enzyme in *T. cruzi* is still unclear and was proposed to be associated with the glycosome as well as the cytoplasm. In our analyses using AnEnPi, we identified three homologous gene copies that encoded this enzyme in *T. cruzi*, corroborating the previous description of this enzyme and its indication as a possible new therapeutic target.

## The energy metabolism

Energy metabolism begins with nutrient uptake. In *T. cruzi,* a single isoform of the hexose transporter THT1 was described and several copies are present in the genome. In *T. brucei*, two isoforms of the transporter were identified (Bringaud & Baltz 1993, Tetaud et al. 1994). The transporter in *T. cruzi* has a high affinity for glucose, but can also transport other monosaccharides, such as D-fructose, which differs from the human transporter GLUT1 (Tetaud et al. 1994, Barrett et al. 1998). More recently, amino-acid transporters were also described for *T. cruzi*, including high affinity systems for arginine (Pereira et al. 1999) and aspartate (Canepa et al. 2005), as well as high and low affinity transporters for proline (Silber et al. 2005). A glutamate transporter was reported in *T. cruzi* and it was demonstrated that the uptake of this amino-acid is five times higher in epimastigotes than in metacyclics (Silber et al. 2006). It was further shown that aspartate, glutamine, asparagine, methionine, oxaloacetate and alpha-ketoglutarate also compete for use of this transport system.

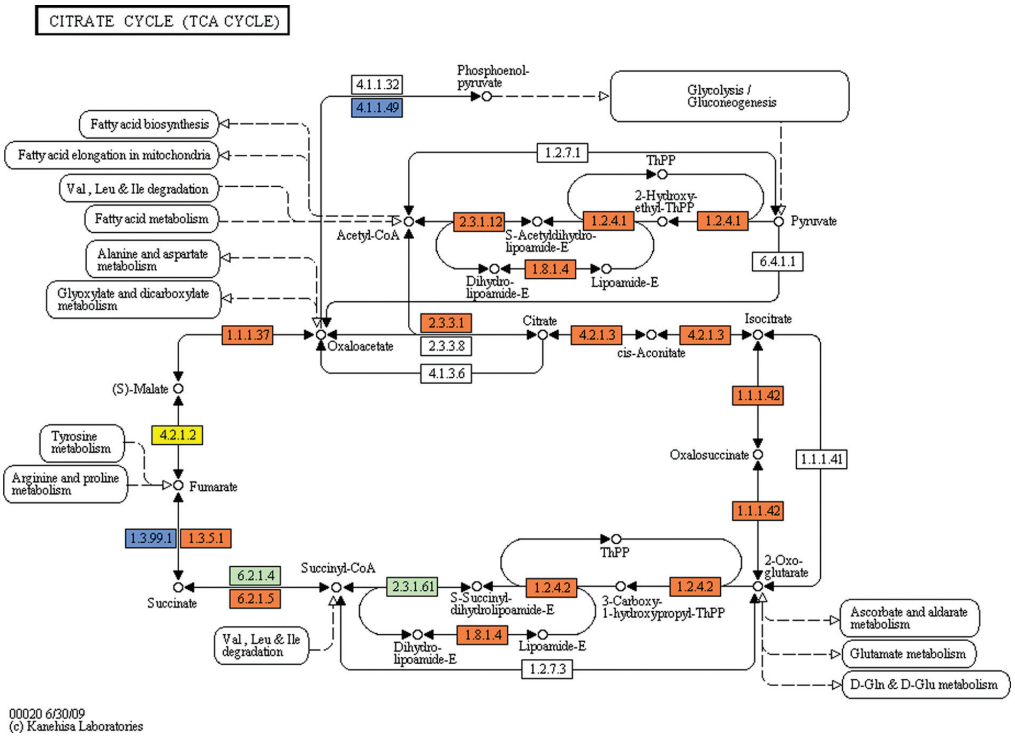



Fig. 2: reconstruction of the Krebs cycle pathway in *Trypanosoma cruzi*. The map was generated using the reference map by KEGG (http://www.genome.jp/kegg/pathway.html). Green boxes: enzymatic activities without putative cases of analogy (for all species in KEGG database); orange boxes: enzymatic activities with putative cases of analogy (also for all species in KEGG database); yellow boxes: enzymatic activities with probable cases of analogy between *Trypanosoma cruzi* and *Homo sapiens*; blue boxes: enzymatic activities specific for *T. cruzi* and absent in *H. sapiens*.

TABLE II

Computational reconstruction of the Krebs cycle and glycolysis-gluconeogenesis pathways in *Ttypanosoma cruzi*: identification of enzymatic functions and analogy with human

| | KEGG annotation | | AnEnPi analysis | |
|---|---|---|---|---|
| | Glycolysis | Krebs cycle | Glycolysis | Krebs cycle |
| Enzymatic functions for *T. cruzi* | 22 | 9 | 23 | 10 |
| Enzymatic activities with analogy | 20 | 9 | 21 | 10 |
| Analogy human - *T. cruzi* | NA | NA | 2 | 1 |
| Functions specific for *T. cruzi* | NA | NA | 2 | 1 |

analysis was as described in Table I, but specific to the glycolysis and gluconeogenesis pathways (Fig. 1) and the Krebs cycle (Fig. 2). An e-value cutoff of $e^{-80}$ was used. NA: not applicable to the KEGG annotation.

*Glycolysis* - The enzymes of the glycolytic pathway in trypanosomatids are organised in two cellular compartments, while in higher eukaryotes these enzymes are present in the cytoplasm. In *T. brucei*, the first seven enzymes of the pathway are present in the glycosome and transform glucose into 3-phosphoglycerate, while the three other enzymes are localised in the cytoplasm (Hannaert et al. 2003, Bringaud et al. 2006). *T. brucei* is an exclusively extracellular parasite and its metabolism is directly related to the concentration of nutrients in the medium. The slender bloodstream forms of the parasite produce energy almost exclusively from glucose. The phosphoglycerol kinase localisation is either glycosomal or cytosolic, as observed in the promastigotes of different *Leishmania* species and in *T. cruzi* epimastigotes and are required to maintain the glycosomal ATP/ADP balance, which offers a rationale for the presence of both phosphoglycerate mutase and enolase in the cytosol (Hannaert et al. 2003).

In anaerobic conditions, *T. brucei* produces mostly pyruvate and re-oxidises NADH through the use of an alternative oxidase (van Hellemond et al. 2005). However,

*T. cruzi* and *Leishmania* species require a more complex energy metabolic system and additional pathways, such as the Krebs cycle and oxidative phosphorylation, are compartmentalised in the mitochondria (Taylor & Gutteridge 1987, Hannaert et al. 2003).

In *T. cruzi*, the first enzyme of the pathway, a hexokinase (HK), was described as having different kinetic characteristics than the human enzyme. It did not suffer inhibition by D-glucose-6-phosphate or by other vertebrate HK regulators, such as fructose-1,6-diphosphate, phosphoenolpyruvate, lactate or citrate; although there is a weak competitive inhibition by ADP with respect to ATP (Racagni et al. 1983, Urbina & Crespo 1984). More recently, additional experiments showed that the enzyme is inhibited in a non-competitive way by inorganic pyrophosphate (PPi) and does not phosphorylate other sugars, such as fructose, mannose and galactose (Cáceres et al. 2003). This suggests that biphosphonates are possible inhibitors of this *T. cruzi* enzyme. Therefore, the synthesis of 42 compounds was described, in which the best of this series demonstrated a $K_i$ of 2.2 μM against amastigotes. Other reports in this direction showed the potent and selective inhibition of HK and inhibition of the proliferation of the clinically relevant intracellular amastigote form of the parasite, using aromatic arinomethylene biphosphonates, which act as non-competitive or mixed inhibitors of HK (Hudock et al. 2006, Sanz-Rodrıguez et al. 2007). However, Cordeiro et al. (2007) showed that the parasite also has a glucokinase with ten times higher affinity for glucose preferentially in a β-D-glucose isomeric form, while the preferential HK substrate is the α-isomer. Using the AnEnPi tool, *in silico* annotation of *T. cruzi* predicted proteins indicated the presence of both HK and glucokinase. Fig. 1 shows the reconstruction of the glycolytic pathway and where these two enzymes are located (EC 2.7.1.1 and EC 2.7.1.2). Unlike the HK in this analysis the glucokinase is clearly analogous to the human enzyme and constitutes a possibly interesting therapeutic target in the parasite. Recently, Chambers et al. (2008) showed that the anti-cancer drug ionidamine was capable of inhibiting one of the HKs (TbHK1) from *T. brucei* and was effective against both the recombinant enzyme as well as the bloodstream and procyclic forms of this parasite.

Another point of regulation in the glycolytic pathway of eukaryotes resides in the catalytic step of the phosphofructokinase. This enzyme displays different characteristics in *T. cruzi,* including the fact that the enzyme is not sensitive to typical modulators, such as the activator fructose-6-P and the inhibitors ATP, citrate and phosphoenolpyruvate (Urbina & Crespo 1984, Adroher et al. 1990).

The triose phosphate isomerase (TIM) is also a central enzyme of the glycolytic pathway that has been studied in *T. cruzi* and in a number of pathogenic protozoa (Pérez-Montfort et al. 1999, Reyes-Vivas et al. 2001, Rodríguez-Romero et al. 2002, Olivares-Illana et al. 2006). Recently, two studies showed the possibility of developing new therapeutic agents against trypanosomes using TIM as a molecular target. The compound 6, 6'-bisbenzothiazole-2, 2' diamine in the low micromolar range was able to specifically inhibit the TIM of trypanosomatids (Olivares-Illana et al. 2006), while the compound dithiodianiline in nanomolar concentrations was able to completely inhibit the recombinant TIM of *T. cruzi* and was trypanocidal for the epimastigote form of the parasite (Olivares-Illana et al. 2007).

The phosphoglycerate kinase (PGK) was described by Concepción et al. (2001) as having two isoforms: one of 56 kDa, which is exclusively glycosomal, and a second of 48kDa, which is expressed both in the cytoplasm and in glycosomes. In the same study, the authors demonstrated that 20% of total PGK activity is found in glycosomes and 80% in the cytoplasm.

The pyruvate kinase in *T. cruzi* was first reported by Juan et al. (1976) and following studies demonstrated the kinetic details, regulation and localisation of this enzyme (Cazzulo et al. 1989, Adroher et al. 1990).

The enzyme alcohol dehydrogenase (EC: 1.1.1.2), which is important in oxidoreduction of alcohol, can be considered a possible target for therapeutic studies against *T. cruzi* because AnEnPi analysis indicates it as being an analogue to the enzyme found in humans (Table II). It was demonstrated that the corresponding enzyme from *Entamoeba histolytica* can be inhibited by cyclopropyl and cyclobutyl carbinols and is considered as a possible new drug target for this parasite as well as for *Giardia lamblia* (Espinosa et al. 2004).

*The pentose phosphate shunt* - The pentose phosphate shunt is the most studied pathway in *T. cruzi* and the participating enzymes have been well characterised both molecularly and biochemically (Igoillo-Esteve et al. 2007). The enzymes of this pathway were identified as cytoplasmic, but with secondary localisation in the glycosome. They play an important role in the generation of NADPH, which is essential in responding to oxidative processes of the host defence system. Maugeri and Cazzulo (2004) showed that under normal conditions 10% of the glucose captured by *T. cruzi* is metabolised in the pentose pathway. All seven enzymes of the pathway are expressed in the four life stages of *T. cruzi* (Maugeri & Cazzulo 2004) and this pathway can be divided into oxidative and non-oxidative branches.

The oxidative branch is responsible for the production of NADPH and ribose-5-phosphate and is regulated by the ratio of NADP/NADPH through the first enzyme in the process, glucose-6-phosphate dehydrogenase (G6PD). This enzyme has several copies in the *T. cruzi* genome, while the 6-phosphogluconolactonase, the 6-phosphogluconate dehydrogenase (6PGD) and the ribose 5-phosphate isomerase are present as only one copy. The latter enzyme is analogous to the human enzyme, which makes it an interesting therapeutic target (Igoillo-Esteve et al. 2007). Recently, Stern et al. (2007), using site directed mutagenesis, revealed details on the reaction mechanism of this enzyme, which acts via the formation of the 1,2-cis-enediol. They also showed that the 4-phospho-D-erythronohydroxamic acid, an analogue to the reaction intermediate, competitively inhibits this enzyme with a $K_i$ = 1.2 mM and an IC50 of 0.7 mM. Mielniczki-Pereira et al. (2007) studied the two enzymes responsible for NADPH production, G6PD and 6PGD and showed differences in the activities of those

enzymes between the Tulahuen 2 and Y strains, as well as the importance of G6PD in protecting the parasite against reactive oxygen species.

The non-oxidative branch complements the pathway and the enzymes involved are ribulose-5-phosphate epimerase, with two gene copies in *T. cruzi* and the transaldolase and transketolase, each with one copy per haploid genome (Igoillo-Esteve et al. 2007). One of the copies of ribulose-5-phophate isomerase presents a C-terminal (PTS1) signal for glycosome targeting but the cytosol also showed activity (Maugeri & Cazzulo 2004). A transaldolase is responsible for the transfer of the dihydroxyacetone group from fructose-6-phosphate to erythrose-4-phosphate, leading to the synthesis of sedoheptulose-7-phosphate and glyceraldeyde-3-P. The biochemical characterisation of the recombinant enzyme was reported (Igoillo-Esteve et al. 2007). The transketolase of *T. cruzi* contains the PTS1 signal peptide, which possibly permits its distribution both in the glycosome and in the cytoplasm. The recombinant enzyme is apparently a dimer of 146 kDa (Igoillo-Esteve et al. 2007).

Using the AnEnPi approach, we identified the sedoheptulose-bisphosphatase (EC: 3.1.3.37) that performs the addition of $H_20$ to sedoheptulose 1,7-bisphosphate, resulting in sedoheptulose 7-phosphate and phosphate. This enzyme is a possible target for the development of drugs because they are specific to the parasite and absent in the human host.

*The Krebs cycle and oxidative phosphorylation* - The intermediate energy metabolism that occurs in the mitochondria of *T. cruzi* was studied by several groups and reported several decades ago (von Brand & Agosin 1955, de Boiso & Stoppani 1973, Docampo et al. 1978, Rogerson & Gutteridge 1980). Cannata and Cazzulo (1984) reviewed the metabolism of carbohydrates and described, in detail, current knowledge of the Krebs cycle, as well as the process of aerobic fermentation in this parasite. Within this context, the main products of this peculiar metabolic pathway would be succinate from the reduction of fumarate coupled with the reduction of oxaloacetate and the generation of L-malate. More recently, the intermediary energy metabolism was reviewed by Urbina et al. (1993), including a discussion on the possibility of regulating the activity of the Krebs cycle according to the action of two glutamate dehydrogenases, one $NAD^+$-dependent and the other $NADP^+$ dependent, which indicates the use of amino acids as a main energy source for the parasite.

The constitutive expression of the ATP-dependent phosphoenolpyruvate carboxykinase (PEPCK, EC 4.1.1.49) was described in all evolutionary forms of *T. cruzi*. The enzyme catalyses the reaction from phosphoenolpyruvate to oxaloacetate yielding ATP. In vertebrates, this enzyme is involved in glycogenesis, while in *T. cruzi* it appears to be involved in catabolism. Additionally, the inhibition of this enzyme by 3-mercaptopicolinic acid has been described in studies with the purified enzyme and in vivo (Urbina et al. 1990). This enzyme was also shown to have a specific requirement for transition metal ions that modulate the reactivity of a single essential thiol group, which differs from the hyper-reactive cysteines present in vertebrates or yeast (Jurado et al. 1996). Trapani et al. (2001) elucidated the 3D and dimeric structures of the PEPCK of *T. cruzi,* permitting progress in the study of this enzyme for the development of new therapeutic agents.

Our data from *in silico* analysis indicate the presence of all the enzymes involved in the Krebs cycle, corroborating the analyses by KEGG, but also indicated a phosphoenolpyruvate carboxykinase (EC: 4.1.1.49) specific for *T. cruzi* and absent in humans (Table II, Fig. 2). Furthermore, the enzyme fumarate reductase (EC: 1.3.99.1) is also specific for *T. cruzi* and fumar-

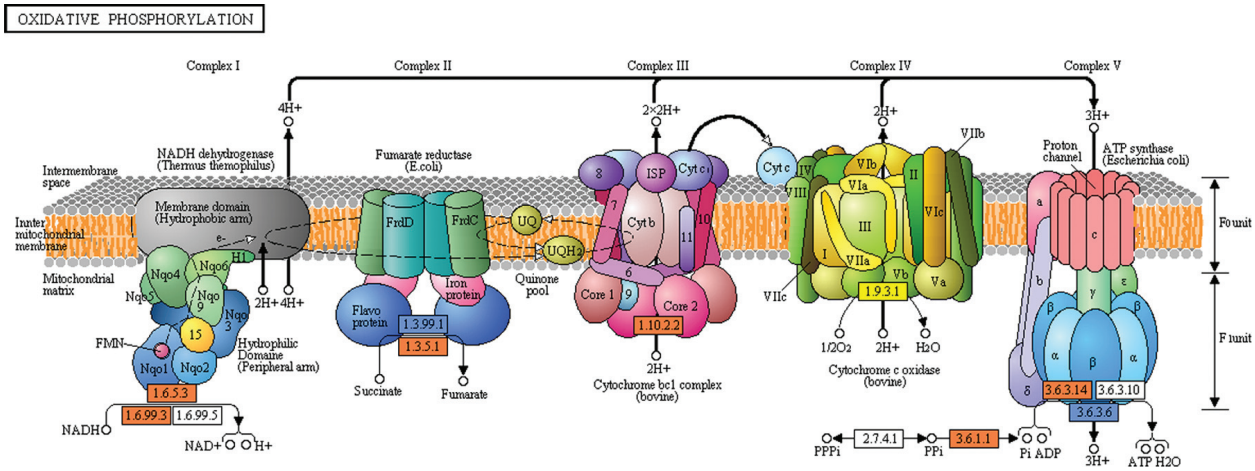



Fig. 3: map of the oxidative phosphorylation metabolism in *Trypanosoma cruzi* with a representation of the multi-enzyme complexes and their corresponding enzymatic function. The map was generated and adapted from the reference map by KEGG (http://www.genome.jp/kegg/pathway.html). Colored boxes symbolize activities annotated by AnEnPi methodology. Green boxes: enzymatic activities without putative cases of analogy (for all species in KEGG database); orange boxes: enzymatic activities with putative cases of analogy (also for all species in KEGG database); yellow boxes: enzymatic activities with probable cases of analogy between *T. cruzi* and *Homo sapiens*; blue boxes: enzymatic activities specific for *T. cruzi* and absent in *H. sapiens*.

ate hydratase (EC: 4.2.1.2) in the parasite is analogous (Fig. 2) to the human protein.

With relation to the mechanism of oxidative phosphorylation and the respiratory chain, our data and those of KEGG indicate the presence of all the multi-enzyme complexes in *T. cruzi* (Fig. 3), with several subunits that are either specific to the parasite or analogous to the human enzymes. Reports in the literature suggest that the process of oxidative phosphorylation in *T. cruzi* does not occur in a manner similar to humans, having been called oxidative fermentation (Cannata & Cazzulo 1984). Affranchino et al. (1985) conducted a study on the regulation of mitochondrial respiration in epimastigotes of *T. cruzi,* which showed, among other peculiarities, that the oxidation of NADH is not controlled by ADP.

*β-oxidation* - The oxidation of fatty acids is an important source of ATP in many organisms, but this is apparently not the case for most parasites (van Hellemond & Tielens 2006). Initial analysis of the *T. brucei* genome (Berriman et al. 2005) but also of *T. cruzi* and *Leishmania* identified homologous genes for the four enzymes responsible for β-oxidation of fatty acids and this pathway probably occurs both in glycosomes and mitochondria (van Hellemond & Tielens 2006). The oxidation rates, however, seem minimal and oxidation can be for very specific fatty acids only or occur only under special conditions (Wiemer et al. 1996).

The enzyme pyruvate phosphate dikinase was described in *T. cruzi* epimastigotes with a glycosomal localisation. The complete function of this enzyme is not completely understood but its reaction probably leads to pyruvate production from phosphoenolpyruvate and PPi (Acosta et al. 2004). The authors showed that palmitoyl-CoA β-oxidation occurs in glycosomes and suggest that this enzyme could be a link between glycolysis, fatty acid oxidation and the biosynthetic PPi-producing pathways in this organelle, as well as replacing pyrophosphatase in its classical thermodynamic role and eliminating the toxic PPi.

*Use of amino acids for energy production* - *T. cruzi* epimastigotes metabolise asparagine, aspartate, glutamine, glutamate, leucine, isoleucine and proline (Sylvester & Krassner 1976). Alanine, aspartate and glutamate provide the Krebs cycle with intermediates (Silber et al. 2005), while the latter is an important intermediate in proline metabolism, which is an important carbon and energy source for the parasite. The amino group of glutamate can be transferred to pyruvate by both alanine aminotransferase and tyrosine aminotransferase, yielding a-ketoglutarate and alanine (Cazzulo 1994) or can be deaminated by glutamate dehydrogenase, which is localised in the cytoplasm and mitochondria (Duschak & Cazzulo 1991). A detailed computational analysis of the *T. cruzi* metabolic pathways involving amino acid metabolism was described previously and many instances of analogous enzymes in the parasite, compared to human, were pointed out (Guimarães et al. 2008). Many questions still remain open regarding energy metabolism in *T. cruzi* and our understanding of the physiological mechanisms involved in the generation of energy in this parasite. Furthermore, the identification of new therapeutic targets needs further work.

## The polyamine metabolism

Due to the importance of polyamines for trypanosomatids, which are highly dependent on spermidine for growth and survival, polyamine metabolism has been well studied and several enzymes, such as arginase, ornithine decarboxylase (ODC), S-adenosylmethionine decarboxylase (AdoMetDC), spermidine synthase, trypanothione synthetase and trypanothione reductase have been targeted for the development of new drugs (Heby et al. 2007). *T. cruzi* does not contain ODC, an effective target of alpha-difluoromethylornithine for the treatment of sleeping sickness. Instead, the parasite salvages putrescine or spermidine from the host. Two candidate aminopropyltransferases have been proposed. Although *T. cruzi* maintains an apparently functional copy of the AdoMetDC, known inhibitors of this enzyme do not have much effect on the parasite (Beswick et al. 2006). The traditional cellular redox couple formed by the otherwise ubiquitous glutathione/glutathione reductase couple is replaced in trypanosomatids by the dithiol bis(glutathionyl)spermidine called trypanothione and the flavoenzyme trypanothione reductase. Trypanothione is the reducing agent of thioredoxin and tryparedoxin, which are small dithiol proteins that are reducing agents for the synthesis of deoxyribonucleotides, as well as for the detoxification of hydroperoxides by different peroxidases. The trypanothione reductase is an essential enzyme for the parasite and its absence in the mammalian host makes it an interesting target for drug development (Krauth-Siegel & Inhoff 2003, Martyn et al. 2007).

More detailed analysis of biochemical pathways and increased high-throughput screening activities, using both synthetic compounds and natural products, set high hopes for the development of new drugs against Chagas disease, leishmaniasis and sleeping sickness. The many ongoing initiatives and the identification of numerous potential targets bring hope that a breakthrough in the treatment of these parasitic diseases will be shortcoming.

## REFERENCES

Acosta H, Dubourdieu M, Quinõnes W, Cáceres A, Bringaud F, Concepción JL 2004. Pyruvate phosphate dikinase and pyrophosphate metabolism in the glycosome of *Trypanosoma cruzi* epimastigotes. *Comp Biochem and Physiol B 138*: 347-356.

Adroher FJ, Osuna A, Lupiáñez JÁ 1990. Differential energetic metabolism during *Trypanosoma cruzi* differentiation. II. Hexokinase, phosphofructokinase and pyruvate kinase. *Mol Cell Biochem 94*: 71-82.

Affranchino JL, De Tarlovsky MN, Stoppani AO 1985. Respiratory control in mitochondria from *Trypanosoma cruzi*. *Mol Biochem Parasitol 16*: 289-298.

Atwood JA 3rd, Weatherly DB, Minning TA, Bundy B, Cavola C, Opperdoes FR, Orlando R, Tarleton RL 2005. The *Trypanosoma cruzi* proteome. *Science 309*: 473-476.

Barrett MP, Tetaud E, Seyfang A, Bringaud F, Baltz T 1998. Trypanosome glucose transporters. *Mol Biochem Parasitol 91*: 195-205.

Beach DH, Goad LJ, Holz GG Jr 1986. Effects of ketoconazole on sterol biosynthesis by *Trypanosoma cruzi* epimastigotes. *Biochem Biophys Res Commun 136*: 851-856.

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, Lennard NJ, Caler E, Hamlin NE, Haas B, Böhme U, Hannick L, Aslett MA, Shallom J, Marcello L, Hou L, Wickstead B, Alsmark UCM, Arrowsmith C, Atkin RJ, Barron AJ, Bringaud F, Brooks K, Carrington M, Cherevach I, Chillingworth T-J, Churcher C, Clark LN, Corton CH, Cronin A, Davies RM, Doggett J, Djikeng A, Feldblyum T, Field MC, Fraser A, Goodhead I, Hance Z, Harper D, Harris BR, Hauser H, Hostetler J, Ivens A, Jagels K, Johnson D, Johnson J, Jones K, Kerhornou AX, Koo H, Larke N, Landfear S, Larkin C, Leech V, Line A, Lord A, MacLeod A, Mooney PJ, Moule S, Martin DMA, Morgan GW, Mungall K, Norbertczak H, Ormond D, Pai G, Peacock CS, Peterson J, Quail MA, Rabbinowitsch E, Rajandream M-A, Reitter C, Salzberg SL, Sanders M, Schobel S, Sharp S, Simmonds M, Simpson AJ, Tallon L, Turner CMR, Tait A, Tivey AR, Van Aken S, Walker D, Wanless D, Wang S, White B, White O, Whitehead S, Woodward J, Wortman J, Adams MD, Embley TM, Gull K, Ullu E, Barry JD, Fairlamb AH, Opperdoes F, Barrell BG, Donelson JE, Hall N, Fraser CM, Melville SE, El-Sayed NM 2005. The genome of the African trypanosome *Trypanosoma brucei*. *Science 309*: 416-422.

Beswick TC, Willert EK, Phillips MA 2006. Mechanisms of allosteric regulation of *Trypanosoma cruzi* S-adenosylmethionine decarboxylase. *Biochemistry 45*: 7797-7807.

Bilate AM, Cunha-Neto E 2008. Chagas disease cardiomyopathy: current concepts of an old disease. *Rev Inst Med Trop Sao Paulo 50*: 67-74.

Brak K, Doyle PS, McKerrow JH, Ellman JA 2008. Identification of a new class of nonpeptidic inhibitors of cruzain. *J Am Chem Soc 130*: 6404-6410.

Bringaud F, Baltz T 1993. Differential regulation of two distinct families of glucose transporter genes in *Trypanosoma brucei*. *Mol Cell Biol 13*: 1146-1154.

Bringaud F, Rivière L, Coustou V 2006. Energy metabolism of trypanosomatids: adaptation to available carbon sources. *Mol Biochem Parasitol 149*: 1-9.

Buhaescu I, Izzedine H 2007. Mevalonate pathway: a review of clinical and therapeutical implications. *Clin Biochem 40*: 575-584.

Cáceres AJ, Portillo R, Acosta H, Rosales D, Quinõnes W, Avilan L, Salazar L, Dubourdieu M, Michels PAM, Concepción JL 2003. Molecular and biochemical characterization of hexokinase from *Trypanosoma cruzi*. *Mol Biochem Parasitol 126*: 251-262.

Canepa GE, Bouvier LA, Urias U, Miranda MR, Colli W, Alves MJ, Pereira CA 2005. Aspartate transport and metabolism in the protozoan parasite *Trypanosoma cruzi*. *FEMS Microbiol Lett 247*: 65-71.

Cannata JJ, Cazzulo JJ 1984. The aerobic fermentation of glucose by *Trypanosoma cruzi*. *Comp Biochem Physiol B 79*: 297-308.

Cavazzuti A, Paglietti G, Hunter WN, Gamarro F, Piras S, Loriga M, Allecca S, Corona P, McLuskey K, Tulloch L, Gibellini F, Ferrari S, Costi MP 2008. Discovery of potente pteridine reductase inhibitors to guide antiparasite drug development. *Proc Natl Acad Sci USA 105*: 1448-1453.

Cazzulo JJ 1994. Intermediate metabolism in *Trypanosoma cruzi*. *J Bioenerg Biomembr 26*: 157-165.

Cazzulo JJ, Cazzulo Franke MC, Franke de Cazzulo BM 1989. On the regulatory properties of the pyruvate kinase from *Trypanosoma cruzi* epimastigotes. *FEMS Microbiol Lett 50*: 259-263.

Chambers JW, Fowler ML, Morris MT, Morris JC 2008. The antitrypanosomal agent lonidamine inhibits *Trypanosoma brucei* hexokinase 1. *Mol Biochem Parasitol 158*: 202-207.

Concepción JL, Adjé CA, Quiñones W, Chevalier N, Dubourdieu M, Michels PA 2001. The expression and intracellular distribution of phosphoglycerate kinase isoenzymes in *Trypanosoma cruzi*. *Mol Biochem Parasitol 118*: 111-121.

Cordeiro AT, Cáceres AJ, Vertommen D, Concepción JL, Michels PA, Versées W 2007. The crystal structure of *Trypanosoma cruzi* glucokinase reveals features determining oligomerization and anomer specificity of hexose-phosphorylating enzymes. *J Mol Biol 372*: 1215-1226.

Coura JR, Castro SL 2002. A critical review on Chagas disease chemotherapy. *Mem Inst Oswaldo Cruz 97*: 3-24.

Croft SL, Seifert K, Duchêne M 2003. Antiprotozoal activities of phospholipid analogues. *Mol Biochem Parasitol 126*: 165-172.

de Boiso JF, Stoppani AO 1973. The mechanism of acetate and pyruvate oxidation by *Trypanosoma cruzi*. *J Protozool 20*: 673-678.

de Souza W 2002. From the cell biology to the development of new chemotherapeutic approaches against trypanosomatids: dreams and reality. *Kinetoplastid Biol Dis 1*: 3.

de Souza W 2005. Microscopy and cytochemistry of the biogenesis of the parasitophorous vacuole. *Histochem Cell Biol 123*: 1-18.

de Souza W, Campanati L, Attias M 2008. Strategies and results of field emission scanning electron microscopy (FE-SEM) in the study of parasitic protozoa. *Micron 39*: 77-87.

Dias JC 2006. Chagas disease: successes and challenges. *Cad Saude Publica 22*: 2020-2021.

Dixon H, Ginger CD, Williamson J 1971. The lipid metabolism of blood and culture forms of *Trypanosoma lewisi* and *Trypanosoma rhodesiense*. *Comp Biochem Physiol B 39*: 247-266.

Dixon H, Ginger CD, Williamson J 1972. Trypanosome sterols and their metabolic origins. *Comp Biochem Physiol B 41*: 1-18.

Docampo R, de Boiso JF, Stoppani AO 1978. Tricarboxylic acid cycle operation at the kinetoplast-mitochondrion complex of *Trypanosoma cruzi*. *Biochim Biophys Acta 502*: 466-476.

Docampo R, de Souza W, Miranda K, Rohloff P, Moreno SN 2005. Acidocalcisomes - conserved from bacteria to man. *Nat Rev Microbiol 3*: 251-261.

Doyle PS, Zhou YM, Engel JC, McKerrow JH 2007. A cysteine protease inhibitor cures Chagas' disease in an immunodeficient-mouse model of infection. *Antimicrob Agents Chemother 51*: 3932-3939.

Duschak VG, Cazzulo JJ 1991. Subcellular localization of glutamate dehydrogenases and alanine aminotransferase in epimastigotes of *Trypanosoma cruzi*. *FEMS Microbiol Lett 67*: 131-135.

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou L, Huang Y, Kindlund E, Klingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarleton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B 2005. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science 309*: 409-415.

Espinosa A, Clark D, Stanley SL Jr 2004. *Entamoeba histolytica* alcohol dehydrogenase 2 (EhADH2) as a target for anti-amoebic agents. *J Antimicrob Chemother 54*: 56-59.

Freitas RF, Prokopczyk IM, Zottis A, Oliva G, Andricopulo AD, Trevisan MT, Vilegas W, Silva MG, Montanari CA 2009. Discovery of novel *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase inhibitors. *Bioorg Med Chem 17*: 2476-2482.

Fricker SP, Mosi RM, Cameron BR, Baird I, Zhu Y, Anastassov V, Cox J, Doyle PS, Hansell E, Lau G, Langille J, Olsen M, Qin L, Skerlj R, Wong RS, Santucci Z, McKerrow JH 2008. Metal compounds for the treatment of parasitic diseases. *J Inorg Biochem 102*: 1839-1845.

Galarreta BC, Sifuentes R, Carrillo AK, Sanchez L, Amado M del R, Maruenda H 2008. The use of natural product scaffolds as leads in the search for trypanothione reductase inhibitors. *Bioorg Med Chem 16*: 6689-6695.

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

Guimarães AC, Otto TD, Alves-Ferreira M, Miranda AB, Degrave WM 2008. *In silico* reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi*. *Genet Mol Res 7*: 872-882.

Haemers T, Wiesner J, Gießmann D, Verbrugghen T, Hillaert U, Ortmann R, Jomaa H, Link A, Schlitzerd M, Van Calenbergha S 2008. Synthesis of b- and c-oxa isosteres of fosmidomycin and FR900098 as antimalarial candidates. *Bio Med Chem 16*: 3361-3371.

Hannaert V, Albert MA, Rigden DJ, da Silva Giotto MT, Thiemann O, Garratt RC, Van Roy J, Opperdoes FR, Michels PA 2003. Kinetic characterization, structure modelling studies and crystallization of *Trypanosoma brucei* enolase. *Eur J Biochem 270*: 3205-3213.

Heby O, Persson L, Rentala M 2007. Targeting the polyamine biosynthetic enzymes: a promising approach to therapy of African sleeping sickness, Chagas' disease and leishmaniasis. *Amino Acids 33*: 359-366.

Hudock MP, Sanz-Rodriguez CE, Song Y, Chan JM, Zhang Y, Odeh S, Kosztowski T, Leon-Rossell A, Concepcion JL, Yardley V, Croft SL, Urbina JA, Oldfield E 2006. Inhibition of *Trypanosoma cruzi* hexokinase by bisphosphonates. *J Med Chem 49*: 215-223.

Hurtado-Guerrrero R, Peña-Díaz J, Montalvetti A, Ruiz-Pérez LM, González-Pacanowska D 2002. Kinetic properties and inhibition of *Trypanosoma cruzi* 3-hydroxy-3-methylglutaryl CoA reductase. *FEBS Letters 510*: 141-144.

Igoillo-Esteve M, Maugeri D, Stern AL, Beluardi P, Cazzulo JJ 2007. The pentose phosphate pathway in *Trypanosoma cruzi*: a potential target for the chemotherapy of Chagas disease. *An Acad Bras Cienc 79*: 649-663.

Jiménez-Jiménez J, Ledesma A, Zaragoza P, González-Barroso MM, Rial E 2006. Fatty acid activation of the uncoupling proteins requires the presence of the central matrix loop from UCP1. *Biochim Biophys Acta 1757*: 1292-1296.

Juan SM, Cazzulo JJ, Segura EC 1976. The pyruvate kinase of *Trypanosoma cruzi*. *Acta Physiol Lat Am 26*: 424-426.

Jurado LA, Machín I, Urbina JA 1996. *Trypanosoma cruzi* phosphoenolpyruvate carboxykinase (ATP-dependent): transition metal ion requirement for activity and sulfhydryl group reactivity. *Biochim Biophys Acta 1292*: 188-196.

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M 2006. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res 34*: D354-D357.

Karp PD, Krummenacker M, Paley S, Wagg J 1999. Integrated pathway-genome databases and their role in drug discovery. *Trends Biotechnol 17*: 275-281.

Korn ED, Greenblatt CL 1963. Synthesis of alpha-linoleic acid by *Leishmania enriettii*. *Science 142*: 1301-1303.

Korn ED, Von Brand T, Tobie EJ 1969. The sterols of *Trypanosoma cruzi* and Crithidia fasciculata. *Comp Biochem Physiol 30*: 601-610.

Krauth-Siegel RL, Inhoff O 2003. Parasite-specific trypanothione reductase as a drug target molecule. *Parasitol Res 90*: S77-S85.

Lee SH, Stephens JL, Paul KS, Englund PT 2006. Fatty acid synthesis by elongases in trypanosomes. *Cell 126*: 691-699.

Liendo A, Visbal G, Piras MM, Piras R, Urbina JA 1999. Sterol composition and biosynthesis in *Trypanosoma cruzi* amastigotes. *Mol Biochem Parasitol 104*: 81-91.

Liñares GE, Ravaschino EL, Rodriguez JB 2006. Progresses in the field of drug design to combat tropical protozoan parasitic diseases. *Curr Med Chem 13*: 335-360.

Lorente SO, Jimenez CJ, Gros L, Yardley V, de Luca-Fradley K, Croft SL, Urbina J, Ruiz-Perez LM, Pacanowska DG, Gilbert IH 2005. Preparation of transition-state analogues of sterol 24-methyl transferase as potential anti-parasitics. *Bioorg Med Chem 13*: 5435-5453.

Magaraci F, Jimenez CJ, Rodrigues C, Rodrigues JC, Braga MV, Yardley V, de Luca-Fradley K, Croft SL, de Souza W, Ruiz-Perez LM, Urbina J, Gonzalez Pacanowska D, Gilbert IH 2003. Azasterols as inhibitors of sterol 24-methyltransferase in *Leishmania* species and *Trypanosoma cruzi*. *J Med Chem 46*: 4714-4727.

Marin-Neto JA, Rassi A Jr, Morillo CA, Avezum A, Connolly SJ, Sosa-Estani S, Rosas F, Yusuf S 2008. Rationale and design of a randomized placebo-controlled trial assessing the effects of etiologic treatment in Chagas' cardiomyopathy: the BENznidazole Evaluation for Interrupting Trypanosomiasis (BENEFIT). *Am Heart J 156*: 37-43.

Martyn DC, Jones DC, Fairlamb AH, Clardy J 2007. High-throughput screening affords novel and selective trypanothione reductase inhibitors with anti-trypanosomal activity. *Bioorg Med Chem Lett 17*: 1280-1283.

Maugeri DA, Cazzulo JJ 2004. The pentose phosphate pathway in *Trypanosoma cruzi*. *FEMS Microbiol Lett 234*: 117-123.

Maya JD, Cassels BK, Iturriaga-Vásquez P, Ferreira J, Faúndez M, Galanti N, Ferreira A, Morello A 2007. Mode of action of natural and synthetic drugs against *Trypanosoma cruzi* and their interaction with the mammalian host. *Comp Biochem Physiol A Mol Integr Physiol 146*: 601-620.

Michels PA, Hannaert V, Bringaud F 2000. Metabolic aspects of glycosomes in trypanosomatidae - new data and views. *Parasitol Today 16*: 482-489.

Mielniczki-Pereira AA, Chiavegatto CM, López JA, Colli W, Alves MJM, Gadelha FR 2007. *Trypanosoma cruzi* strains, Tulahuen 2 and Y, besides the difference in resistance to oxidative stress, display differential glucose-6-phosphate and 6-phosphogluconate dehydrogenases activities. *Acta Tropica 101*: 54-60.

Montalvetti A, Peña-Díaz J, Hurtado R, Ruiz-Pérez LM, González-Pacanowska D 2000. Characterization and regulation of *Leishmania major* 3-hydroxy-3-methylglutaryl-CoA reductase. *Biochem J 349*: 27-34.

Moyersoen J, Choe J, Fan E, Hol WG, Michels PA 2004. Biogenesis of peroxisomes and glycosomes: trypanosomatid glycosome assembly is a promising new drug target. *FEMS Microbiol Rev 28*: 603-643.

Olivares-Illana V, Pérez-Montfort R, López-Calahorra F, Costas M, Rodríguez-Romero A, Tuena de Gómez-Puyou M, Gómez Puyou A 2006. Structural differences in triosephosphate isomerase from different species and discovery of a multitrypanosomatid inhibitor. *Biochemistry 45*: 2556-2560.

Olivares-Illana V, Rodríguez-Romero A, Becker I, Berzunza M, García J, Pérez-Montfort R, Cabrera N, López-Calahorra F, Gómez-Puyou MT, Gómez-Puyou A 2007. Perturbation of the

dimer interface of triosephosphate isomerase and its effect on *Trypanosoma cruzi*. *PLoS Negl Trop Dis 1*: e1.

Otto TD, Guimarães AC, Degrave WM, de Miranda AB 2008. AnEnPi: identification and annotation of analogous enzymes. *BMC Bioinformatics 9*: 544.

Paulino M, Iribarne F, Dubin M, Aguilera-Morales S, Tapia O, Stoppani AO 2005. The chemotherapy of Chagas' disease: an overview. *Mini Rev Med Chem 5*: 499-519.

Peña-Díaz J, Montalvetti A, Camacho A, Gallego C, Ruiz-Perez LM, Gonzalez-Pacanowska D 1997. A soluble 3-hydroxy-3-methylglutaryl-CoA reductase in the protozoan *Trypanosoma cruzi*. *Biochem J 324*: 619-626.

Pereira CA, Alonso GD, Paveto MC, Flawiá MM, Torres HN 1999. L-arginine uptake and L-phosphoarginine synthesis in *Trypanosoma cruzi*. *J Eukaryot Microbiol 46*: 566-570.

Pérez-Montfort R, Garza-Ramos G, Alcántara GH, Reyes-Vivas H, Gao XG, Maldonado E, de Gómez-Puyou MT, Gómez-Puyou A 1999. Derivatization of the interface cysteine of triosephosphate isomerase from *Trypanosoma brucei* and *Trypanosoma cruzi* as probe of the interrelationship between the catalytic sites and the dimer interface. *Biochemistry 38*: 4114-4120.

Perez-Pineiro R, Burgos A, Jones DC, Andrew LC, Rodriguez H, Suarez M, Fairlamb AH, Wishart DS 2009. Development of a novel virtual screening cascade protocol to identify potential trypanothione reductase inhibitors. *J Med Chem 52*: 1670-1680.

Pourshafie M, Morand S, Virion A, Rakotomanga M, Dupuy C, Loiseau PM 2004. Cloning of S-adenosyl-L-methionine: C-24-Delta-sterol-methyltransferase (ERG6) from *Leishmania donovani* and characterization of mRNAs in wild-type and amphotericin B-Resistant promastigotes. *Antimicrob Agents Chemother 48*: 2409-2414.

Racagni GE, Machado de Domenech EE 1983. Characterization of *Trypanosoma cruzi* hexokinase. *Mol Biochem Parasitol 9*: 181-188.

Reyes-Vivas H, Hernández-Alcantara G, López-Velazquez G, Cabrera N, Pérez-Montfort R, de Gómez-Puyou MT, Gómez-Puyou A 2001. Factors that control the reactivity of the interface cysteine of triosephosphate isomerase from *Trypanosoma brucei* and *Trypanosoma cruzi*. *Biochemistry 40*: 3134-3140.

Roberts CW, McLeod R, Rice DW, Ginger M, Chance ML, Goad LJ 2003. Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa. *Mol Biochem Parasitol 126*: 129-142.

Rocha MO, Teixeira MM, Ribeiro AL 2007. An update on the management of Chagas cardiomyopathy. *Expert Rev Anti Infect Ther 5*: 727-743.

Rodríguez-Romero A, Hernández-Santoyo A, del Pozo Yauner L, Kornhauser A, Fernández-Velasco DA 2002. Structure and inactivation of triosephosphate isomerase from *Entamoeba histolytica*. *J Mol Biol 322*: 669-675.

Rogerson GW, Gutteridge WE 1980. Catabolic metabolism in *Trypanosoma cruzi*. *Int J Parasitol 10*: 131-135.

Sanz-Rodriguez CE, Concepcion JL, Pekerar S, Oldfield E, Urbina JA 2007. Bisphosphonates as inhibitors of *Trypanosoma cruzi* hexokinase: kinetic and metabolic studies. *J Biol Chem 282*: 12377-12387.

Schmunis GA 2007. Epidemiology of Chagas disease in non-endemic countries: the role of international migration. *Mem Inst Oswaldo Cruz 102*: 75-85.

Sgraja T, Smith TK, Hunter WN 2007. Structure, substrate recognition and reactivity of *Leishmania major* mevalonate kinase. *BMC Struct Biol* 7: 20.

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

Silber AM, Rojas RL, Urias U, Colli W, Alves MJ 2006. Biochemical characterization of the glutamate transport in *Trypanosoma cruzi*. *Int J Parasitol 36*: 157-163.

Stephens JL, Lee SH, Paul KS, Englund PT 2007. Mitochondrial fatty acid synthesis in *Trypanosoma brucei*. *J Biol Chem 282*: 4427-4436.

Stern AL, Burgos E, Salmon L, Cazzulo JJ 2007. Ribose 5-phosphate isomerase type B from *Trypanosoma cruzi*: kinetic properties and site-directed mutagenesis reveal information about the reaction mechanism. *Biochem J 401*: 279-285.

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

Tarleton RL 2007. Immune system recognition of *Trypanosoma cruzi*. *Curr Opin Immunol 19*: 430-434.

Taylor MB, Gutteridge WE 1987. *Trypanosoma cruzi*: subcellular distribution of glycolytic and some related enzymes of epimastigotes. *Expt Parasitol 63*: 84-97.

Tetaud E, Bringaud F, Chabas S, Barrett MP, Baltz T 1994. Characterization of glucose transport and cloning of a hexose transporter gene in *Trypanosoma cruzi*. *Proc Natl Acad Sci USA 91*: 8278-8282.

Tielens AG, Van Hellemond JJ 1998. Differences in energy metabolism between trypanosomatidae. *Parasitol Today 14*: 265-272.

Trapani S, Linss J, Goldenberg S, Fischer H, Craievich AF, Oliva G 2001. Crystal structure of the dimeric phosphoenolpyruvate carboxykinase (PEPCK) from *Trypanosoma cruzi* at 2 A resolution. *J Mol Biol 313*: 1059-1072.

Urbina JA 1997. Lipid biosynthesis pathways as chemotherapeutic targets in kinetoplastid parasites. *Parasitology 114*: S91-S99.

Urbina JA, Crespo A 1984. Regulation of energy metabolism in *Trypanosoma (Schizotrypanum) cruzi* epimastigotes. I. Hexokinase and phosphofructokinase. *Mol Biochem Parasitol 11*: 225-239.

Urbina JA, Docampo R 2003. Specific chemotherapy of Chagas disease: controversies and advances. *Trends Parasitol 19*: 495-501.

Urbina JA, Machin I, Jurado L 1993. The limitations of paradigms: studies on the intermediary metabolism of *Trypanosoma cruzi*. *Biol Res 26*: 81-88.

Urbina JA, Osorno CE, Rojas A 1990. Inhibition of phosphoenolpyruvate carboxykinase from *Trypanosoma (Schizotrypanum) cruzi* epimastigotes by 3-mercaptopicolinic acid: *in vitro* and *in vivo* studies. *Arch Biochem Biophys 282*: 91-99.

van Hellemond JJ, Opperdoes FR, Tielens AG 2005. The extraordinary mitochondrion and unusual citric acid cycle in *Trypanosoma brucei*. *Biochem Soc Trans 33*: 967-971.

van Hellemond JJ, Tielens AG 2006. Adaptations in the lipid metabolism of the protozoan parasite *Trypanosoma brucei*. *FEBS Lett 580*: 5552-5558.

van Weelden SW, van Hellemond JJ, Opperdoes FR, Tielens AG 2005. New functions for parts of the Krebs cycle in procyclic *Trypanosoma brucei*, a cycle not operating as a cycle. *J Biol Chem 280*: 12451-12460.

Von Brand T, Agosin M 1955. The utilization of Krebs cycle intermediates by the culture forms of *Trypanosoma cruzi* and *Leishmania tropica*. *J Infect Dis 97*: 274-279.

Wiemer EA, IJlst L, van Roy J, Wanders RJ, Opperdoes FR 1996. Identification of 2-enoyl coenzyme A hydratase and NADP(+)-dependent 3-hydroxyacyl-CoA dehydrogenase activity in glycosomes of procyclic *Trypanosoma brucei*. *Mol Biochem Parasitol 82*: 107-111.

Wiesner J, Jomaa H 2007. Isoprenoid biosynthesis of the apicoplast as drug target. *Curr Drug Targets 8*: 3-13.

# Capítulo IV – Análise comparativa e modelagem estrutural de proteínas análogas

Neste capítulo, foi aplicada uma série de técnicas de modelagem comparativa para a obtenção de modelos tridimensionais para algumas proteínas preditas do genoma de *T. cruzi* e comparação com modelos estruturais de proteínas de *Homo sapiens*. Essas proteínas foram originadas a partir da comparação de análogos funcionais entre estes dois organismos, na tentativa de buscar moléculas estruturalmente distintas com a finalidade de encontrar alvos moleculares para o desenvolvimento de novas drogas contra o parasita. Estes alvos potenciais podem ser posteriormente refinados para o desenvolvimento de novas abordagens para o tratamento da doença de Chagas.

Os dados provenientes destas análises estão no manuscrito "Structural Modelling and Comparative Analysis of Homologous, Analogous and Specific Proteins from *Trypanosoma cruzi* versus *Homo sapiens*: Putative Drug Targets for Chagas' Disease Treatment", submetido para a revista *BMC Genomics*.

# Structural Modelling and Comparative Analysis of Homologous, Analogous and Specific Proteins from *Trypanosoma cruzi* versus *Homo sapiens*: Putative Drug Targets for Chagas' Disease Treatment

Priscila V. S. Z. Capriles*[1,4], Ana Carolina R. Guimarães*[2,4], Thomas Dan Otto[2,3], Antonio B. de Miranda[2], Laurent E. Dardenne[1] and Wim M. Degrave[2]

[1]Grupo de Modelagem Molecular de Sistemas Biológicos - Laboratório Nacional de Computação Científica - LNCC/MCT - Petrópolis, CEP 25651-075 - Brazil
[2]Laboratório de Genômica Funcional e Bioinformática - Instituto Oswaldo Cruz - IOC/FIOCRUZ - Rio de Janeiro, CEP 21045-900 - Brazil
[3]Parasite Genomics - Wellcome Trust Sanger Institute - Wellcome Trust Genome Campus - Cambridge, CB10 1SA - United Kingdom

Email: Priscila V. S. Z. Capriles*- capriles@lncc.br; Ana Carolina R. Guimarães*- carol@fiocruz.br; Thomas Dan Otto - otto@fiocruz.br; Antonio B. de Miranda - antonio@fiocruz.br; Laurent E. Dardenne - dardenne@lncc.br; Wim M. Degrave - wdegrave@fiocruz.br;

*Corresponding author

[4]Both authors contributed equally to this work and should be considered co-first.

## Abstract

**Background:** *Trypanosoma cruzi* is the etiological agent of Chagas' disease, an endemic infection that causes thousands of deaths every year in Latin America. Therapeutic options remain inefficient, demanding the search for new drugs and/or new molecular targets. Such efforts can focus on proteins that are specific for the parasite, but analogous enzymes and enzymes with a three-dimensional (3D) structure sufficiently different from the corresponding host proteins may represent equally interesting targets. To find these targets we used the workflows MHOLline and AnEnΠ obtaining the 3D models from homologous, analogous and specific proteins from *Trypanosoma cruzi* versus *Homo sapiens*.

**Results:** We applied genome wide comparative modelling techniques to obtain 3D models for 3,286 predicted proteins of *T. cruzi*. In combination with comparative genome analysis to *Homo sapiens*, we could identify a subset of 397 enzyme sequences of which 356 are homologous, 3 analogous and 38 specific for the parasite.

**Conclusions:** In this work, we presented a set of 397 enzyme models of *T. cruzi* that can constitute potential structure-based drug targets to be investigated for the development of new strategies to fight Chagas' disease. The strategies presented here support the concept of structural analysis together with protein functional analysis could be an interesting computational methodology to detect potential targets for structure-based rational drug design. For example, the analogous enzymes 2,4-dienoyl-CoA reductase (EC 1.3.1.34) and triacylglycerol lipase (EC 3.1.1.3) were identified as new potential molecular targets.

---

## Background

Chagas' disease constitutes a significant health and socio-economic problem in most of Central and South America and Mexico [1, 2]. About 18 million people are infected resulting in an estimated 21,000 deaths per year (WHO, 2002). Cases have also been described in Canada, United States [3–5], Europe and Australia [6–8].

Hundred years after the discovery of Chagas' disease, caused by the haemoflagellate protozoan *Trypanosoma cruzi*, there are still no appropriate therapies that lead to consistent cure in the chronic phase of the disease. The incidence and death rates, the toxicity of the current drugs benznidazol and nifutimox, and the parasite's ability to develop drug resistance [9, 10], all reinforce the importance of developing new effective chemotherapies [11]. The analysis of the *T. cruzi* genome [12] opens new opportunities to develop more effective and less toxic drugs against the parasite. In biological systems, therapeutic agents can interact with proteins, polysaccharides, lipids and nucleic acids. However, protein inhibitors, particularly enzyme inhibitors, comprise about 47% of all drugs against pharmacological targets with commercial interest [13].

Metabolic pathways that are common to many diverse organisms are mostly made up of enzymatic reactions that are catalysed by conserved proteins. Enzymes performing similar chemical reactions usually share similar structures. However, analogous enzymes have little or no structural similarity, while sharing the same catalytic activity, and are thought to be evolutionarily unrelated [14]. *In silico* sequence analysis and comparisons of the primary and secondary structures *per se* cannot prove that two sequences are unrelated from an evolutionary point of view. A common origin can be inferred from protein structure conservation, even when the amino acid sequence homology was completely washed out by divergence. The

possibility of a common origin can only be considered highly unlikely by additional confirmation that two proteins have different three-dimensional (3D) structures [15]. Furthermore, these differences of 3D structures contribute as an important factor in selecting a protein as a potential therapeutic target [16]. During the process of the development of a new drug, often many synthetic compounds or natural products are tested. The efforts to isolate, purify, characterise, and synthesise active compounds and perform pre- and clinical tests takes many years and can cost billions of dollars [17, 18]. When an active compound is discovered, its mechanism of action is often unknown at the start. Structure-based rational drug design intends to accelerate the steps of identification and comprehension of the molecular interactions between receptor and ligand using computational methods [19]. In this context, bioinformatics and molecular modelling tools can play an important role in the identification and structural investigation of molecular targets that are essential for the survival of *T. cruzi*. Indeed, candidate targets must be essential for the parasite's infectivity and/or survival, without affecting the (human) host [20], and inhibitors should be efficient, soluble and bio-available and administrable in an acceptable way, having the potential for a chemotherapeutic development [21].

Using comparative modelling technique, it is possible to obtain protein models accurate enough to be used in structure-based rational drug design studies. Models building based on templates of proteins that had their 3D structure experimentally determined by X-Ray or Nuclear Magnetic Resonance and that share a reasonable degree of sequence identity with the target have been useful for drug design, and they can guide the development of more specific non-natural inhibitors for variants of a given enzyme or receptor [22–24]. Models building based on a low and medium similarity between the target and template sequences can be useful for functional inference and for the design of rational mutagenesis experiments and molecular replacement in crystallography. Thus, structural biology has been helpful in directing target identification and discovery, using high-throughput methods of structure determination, and providing an important tool for initial drug target screening and further optimisation [19].

A high-throughput functional genomics approach has been used to bridge the gap between raw genomic information and the identification of possible viable drug targets using techniques in biochemistry, molecular and cell biology, and bioinformatics [25]. This approach allows a better understanding of the role played by the steps in biological pathways involved in a variety of diseases.

The search for suitable targets for the development of new drugs is usually based on the identification of parasite specific enzymes from metabolic pathways. But data about the frequency and distribution of analogous enzymes suggests that they may represent an untapped resource for such targets, since

analogous enzymes share the same activity but possess different tertiary structures, an interesting attribute for drug development.

In previous studies, the existence of functional analogues was observed in several important steps in the metabolism of *T. cruzi*, such as the energetic [26] and amino acids pathways [27]. These works show enzymes that are analogous to those found in the human host, listed as possible new therapeutic targets to be studied. Other studies involving the analysis of similar enzymes have suggested the presence of several events of independent origin to about 25% of the total enzymatic activity of an organism [28].

In this work, the protein sequences that have been predicted from the *T. cruzi* genome sequence were analysed with the objective to improve the annotation of their putative biological functions, and to model their probable three-dimensional structures. We used a high-throughput computational environment that uses the comparative modelling technique for the 3D protein structure prediction. From the comparison between *T. cruzi* and *Homo sapiens* enzyme sequences, we could identify and model the 3D structure of 356 homologous, 3 analogous and 38 specific *T. cruzi* putative enzymes, that can be investigated as potential drug targets for Chagas' disease treatment.

## Results and Discussion

### Analysis of Enzymatic Functions of *Trypanosoma cruzi* and Construction of Three-Dimensional Models

We intended to perform a comparative analysis of the 3D structures available or that be constructed through comparative modelling, for *T. cruzi* and human enzymes, in order to detect significant differences that can be exploited and justify these enzymes as potential drug targets. As a starting point, we used the *T. cruzi* CL-Brener database (http://tcruzidb.org/tcruzidb/)of predicted proteins, containing 19,607 entries (translated CDS – Coding Sequences). To remove redundant and very similar sequences, an all-against-all BLAST analysis was done and the output was submitted to BioParser [29], and from multiple sequences with more than 95% identity only one member was kept, resulting in a dataset of 12,348 protein sequences. These were submitted to the MHOLline workflow (http://www.mholline.lncc.br) to construct 3D protein structure models by comparative modelling. This analysis resulted in 3,286 models, presented in Table 1, that were classified according to the criterion described in Table 7.

### Inference of Functional Annotation of *Trypanosoma cruzi* Predicted Proteins

We previously reported results [26, 27] on the inference of function to the proteins predicted from the *T. cruzi* CL-Brener genome initiative (http://tcruzidb.org/tcruzidb/) using the annotation module in the

AnEnΠ pipeline [28]. In addition to aforementioned analysis, here we added enzymatic functions specified in the Swiss-Prot that were absent in the KEGG database, in order to increase the number of enzymatic functions to be analysed. It was performed due to the fact that there are enzymatic functions that are not represented in the metabolic pathways described in the KEGG database (*e.g.* prolineracemase - EC 5.1.1.4).

The choice of the cut-off remains a critical point in this procedure and for this reason we investigated different e-values as cut-off ($10e^{-20}$, $10e^{-40}$ and $10e^{-80}$)(Table 2). In order to confer a high degree of reliability to our analysis we adopted the cut-off of $10e^{-80}$ for the next steps.

To establish a good cut-off we should analyse groups of protein families separately and take into account other parameters like coverage, bit-score and identity, but this is not yet available in AnEnΠ. The inferred protein functions of *T. cruzi* were used to find analogy between the parasite sequences and the predicted proteins of *Homo sapiens*.

### Comparison Between *Homo sapiens* and *Trypanosoma cruzi* Enzymatic Functions

Using AnEnΠ, we analysed and compared the predicted protein sequences from *Homo sapiens* and *Trypanosoma cruzi* to establish possible cases of analogy between these two species. For some enzymatic functions, the sequences of *H. sapiens* and *T. cruzi* were allocated in different clusters, representing probable cases of analogy (see the Methods for more details), while sequences allocated in the same cluster were considered homologous. We expect the 3D structures to be dissimilar in the first case, and probably similar in the latter case. This is indeed true in some cases, and it is exemplified in Figure 1. Also, some sequences are specific of *T. cruzi* and absent in *H. sapiens*. The results are summarised in Table 3 and were acquired using as final dataset the 478 entries obtained by the combination of both KEGG and Swiss-Prot databases, considering the complete four-digit EC number.

Figure 1 shows examples of comparison between *T. cruzi* and *H. sapiens* protein structures, using functional classification determined by AnEnΠ. Figure 1(a) presents the structural alignment ($RMSD_{C_\alpha} = 0.65\mathring{A}$) between the *T. cruzi* protein model (yellow), obtained with MHOLline, and the homologous structure (PDB1F14) of L-3-Hydroxiacyl-CoA Dehydrogenase from *Homo sapiens* (blue). The structure of the active site (S137, H158 and N208) of the human protein, according to [30], it is quite similar to the structure of modelled *T. cruzi* protein. Figure 1(b) shows the same *T. cruzi* model (yellow) and the analogous enzyme (PDB1SO8) 3-Hydroxiacyl-CoA Dehydrogenase Type II from *H. sapiens* (green). In this figure, the dissimilarity ($RMSD_{C_\alpha} = 13.81\mathring{A}$) between these two structures is evident. The

$RMSD_{C_\alpha}$ were calculated using Swiss-PDB Viewer (v4.0.1) program [31].

### Functional Classification of Modelled Enzymes

In a next step, we combined the results presented in Table 1 and Table 3, and could identify a set of 397 predicted proteins from *Trypanosoma cruzi* to which an enzymatic function could be assigned with the AnEnΠ tool, and for which a structure model could be obtained using MHOLline. These functions have 93 distinct EC numbers assigned to them, as showed in Table 4.

Table 5 summarises the results of the overall analysis in this work. The modelled proteins associated to an EC number were grouped as follows with regard to the comparison between *T. cruzi* and *H. sapiens*: (i) Homologous enzymes; (ii) Analogous enzymes; (iii) Specific of *T. cruzi* and (iv) Undetermined enzymes – enzymes with conflicting clustering depending on the KEGG or Swiss-Prot database used for inicial clustering. Moreover, these protein sequences were classified according to IUBMB Nomenclature with regard to the first EC number digit. Additionally, they were classified from 1 to 7 according to the MHOLline model quality proposed in the Methods.

## Discussion and Conclusions

Knowledge of the three-dimensional structures of proteins opens the way to accelerate drug discovery [19]. Theoretical predictions of 3D protein structures and protein folding patterns, even on a genome scale, can provide valuable information to infer possible protein function and contribute to identification of potential drug targets [32]. It is believed that evolution tends to conserve functions based more on the preservation of the 3D structure than on that of the primary sequence. The protein folding permits that distant amino acids in the sequence interact in the same spatial region, forming distinct conformational patterns, depending on the ambient conditions. A 3D alignment between structural relatives, even (or mainly) comprising a small number of residues within a protein active site, can be a powerful method to infer function [33].

Using the 19,607 predicted protein sequences from *Trypanosoma cruzi* CL-Brener genome as initial dataset, we produced a non-redundant dataset comprising 12,348 sequences. After, submitting these sequences to the MHOLline workflow, we were able to construct models for 3,286 sequences (26.6% of the total), of which 1,164 models (35.4%) have a "medium to good" to a "very high" quality (presented in Table 1), being adequate to structure-based drug design projects.

It is important to note that there are problems in the processes of assembly and annotation of genomes,

involving for example the quality of the produced sequences, errors derived from automatic gene prediction, the presence of repetitive regions, lack of usage of controlled vocabulary terms (ontology) and propagation of previous annotation errors.

Due to the high repetitive gene content and the heterozygosity of the *T. cruzi* strain at hand, until now the genome of *T. cruzi* has not been completely assembled, and many predicted proteins have unknown or putative functions which hinder the correct identification of proteins and consequently the elucidation of the parasite's metabolism. To minimise some of these problems, we used the AnEnΠ pipeline to annotate the *T. cruzi* genome and to identify enzymatic functions using KEGG and Swiss-Prot databases (Table 2). From the comparison between *T. cruzi* and *Homo sapiens* enzymatic functions, we identified a set of 397 *T. cruzi* modelled sequences, comprising 93 distinct EC numbers (Table 4). Six sequences originally annotated (by GeneDB) as hypothetical proteins could be associated to an enzymatic function by AnEnΠ (more details in Supplementary Material).

An important result of this work was the identification and construction of 3D protein models for three sequences classified as analogous and 38 sequences as specific for *T. cruzi* (listed on Table 6), which can be considered as possibly interesting molecular targets for the development of drugs against Chagas' disease. Among the specific enzymes, we could identify some proteins that are already being studied as drug targets (*e.g.* cruzipain and trypanothione-disulfide reductase). In general, to confirm the potential of these 41 proteins as structure-based drug targets, it is necessary to take into account the importance of the metabolic pathway involved for parasite survival, the existence of possible isoforms and alternative metabolic pathways, data about enzymatic assays and the quality os constructed model for further structural analysis, and other information that could help in the understanding of physico-chemical properties, catalytic sites and pharmacological inhibitors of these proteins. Of course, one should not discard the 356 sequences classified as homologous proteins in relation to *H. sapiens* as, for example, the glyceraldehyde-3-phosphate dehydrogenase [34], an important known drug target.

We have further analysed the models for the *T. cruzi* analogous enzymes (presented in Table 6) 2,4-dienoyl-CoA reductase (EC 1.3.1.34) and triacylglycerol lipase (EC 3.1.1.3), which are involved in the metabolism of lipids. The major aspects of lipid metabolism concern fatty acid oxidation to produce energy, and the synthesis of lipids. Knowledge about the oxidation of fatty acids as a source of ATP for trypanosomatides remains scare. Previous analysis of *T.brucei*, *T.cruzi* and *Leishmania* genomes identified orthologous genes encoding enzymes involved in the $\beta$-oxidation of fatty acids, and this pathway probably occurs both in glycosomes and mitochondria [35].

The oxidation of polyunsaturated fatty acids requires an auxiliary enzyme (2,4-dienoyl-CoA reductase) that removes the double bonds in the fatty acids. This enzyme (combined with enoyl-CoA isomerase) is essential to allow beta-oxidation and consequently energy production for the parasite [36]. It is possible that this reaction occurs in the opposite direction, generating an unsaturation which could be important in the synthesis of a compound produced in the parasite, whenever the parasite requires it in the composition of unsaturated fatty acids. The sequence and structure alignment between the two isoforms of 2,4-dienoyl-CoA reductase from *T. cruzi* suggest that these proteins are paralogous. Figure 2 presents the difference between the primary and tertiary structures of the paralogous enzymes of *T. cruzi* and the 2,4-dienoyl CoA reductase 1 (DECR1 - mitochondrial) and 2,4-dienoyl CoA reductase 2 (DECR2 - peroxisomal) of *H. sapiens*.

The other analogous enzyme, triacylglycerol lipase, converts triacylglycerol and $H_2O$ in diacylglycerol and a carboxylate. This reaction is important to glycerolipid metabolism. [37] showed that the parasite is able to take up by endocytosis LDL cholesterol, a molecule that has triglycerides in it composition, justifying the presence of this enzyme in the parasite. Furthermore, the product of the reaction is composed of diacylglycerol, an important molecule for the synthesis of membrane lipids (phospholipids and glycolipids). Taking into account the presented results and the importance of the two enzymatic activities in the oxidation of polyunsaturated fatty acids and glycerolipids metabolism, these analogous enzymes might be an interesting choice for further studies for drug development against Chagas' disease.

The most widely used paradigm in the search of new drug targets is to look for pathogen specific molecules, where the development of ligands that result in the inactivation of the function but without affecting the host [20]. However, data on the frequency and distribution of analogous enzymes suggest that these enzymes should be studied as additional targets since it is expected that they share the same enzymatic activity with sufficiently different tertiary structures, a prerequisite for the development of drugs [20].

The results presented in this work corroborate the idea that structural analysis could be an attractive computational methodology for predicting protein functions [38]. The combination of MHOLline workflow with AnEnΠ pipeline was effective to infer protein function and to detect and construct structure models of proteins in high-throughput analysis. Thus, we were able to identify a list of *T. cruzi* specific or analogous enzymes that can be considered as target candidates suitable to be used in further structure-based drug design projects against Chagas' disease (a complete list of proteins is provided in Supplementary Material).

## Methods

### Datasets

In this work, we used the dataset composed of 19,607 predicted protein sequences from the *Trypanosoma cruzi* genome (CL-Brener strain). This dataset was obtained from TcruziDB (http://tcruzidb.org/common/downloads/release-5.0/Tcruzi/TcruziAnnotatedProtein.fas - version 5.0). AnEnΠ is a tool for identification and annotation of analogous enzymes [28]. We have used the dataset contained in AnEnΠ (Analogous Enzyme Pipeline), which was obtained from KEGG (Kyoto Encyclopedia of Genes and Genomes) database (from ftp://ftp.genome.ad.jp/pub/kegg/ of December, 2006) [39]. To increase the number of identifiable enzymatic functions by AnEnΠ, we incorporated data from Swiss-Prot [40] (from http://www.expasy.org/sprot/ of May, 2008), resulting in a final dataset composed by 478 four-digit EC numbers. Each *T. cruzi* enzyme function obtained (considering the complete four-digit EC number) was compared with the original genome function annotation list from GeneDB (from http://www.genedb.org/ of October, 2007).

The structures used as templates to provide 3D models of predicted proteins from *T. cruzi* were obtained from the Protein Data Bank (PDB) (44,700 sequences from ftp://ftp.rcsb.org/pub/pdb/derived_data/ of December, 2006). These models were constructed by comparative modelling method using the workflow MHOLline, as described in the Methods.

### High-Throughput Comparative Modelling

To construct 3D structure models of the predicted proteins from *T. cruzi* genome we used the MHOLline software (http://www.lncc.br). It is a biological workflow that combines a specific set of programs for automated protein structure prediction, detection of transmenbrane regions, and EC number association. It extracts distinct and valuable structural information about protein sequences even in large-scale genome annotation projects

MHOLline uses the HMMTOP program to identify transmembrane regions. BLAST algorithm is used for template structure identification by performing searches against the Protein Data Bank [41]. Refinements in the template search – a key step for the model construction – were implemented with the development of a program called BATS (Blast Automatic Targeting for Structures). BATS identifies the sequences where comparative modelling techniques can be applied, by choosing template sequences from the BLAST output file using their scores, expectation values, identity and sequence similarity as criteria. It also consider the number of gaps and the alignment coverage.

BATS also selects the best template for 3D model construction and generates the files for the automated alignment used by the Modeller program [42]. Statistically, there is no relevant correlation between the number of templates used during model construction and the overall quality of a model [43]. The generated models are evaluated by stereochemical quality using the Procheck program [44]. In summary, for each submitted sequence, MHOLline generates and aggregates structural information, returns a 3D model, a Ramachandran plot and comments about structure quality and enzymatic function.

*Sequence Filtering and Generation of Distinct Quality Protein Models*

To exclude possible redundant and very similar sequences, an all-against-all BLAST analysis was performed in the dataset composed of all *T. cruzi* translated CDS, using the BLOSUM62 matrix and an e-value $\leq 10e^{-5}$ as cutoff. The result was automatically filtered by identity ($\leq 95\%$) using the BioParser tool [29]. This non redundant dataset of *T. cruzi* was submitted to the MHOLline workflow to construct the 3D protein models. Sequences were locally aligned by MHOLline (using BLASTP) with protein sequences from PDB using an e-value $\leq 10e^{-5}$. The MHOLline program filtered the new set of aligned sequences with the BATS program and the Filters tool, and it constructed the protein structure models using the Modeller program. Table 7 displays the criteria used for the classification of the obtained models.

*Trypanosoma cruzi* **Protein Function Inference**

AnEnΠ uses the similarity score of BLASTP pair wise comparisons between all proteins included in a previously determined group to assign these proteins to separate clusters for each enzymatic function (EC numbers). Enzymes inside a cluster are considered homologous, while enzymes in different clusters (of the same group/function) are considered analogous.

With the purpose of annotation and identification, users can perform similarity searches by BLASTP. In this case, the database is composed of the sequences belonging to each cluster. In this study, AnEnΠ was used for the identification of predicted proteins of *Trypanosoma cruzi* using different e-values as cutoff ($10e^{-20}$, $10e^{-40}$ and $10e^{-80}$).

## Authors contributions

PVSZC and ACRG project the studies, performed all computational analysis, and drafted the manuscript. TDO performed some computational analysis related to analogies. ABM, LED and WMD planned and supervised the study. All authors read and approved the final manuscript.

## Acknowledgements

We thank Shaila Rössle from Department of Geo- and Enviroment Sciences - University of Munich (Ludwig - Maximilians - Muenchen - Germany) and Damásio A. A. Ferreira to introduce us to MHOLline workflow, and Marcos Catanho from Laboratório de Genômica Funcional e Bioinformática - IOC/FIOCRUZ (Rio de Janeiro - RJ - Brazil) to help us with BioParser. We thank the Brazilian National Council of Research (CNPq) and the FAPERJ Foundation for support this work. Contract grants: E26/170.648/2004, E26/102.443/2009, CNPq/IM-INOFAR 420.015/2005-1, CNPq/MS-SCTIE-DECIT 41.0544/2006-0, CNPq/MS-SCTIE-DECIT 409078/2006-9, CNPq/MCT 15/2007-Universal.

## References

1. Dias JC, Machado EM, Fernandes AL, Vinhaes MC: **General situation and perspectives of chagas disease in Northeastern Region, Brazil**. *Cadernos de Saúde Pública* 2000, **16**(2):13–34.
2. Kirchhoff LV, Paredes P, Lomeli-Guerrero A, Paredes-Espinoza M, Ron-Guerrero CS, Delgado-Mejia M, Pena-Munoz JG: **Transfusion-associated Chagas disease (American trypanosomiasis) in Mexico: implications for transfusion medicine in the United States**. *Transfusion* 2006, **46**(2):298–304.
3. Leiby DA, Herron RM, Read EJ, Lenes BA, Stumpf RJ: *Trypanosoma cruzi* **in Los Angeles and Miami blood donors: impact of evolving donor demographics on seroprevalence an d implications for transfusion transmission**. *Transfusion* 2002, **42**(5):549–555.
4. Beard CB, Pye G, Steurer FJ, Rodriguez R, Campman R, Peterson AT, Ramsey J, Wirtz RA, Robinson LE: **Chagas Disease in a Domestic Transmission Cycle in Southern Texas, USA**. *Emerging Infectious Diseases* 2003, **9**:103–105.
5. Milei J, Guerri-Guttenberg RA, Grana DR, Storino R: **Prognostic impact of Chagas disease in the United States**. *American Heart Journal* 2008, **157**:22–29.
6. Reesink HW: **European Strategies Against the Parasite Transfusion Risk**. *Transfusion Clinique et Biologique* 2005, **12**:1–4.
7. Kerleguer A, Massard S, Janus G, Joussemet M: **Chagas disease: screening tests evaluation in a blood military center, prevalence in the French Army**. *Pathologie Biologie* 2007, **55**:534–538.
8. Schmunis GA: **Epidemiology of Chagas disease in non-endemic countries: the role of international migration**. *Memórias do Instituto Oswaldo Cruz* 2007, **102**(Suppl. I):75–85.
9. Gelb MH, Hol WGJ: **Drugs to Combat Tropical Protozoan Parasites**. *Science* 2002, **297**(19):343–344.
10. Wilkinson SR, Taylor MC, Horn D, Kelly JM, Cheeseman I: **A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes**. *PNAS* 2008, **105**(13):5022–5027.
11. Coura JR: **Chagas disease: what is known and what is needed - A background article**. *Memórias do Instituto Oswaldo Cruz* 2007, **102**(Suppl. I):113–122.
12. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G: **The Genome Sequence of** *Trypanosoma cruzi***, Etiologic Agent of Chagas Disease**. *Science* 2005, **309**(15):409–415.
13. Hopkins AL, Groom CR: **The druggable genome**. *Nature Reviews* 2002, **1**:727–730.
14. Fitch WM: **Distinguishing homologous from analogous proteins**. *Systematic Zoology* 1970, **19**(2):99–113.
15. Galperin MY, Walker DR, Koonin EV: **Analogous enzymes: Independent inventions in enzyme evolution**. *Genome Research* 1998, **8**:779–790.
16. Doolittle RF: **Convergent evolution: the need to be explicit**. *Trends in Biochemical Sciences* 1994, **19**:15–18.

17. Kola I, Landis J: **Can the pharmaceutical industry reduce attrition rates?** *Nature Reviews* 2004, **3**:711–715.

18. Adams CP, Brantner VV: **Estimating The Cost Of New Drug Development: Is It Really $802 Million?** *Health Tracking* 2006, **25**(2):420–428.

19. Congreve M, Murray CW, Blundell TL: **Structural biology and drug discovery**. *Drug Discovery Today* 2005, **10**(13):895–907.

20. Karp PD, Krummenacker M, Paley S, Wagg J: **Integrated pathway-genome databases and their role in drug discovery**. *Trends in Biotechnology* 1999, **17**(7):275–281.

21. Kramer R, Cohen D: **Functional genomics to new drug targets**. *Nature Reviews Drug Discovery* 2004, **3**(11):965–972.

22. Sánchez R, Pieper U, Melo F, Eswar N, Martí-Renom MA, Madhusudhan MS, Mirkovi NC, Sali A: **Protein structure modeling for structural genomics**. *Nature Structural Biology* 2000, **7**:986–990.

23. Hillisch A, Pineda LF, Hilgenfeld R: **Utility of homology models in the drug discovery process**. *Drug Discovery Today* 2004, **9**(15):659–669.

24. Cavasotto CN, Phatak SS: **Homology modeling in drug discovery: current trends and applications**. *Drug Discovery Today* 2009, **4**(13-14):676–683.

25. Lindsay MA: **Target discovery**. *Nature Reviews Drug Discovery* 2003, **2**:831–838.

26. Alves-Ferreira M, Guimarães ACR, Capriles PVSZ, Dardenne LE, Degrave WM: **A new approach for potential drug target discovery through in silico metabolic pathway analysis using** *Trypanosoma cruzi* **genome information**. *Mem Inst Oswaldo Cruz* 2009, **104**(8):1100–1110.

27. Guimarães ACR, Otto TD, Alves-Ferreira M, Miranda AB, Degrave WM: **In silico reconstruction of the amino acid metabolic pathways of** *Trypanosoma cruzi*. *Genetics and Molecular Research* 2008, **7**(3):872–882.

28. Otto T, Guimarães A, Degrave W, Miranda A: **AnEnPi: identification and annotation of analogous enzymes**. *BMC Bioinformatics* 2008, **9**:544.

29. Catanho M, Mascarenhas D, Degrave W, de Miranda AB: **BioParser: A tool for processing of sequence similarity analysis reports**. *Applied Bioinformatics* 2006, **5**:49–53.

30. Barycki JJ, O'Brien LK, Strauss AW, Banaszak LJ: **Sequestration of the Active Site by Interdomain Shifting**. *The Journal of Biological Chemistry* 2000, **275**(35):27186–27196.

31. Guex N, Peitsch MC: **SWISS-MODEL and Swiss-Pdb Viewer: An environment for comparative protein modeling**. *Electrophoresis* 1997, **18**:2714.

32. Cherkasov A, Sui SJH, Brunham RC, Jones SJ: **Structural characterization of genomes by large scale sequence-structure threading: application of reliability analysis in structural genomics**. *BMC Bioinformatics* 2004, **5**(37):1–16.

33. Marsden RL, Maibaum DLM, Yeats C, Orengoe CA: **Comprehensive genome analysis of 203 genomes provides structural genomics with new insights into protein family space**. *Nucleic Acids Research* 2006, **34**(3):1066–1080.

34. Freitas RF, Prokopczyk IM, Zottis A, Oliva G, Andricopulo AD, Trevisan MTS, Vilegas W, Silva MGV, Montanari CA: **Discovery of novel** *Trypanosoma cruzi* **glyceraldehyde-3-phosphate dehydrogenase inhibitors**. *Bioorganic & Medicinal Chemistry* 2009, **17**(6):2476–2482. [Special Issue: Natural Products in Medicinal Chemistry].

35. van Hellemond JJ, Tielens AG: **Adaptations in the lipid metabolism of the protozoan parasite** *Trypanosoma brucei*. *FEBS Letters* 2006, **580**(23):5552–5558.

36. Hubbard PA, Liang X, Schulz H, Kim JJP: **The Crystal Structure and Reaction Mechanism of** *E. coli* **2, 4 - Dienoyl CoA Reductase**. *Journal of Biological Chemistry* 2003, **278**(39):37553–37560.

37. Soares MJ, de Souza W: **Endocytosis of gold-labeled proteins and LDL by** *Trypanosoma cruzi*. *Parasitology Research* 1991, **77**:461–468.

38. Lee D, Redfern O, Orengo C: **Predicting protein function from sequence and structure**. *Nature Reviews Molecular Cell Biology* 2007, **8**:995–1005.

39. Kanehisa M, Goto S, Hattori M, Aoki-Hinoshita K, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M: **From genomics to chemical genomics: new developments in KEGG**. *Nucleic Acids Research* 2006, **34**(D):354–357.

40. Bairoch A, Boeckmann B, Ferro S, Gasteiger E: **Swiss-Prot: juggling between evolution and stability**. *Briefings in Bioinformatics* 2004, **5**:39–55.

41. Westbrook J, Feng Z, Jain S, Bhat TN, Thanki N, Ravichandran V, Gilliland GL, Bluhm WF, Weissig H, Greer DS, Bourne PE, Berman HM: **The Protein Data Bank: unifying the archive**. *Nucleic Acids Research* 2002, **30**:245–248.

42. Sánchez R, Sali A: **Evaluation of comparative protein structure modeling by MODELLER-3**. *PROTEINS: Structure, Function, and Genetics* 1997, **29**(S1):50–58.

43. Peitsch MC: **About the use of protein models**. *Bioinformatics* 2002, **18**:934–938.

44. Laskowski RA, MacArthur MW, Moss DS, Thornton JM: **PROCHECK: a program to check the stereochemical quality of protein structures**. *Journal of Applied Crystallography* 1993, **26**(2):283–291.

45. Lustbader JW, Cirilli M, Lin C, Xu HW, Takuma K, Wang N, Caspersen C, Chen X, Pollak S, Chaney M, Trinchese F, Liu S, Gunn-Moore F, Lue LF, Walker DG, Kuppusamy P, Zewier ZL, Arancio O, Stern D, Yan SS, Wu H: **ABAD Directly Links A$\beta$ to Mitochondrial Toxicity in Alzheimer's Disease**. *Science* 2004, **304**:448–452.

46. Humphrey W, Dalke A, Schulten K: **VMD – Visual Molecular Dynamics**. *Journal of Molecular Graphics* 1996, **14**:33–38.

47. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Research* 1997, **24**:4876–4882.

48. Alphey MS, Yu W, Byers E, Li D, Hunter WN: **Structure and Reactivity of Human Mitochondrial 2,4-Dienoyl-CoA Reductase: Enzyme-Ligand Interactions in a Distinctive Short-Chain Reductase Active Site**. *The Journal of Biological Chemistry* 2005, **280**(4):3068–3077.

## Figures

**Figure 1 - Structural comparison between a medium to high quality model of 3-Hydroxiacyl-CoA Dehydrogenase from *Trypanosoma cruzi* and one homologous and one analogous structures from PDB (classified according to the AnEn$\Pi$ pipeline).**

File format: PNG

Title: Figure1.png

Subtitle: 1(a): structural alignment of the *T. cruzi* (Tc00.1047053510105.240) protein model (yellow) and a homologous protein (PDB1F14) from *Homo sapiens* (blue), detailing its active site residues S137, H158 and N208 according to [30]. The alignment was performed by Swiss-PDB Viewer (v4.0.1) program [31]. 1(b): structure of *T. cruzi* (Tc00.1047053510105.240) model (yellow) and the analogous enzyme (PDB1SO8) from *H. sapiens* (green). In detail, it is presented the putative active site residues S154, H175 and N225 of *T. cruzi* protein (yellow), inferred by the alignment in Figure 1(a), and the *H. sapiens* (green) active site (S155, Y168 and K172) from [45]. The images were generated using VMD (Visual Molecular Dynamics - v1.8.6) software [46]

Figure 1

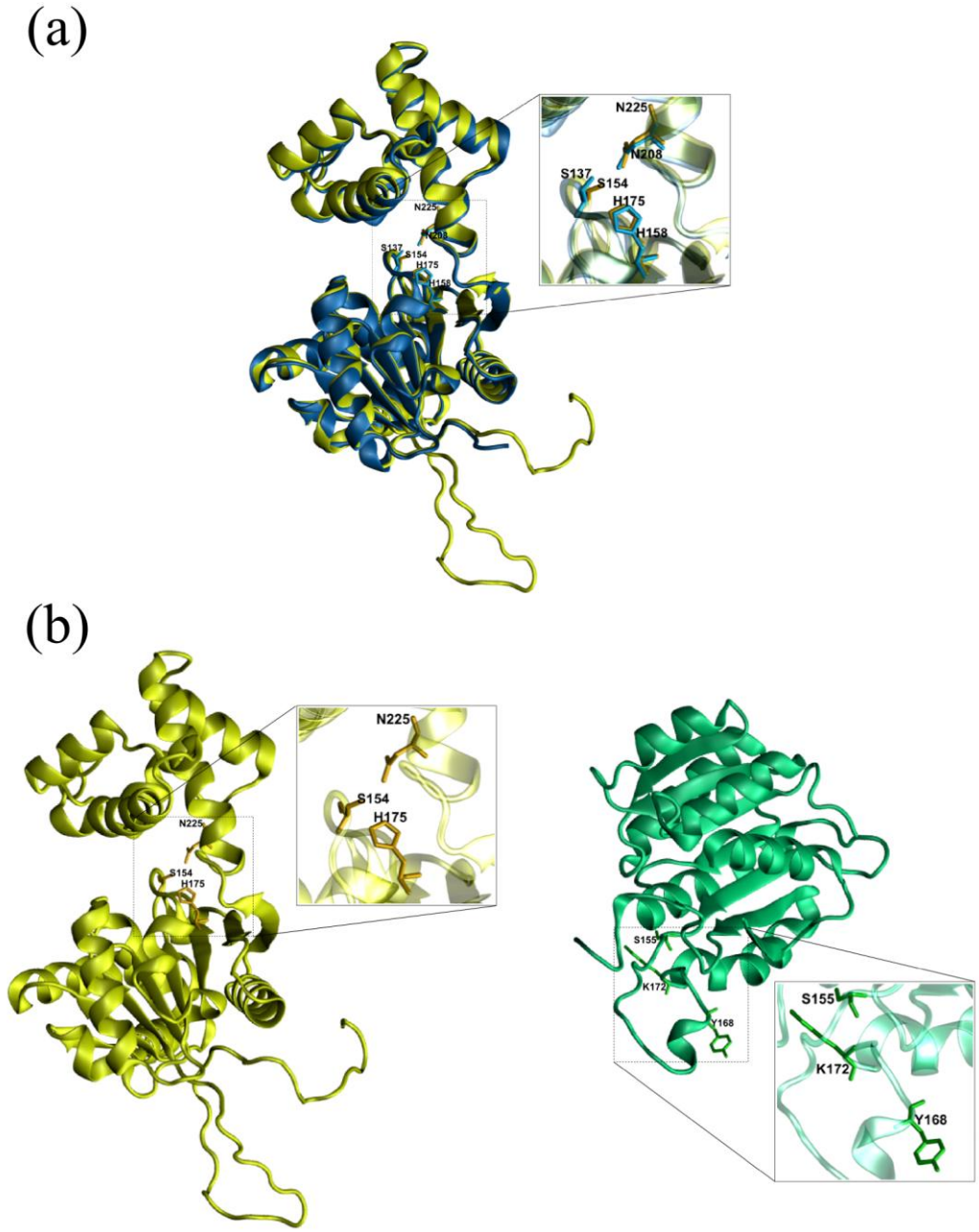

(a)
N225
N208
S137
S154
H175
H158
(b)
N225
S154
H175
S155
K172
Y168

**Figure 2 - Structural and sequence comparison between 2,4-dienoyl CoA reductase (DECR) from *Trypanosoma cruzi* and *Homo sapiens*, analogous enzymes.**

File format: PNG

Title: Figure2.png

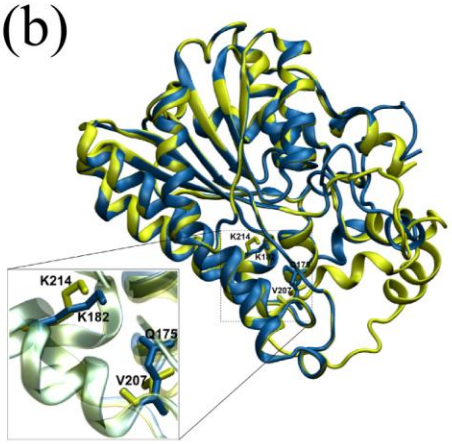
Subtitle: 2(a): sequence alignment between putative paralogous DECR enzymes from *T. cruzi* and mitochondrial DECR1 and peroxisomal DECR2 enzymes from *H. sapiens*. The alignment was performed by ClustalX (v1.83) program [47]. 2(b): structural alignment between the *H. sapiens* DECR1 (reconstructed PDB:1W6U) (yellow) and DECR2 model (blue), constructed using PDB:1W6U as template. In detail, it is presented the active site residues V207 and K214 of DECR1 (yellow), according to [48], and Q175 and K182 of DECR2 (blue). 2(c): structural alignment between DECR enzymes of *T. cruzi*. In detail, it is presented the putative active site constituted by Y175 and H261 of Tc00.1047053509941.100 (yellow) and, G180 and H269 of Tc00.1047053510303.210 (blue). The active sites of *T. cruzi* DECR were inferred by their structural alignment (not presented) with the DECR protein (PDB:1PS9) from *Escherichia coli*, used as template.Its active site residues Y166 and H252 is described by [36]. The alignments were performed by Swiss-PDB Viewer (v4.0.1) program [31] and the images were generated using VMD (Visual Molecular Dynamics - v1.8.6) software [46]

Figure 2

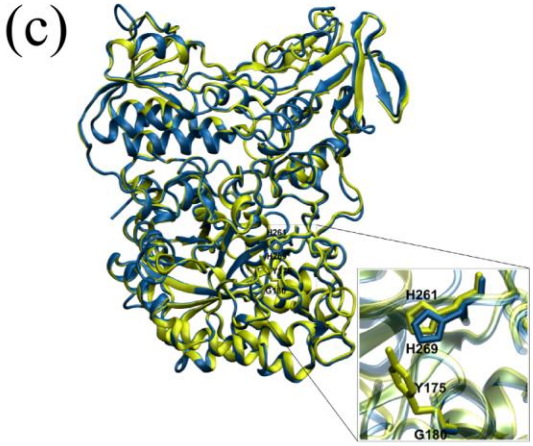
(a)

```
hsa_26063|DECR2 peroxisomal                                    MAQPPPDVEGDDCLPAYRHLFCPDLLRDKVAFITGGGSGIGFRIAEIFMR 50
hsa_1666|DECR1 mitochondrial       MKLPARVFFTLGSRLPCGLAPRRFFSYGTKILYQNTEALQSKFFSPLQKAMLPPNSFQGKVAFITGGGTGLGKGMTTLLSS 81
Tc00.1047053510303.210             MTGKISPLFLPICIGAKRHLELPNRFLMLPMYLNMEQELDLWSDAHMDAMGSFYAERARHGAKLIVCGGLGVSRIGKWRIDS 82
Tc00.1047053509941.100             MSLMYEKMFSPLSLGFT   TLRNRIVMG    SMHTGLEDLKKDRIERMCDFYGERAKGGCALIVTGGFSPNFEGKLNPFA 75
                                                                    .         :  :  :    :*. ** . .

hsa_26063|DECR2 peroxisomal        HGCHTVIASRSLPRVLTAARKLAGATGRRCLPLSMDVRAPPAVMAA                                VDQ 99
hsa_1666|DECR1 mitochondrial       LGAQCVIASRKMDVLKATAEQISSQTGNKVHAIQCDVRDPDMVQNT                                VSE 130
Tc00.1047053510303.210             MALGTYDAAKAFSRVTHAVHREGGAILAQAFHPGRAARRRWFISATSTPSSVQPFR  DAHPWRIPGLLVEYIVSEHARFAR 162
Tc00.1047053509941.100             PCISSKSHALEYKPVADAVHREGGKIVMQILHAGRYAYSPLCVAPSGIPSPIWPFKGYHLRPIALPKVWINKTIKDFARAAA 157
                                          :     :  :..: ..    :        .     :  :

hsa_26063|DECR2 peroxisomal        ALKEFGRIDILINCAAG    NFLCPAGALSFNAFK                                TVMDIDTSGTFNV 144
hsa_1666|DECR1 mitochondrial       LIKVAGHPNIVINNAAG    NFISPTERLSPNAWK                                TITDIVLNGTAFV 175
Tc00.1047053510303.210             LAEAAGFDGIEIPVSEGGLLHNFLSTAVNDRSDYYGRSLEGRLEATLRVLDAVRRSLVEPEKFLLSVRLCLHDLKVGGTPMT 244
Tc00.1047053509941.100             LVKEAGFDGVEIMASEGYLLNEFIVKHTNKRTDEYGGSYENRMRFPLEVLRGVRNAVG  KEFIIIFRLSMLDLIPNGSTHE 237
                                     :  *  .: *  : *    :*:        : :                              : *:  .*:

hsa_26063|DECR2 peroxisomal        SRVLYEKFFRDHG GVIVNITATLGNRGQALQVHAGSAKAAVDAMTRHLAVEWGPQNIRVNSLAPGPISGTEGLRRLGGP   223
hsa_1666|DECR1 mitochondrial       TLEIGKQLIKAQKGAAFLSITTIYAETGSGFVVPSASAKAGVEAMSKSLAAEWGKYGMRFNVIQPGPIKTKGAFSRLDPT   255
Tc00.1047053510303.210             ETLVAAEAIARFGGVDILSTSVGMHDSP VQTMASYVPQAVFARCVQKVRERLQSHGLNVPVVATHRVHSIAVAERLVREGV 325
Tc00.1047053509941.100             EVCKLAEEVANSG ADIINTGIGWHEAR VPTIATSVPRAGFTWVTAACRKYLRSKGISIPFMTANRVNHPDVIERVLSK D 316
                                         : .       ::.      :      : :  .:* .            .: .  : .  :       *:

hsa_26063|DECR2 peroxisomal                                       QASLSTKVTASPLQRLGNKTEIAHSVLYLASPLASYVTG    262
hsa_1666|DECR1 mitochondrial                                      GTFEKEMIGRIPCGRLGTVEELANLAAFLCSDYASWING    294
Tc00.1047053510303.210             CDIVGLARPLLSDPQMIRKAEEGHEDAIIPCIACN HCVNRLYKHQRITCALNPTSGYERERQWAATRYKKNVAVIGAGAAG 406
Tc00.1047053509941.100             ADLVAMARPFLSDPFFVKKTEEGNVHNINICIACNQACLDHIFKGQISSCLVNPTACHEKERAVIPTKQKKRIAVIGGGPSG 398
                                                                                  *        .  *    *

hsa_26063|DECR2 peroxisomal           AVLVADGGAWLTFP      NGVKGLPDFASFSAKL                                   292
hsa_1666|DECR1 mitochondrial          AVIKFDGGEEVLIS      GEFNDLRKVTKEQMDTIEELIRKTKGS                        335
Tc00.1047053510303.210             VTCALTLWRRGHDVTLFEKSNFIGGQLNLAKVVPGKESYQEVLEYWTRQLRQSSINVRLGVEFVREDMSRNHQFFHALVLCC 488
Tc00.1047053509941.100             ASCALVLAQRGHEVTIYEKEDCLGGQFNLAKKIPGKAEYESSIRYWTHSLT TLNNVKLHLKTEVDASLLNKTAYDEVIVAT 479
                                      *:     *  : :       .   .* .    . .

hsa_26063|DECR2 peroxisomal                                                                                   292
hsa_1666|DECR1 mitochondrial                                                                                  335
Tc00.1047053510303.210             GSVPRPATSHQIPGTSECPMVVPFEKILNGSVKAGRRVVIIGNGAISHDVASYLLHDPRVS       RSVEAYCDEWGVNL 563
Tc00.1047053509941.100             GCLPKPIKNSVIKGVEGLKNVFSYVDVLSGKAHVGDRVAIIGGGGIGFDVAYFLVENHDLLGCTGYQKENLEQFAKRWGIDL 561

hsa_26063|DECR2 peroxisomal                                                                                   292
hsa_1666|DECR1 mitochondrial                                                                                  335
Tc00.1047053510303.210             EDGTLDPHAAERAPRNNRDVVLFNKAD  KDADLSRGKGWTQKLNIRNHGGTIIKHGLIENIDRNGVHISMVAPDSRKYFVE 643
Tc00.1047053509941.100             EIASPGGLRKPNIGQSARAVTMLQRSTGKMGRRLGATTGWIHRLELRSHNVKQIIGVTYDSFDGKTLSYSIGD   KKHNLD 640

hsa_26063|DECR2 peroxisomal                                                                                292
hsa_1666|DECR1 mitochondrial                                                                               335
Tc00.1047053510303.210             CDTIVWAYGMLPNISVGTWIYEWMKDGAKERGEMISDFSIYTAGSCRDSYTGEGHGEQDLLRAVHEGYEIGYKI 717
Tc00.1047053509941.100             VDSIVFCHGQQS                DQRLLSGLQNVKYHLIGGCFEA      AELDAKRAILQGHELALKI 693
```

(b)




(c)



## Tables

**Table 1 - *Trypanosoma cruzi* 3D protein models.**

| Quality | TOTAL |
|---|---|
| **1. Very High** | 50 |
| **2. High** | 200 |
| **3. Good** | 79 |
| **4. Medium to Good** | 835 |
| **5. Medium to Low** | 873 |
| **6. Low** | 759 |
| **7. Very Low** | 490 |
| **TOTAL** | **3,286** |

Number of *Trypanosoma cruzi* proteins that could be modeled by comparative modelling using the MHOLline workflow and their respective quality. The quality of models depends on sequence identity and coverage (See the Table 7 in the Methods for detailed description).

**Table 2 - Predicted proteins and enzymatic functions of *Trypanosoma cruzi* using different cutoffs and KEGG and Swiss-Prot databases.**

| **Cutoff** | $10e^{-20}$ | | $10e^{-40}$ | | $10e^{-80}$ | |
|---|---|---|---|---|---|---|
| **Database** | **KEGG** | **Swiss-Prot** | **KEGG** | **Swiss-Prot** | **KEGG** | **Swiss-Prot** |
| **Predicted Functions**[a] | 3,625 | 2,743 | 2,805 | 1,924 | 1,751 | 762 |
| **Enzymatic Functions**[b] | 1,027 | 749 | 770 | 523 | 517 | 246 |

[a]Total number of predicted proteins with functions inferred by AnEnΠ.
[b]Total number of distinct enzymatic functions (EC number) from predicted proteins in [a].

**Table 3 - Comparison between *Homo sapiens* and *Trypanosoma cruzi* functions obtained from KEGG and Swiss-Prot databases.**

| AnEnΠ Classification | KEGG | Swiss-Prot |
|---|---|---|
| **Homologous**[a] | 356(107) | 194 (71) |
| **Analogous** | 28 (5) | 8 (6) |
| **Specific of** *T. cruzi* | 133 (6) | 44 (7) |

Numbers in parenthesis represent the number of enzymatic functions (EC number) found among the modelled proteins from *T. cruzi*, using a cut-off of $10e^{-80}$.
[a]In some cases, a given protein of the parasite is analogous to a human protein but has also an homologous counterpart; these cases were included here.

**Table 4 - Enzyme Commission Numbers (EC) associated to modelled *Trypanosoma cruzi* proteins.**

| EC[a] | Description | Total ID[b] | EC[a] | Description | Total ID[b] |
|---|---|---|---|---|---|
| 1.1.1.205 | IMP dehydrogenase | 2 | 3.1.1.3 | Triacylglycerol lipase | 1 |
| 1.1.1.271 | GDP-L-fucose synthase | 1 | 3.1.1.29 | Aminoacyl-tRNA hydrolase | 1 |
| 1.1.1.35 | 3-hydroxyacyl-CoA dehydrogenase | 2 | 3.1.2.6 | Hydroxyacylglutathione hydrolase | 2 |
| 1.1.1.37 | Malate dehydrogenase | 6 | 3.1.2.15 | Ubiquitin thiolesterase | 9 |
| 1.1.1.40 | Malate dehydrogenase (oxaloacetate-decarboxylating)(NADP(+)) | 4 | 3.1.3.11 | Fructose-bisphosphatase | 2 |
| 1.1.1.42 | Isocitrate dehydrogenase (NADP(+)) | 2 | 3.1.3.16 | Phosphoprotein phosphatase | 23 |
| 1.1.1.44 | Phosphogluconate dehydrogenase (decarboxylating) | 1 | 3.1.3.25 | Inositol-phosphate phosphatase | 2 |
| 1.11.1.9 | Glutathione peroxidase | 1 | 3.1.3.48 | Protein-tyrosine-phosphatase | 11 |
| 1.11.1.15 | Peroxiredoxin | 1 | 3.2.1.18 | Exo-alpha-sialidase | 1 |
| 1.15.1.1 | Superoxide dismutase | 5 | 3.4.11.1 | Leucyl aminopeptidase | 2 |
| 1.2.1.3 | Aldehyde dehydrogenase (NAD(+)) | 1 | 3.4.11.18 | Methionyl aminopeptidase | 4 |
| 1.2.1.5 | Aldehyde dehydrogenase (NAD(P)(+)) | 2 | 3.4.21.26 | Prolyl oligopeptidase | 1 |
| 1.2.1.12 | Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) | 10 | 3.4.22.51 | Cruzipain | 8 |
| 1.2.4.1 | Pyruvate dehydrogenase (acetyl-transferring) | 4 | 3.4.24.16 | Neurolysin | 2 |
| 1.2.4.4 | 3-methyl-2-oxobutanoate dehydrogenase (2-methylpropanoyl-transferring) | 2 | 3.4.24.36 | Leishmanolysin | 56 |
| 1.3.1.34 | 2,4-dienoyl-CoA reductase(NADPH) | 2 | 3.4.24.56 | Insulysin | 2 |
| 1.3.99.3 | Acyl-CoA dehydrogenase | 2 | 3.4.25.1 | Proteasome endopeptidase complex | 4 |
| 1.6.2.2 | Cytochrome-b5 reductase | 7 | 3.5.1.88 | Peptide deformylase | 3 |
| 1.6.2.4 | NADPH–hemoprotein reductase | 2 | 3.5.4.6 | AMP deaminase | 9 |
| 1.8.1.4 | Dihydrolipoyl dehydrogenase | 4 | 3.6.1.1 | Inorganic diphosphatase | 2 |
| 1.8.1.12 | Trypanothione-disulfide reductase | 2 | 3.6.3.14 | H(+)-transporting two-sector ATPase | 1 |
| 2.3.1.7 | Carnitine O-acetyltransferase | 1 | 3.6.3.6 | Proton-exporting ATPase | 2 |
| 2.3.1.12 | Dihydrolipoyllysine-residue acetyltransferase | 1 | 3.6.4.6 | Vesicle-fusing ATPase | 2 |
| 2.3.1.29 | Glycine C-acetyltransferase | 4 | 4.1.1.35 | UDP-glucuronate decarboxylase | 2 |
| 2.3.1.61 | Dihydrolipoyllysine-residue succinyltransferase | 2 | 4.1.1.50 | Adenosylmethionine decarboxylase | 1 |
| 2.3.2.5 | Glutaminyl-peptide cyclotransferase | 2 | 4.2.1.17 | Enoyl-CoA hydratase | 2 |
| 2.4.2.29 | tRNA-guanine transglycosylase | 2 | 4.2.1.22 | Cystathionine beta-synthase | 10 |
| 2.4.2.30 | NAD(+)ADP-ribosyl transferase | 2 | 4.2.1.47 | GDP-mannose 4,6-dehydratase | 1 |
| 2.5.1.46 | Deoxyhypusine synthase | 1 | 4.2.99.18 | DNA-(apurinic or apyrimidinic site) lyase | 4 |
| 2.5.1.47 | Cysteine synthase | 2 | 4.6.1.1 | Adenylate cyclase | 3 |
| 2.5.1.60 | Protein geranylgeranyltransferase typeII | 1 | 5.1.1.4 | Proline racemase | 1 |
| 2.6.1.1 | Aspartate transaminase | 3 | 5.1.3.1 | Ribulose-phosphate 3-epimerase | 2 |
| 2.6.1.5 | Tyrosine transaminase | 7 | 5.2.1.8 | Peptidylprolyl isomerase | 12 |
| 2.7.1.24 | Dephospho-CoA kinase | 1 | 5.3.3.8 | Dodecenoyl-CoA isomerase | 2 |
| 2.7.1.30 | Glycerol kinase | 1 | 5.3.4.1 | Protein disulfide-isomerase | 3 |
| 2.7.1.105 | 6-phosphofructo-2-kinase | 2 | 5.99.1.3 | DNA topoisomerase (ATP-hydrolyzing) | 2 |
| 2.7.2.3 | Phosphoglycerate kinase | 6 | 6.1.1.2 | Tryptophan–tRNA ligase | 5 |
| 2.7.4.3 | Adenylate kinase | 11 | 6.1.1.6 | Lysine–tRNA ligase | 3 |
| 2.7.4.6 | Nucleoside-diphosphate kinase | 1 | 6.1.1.7 | Alanine–tRNA ligase | 2 |
| 2.7.4.8 | Guanylate kinase | 1 | 6.1.1.9 | Valine–tRNA ligase | 2 |
| 2.7.6.1 | Ribose-phosphate diphosphokinase | 3 | 6.1.1.12 | Aspartate–tRNA ligase | 3 |
| 2.7.7.6 | DNA-directed RNA polymerase | 4 | 6.1.1.18 | Glutamine–tRNA ligase | 2 |
| 2.7.7.7 | DNA-directed DNA polymerase | 9 | 6.1.1.22 | Asparagine–tRNA ligase | 1 |
| 2.7.11.1 | Non-specific serine/threonine protein kinase | 23 | 6.2.1.1 | Acetate–CoA ligase | 2 |
| 2.7.11.22 | Cyclin-dependent kinase | 7 | 6.3.2.19 | Ubiquitin–protein ligase | 13 |
| 2.7.11.24 | Mitogen-activated protein kinase | 5 | | | |
| 2.7.11.26 | [Tau protein] kinase | 1 | | | |
| 2.7.12.1 | Dual-specificity kinase | 3 | | | |

[a] EC number determined by AnEnΠ methodology.
[b] Amount of *Trypanosoma cruzi* identification number according to TcruziDB (version 5.0).

**Table 5 - Protein Models: AnEn$\pi$ and enzyme classifications, and model quality.**

| AnEnΠ | Enzyme Classes | Quality Models | | | | | | | TOTAL |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| **1. Homologous** | **Oxidorreductases** | 5 | 16 | - | 25 | 8 | 4 | 1 | **59** (trypanothione-disulfide reductase) |
| | **Transferases** | 7 | 17 | 3 | 41 | 12[b] | 15 | 9 | **104** (protein kinases, polymerases) |
| | **Hydrolases** | - | 9 | 5 | 38 | 17 | 14 | 10 | **93** (trans-sialidase, endopeptidases) |
| | **Lyases** | - | 12 | 1 | 1 | 4 | 2 | - | **20** (hydratases, endonucleases) |
| | **Isomerases** | - | 1 | 3 | 8 | 1 | 7 | 2 | **22** (peptidylprolyl isomerase) |
| | **Ligases** | - | 8 | 1 | 14 | - | 5 | 5 | **33** (glutathione synthase, ubiquitins) |
| **2. Analogous** | **Oxidorreductases** | - | - | - | 1 | 1 | - | - | **2** (dehydrogenases) |
| | **Hydrolases** | - | - | - | 1 | - | - | - | **1** (phosphatases) |
| **3. Specific of** *T. cruzi* | | | | | | | | | |
| | **Oxidorreductases** | 1 | - | 1 | - | - | - | - | **2** (trypanothione-disulfide reductase) |
| | **Transferases** | - | - | - | 2 | - | - | - | **2** (protein kinases, polymerases) |
| | **Hydrolases** | - | - | 2 | 23 | 1 | 4 | 4 | **34** (cruzipain, leishmanolisin) |
| **4. Undetermined**[a] | | | | | | | | | |
| | **Hydrolases** | - | - | - | 22 | - | - | - | **22** (leishmanolisin) |
| | **Lyases** | - | - | - | - | - | - | 3 | **3** (hydratases, endonucleases) |
| **TOTAL** | | **13** | **63** | **16** | **176** | **44** | **51** | **34** | **397** |

In parenthesis are presented examples of proteins found in final dataset.
[a] Conflicting clustering between results obtained by KEGG and Swiss-Prot databases using AnEnΠ methodology.
[b] Two sequences were identified as conflicting annotation between the methodology proposed in this work and GeneDB.

**Table 6 - List of modelled sequences classified by AnEnΠ as analogous or specific of *Trypanosoma cruzi*, in relation to *Homo sapiens*.**

| Categories | Quality Models | EC[a] | Description[b] |
|---|---|---|---|
| **A. Analogous** | 4 | 1.3.1.34 | 2,4-dienoyl-CoA reductase(NADPH) (ID[c]: Tc00.1047053509941.100) |
| | 5 | 1.3.1.34 | 2,4-dienoyl-CoA reductase(NADPH) (ID[c]: Tc00.1047053510303.210) |
| | 6 | 3.1.1.3 | Triacylglycerol lipase (ID[c]: Tc00.1047053509005.50) |
| **B. Specific of** *T. cruzi* | 1 | 1.8.1.12 | Trypanothione-disulfide reductase (ID[c]: Tc00.1047053503555.30) |
| | 3 | 1.8.1.12 | Trypanothione-disulfide reductase (ID[c]: Tc00.1047053504507.5) |
| | 4 | 2.5.1.47 | Cysteine synthase (ID[c]: Tc00.1047053507165.50, Tc00.1047053507793.20) |
| | 3 | 3.4.22.51 | Cruzipain (ID[c]: Tc00.1047053508595.50, Tc00.1047053507297.10) |
| | 6 | 3.4.22.51 | Cruzipain (ID[c]: Tc00.1047053506529.550, Tc00.1047053507537.20) |
| | 7 | 3.4.22.51 | Cruzipain (ID[c]: Tc00.1047053509429.320, Tc00.1047053507603.260, Tc00.1047053507603.270, Tc00.1047053509401.30) |
| | 5 | 3.6.3.6 | Proton-exporting ATPase (ID[c]: Tc00.1047053506649.20 ) |
| | 6 | 3.6.3.6 | Proton-exporting ATPase (ID[c]: Tc00.1047053505763.19 ) |
| | 4 | 3.4.24.36 | Leishmanolysin (ID[c]: Tc00.1047053511211.90, Tc00.1047053510565.150, Tc00.1047053507623.110, Tc00.1047053508699.100, Tc00.1047053508699.90, Tc00.1047053509011.80, Tc00.1047053506587.100, Tc00.1047053509205.100, Tc00.1047053506163.10, Tc00.1047053506163.20, Tc00.1047053508813.40, Tc00.1047053505965.10, Tc00.1047053506257.50, Tc00.1047053510899.10, Tc00.1047053505931.10, Tc00.1047053505931.20, Tc00.1047053511203.10, Tc00.1047053504397.20, Tc00.1047053506921.10, Tc00.1047053508475.30, Tc00.1047053505615.10, Tc00.1047053508825.10, Tc00.1047053510873.20, Tc00.1047053507197.10) |

[a] EC number determined by AnEnΠ methodology.
[b] EC number description obtained from Swiss-Prot database.
[c] *Trypanosoma cruzi* identification number according to TcruziDB (version 5.0).

**Table 7 - Classification according to the quality of the models built based on BLAST sequence identity and BATS coverage of the template in relation to the target.**

| Quality | Identity | Coverage |
|---|---|---|
| **1.Very High** | $\geq 75\%$ | $\geq 90\%$ |
| **2.High** | $\geq 50\%$ and $< 75\%$ | $\geq 90\%$ |
| **3.Good** | $\geq 50\%$ | $\geq 70\%$ and $< 90\%$ |
| **4.Medium to Good** | $\geq 35\%$ and $< 50\%$ | $\geq 70\%$ |
| **5.Medium to Low** | $\geq 25\%$ and $< 35\%$ | $\geq 70\%$ |
| **6.Low** | $\geq 25\%$ | $\geq 50\%$ and $< 70\%$ |
| **7.Very Low** | $\geq 25\%$ | $\geq 30\%$ and $< 50\%$ |

## Additional Files

**Table 1S - Complete list of homologous, analogous and specific 3D protein models of** *Trypanosoma cruzi* **versus** *Homo sapiens*.

File format: PDF

Title: Table1S.pdf

Description: Table 1S - Complete list of homologous, analogous and specific 3D protein models of *Trypanosoma cruzi* versus *Homo sapiens*.

# *Discussão*

A utilização da Bioinformática nos últimos anos levou ao aumento na disponibilização de informações, cuja compilação e divulgação costumavam ser bastante demoradas. Esses dados podem agora servir como ponto de partida para a elucidação de muitos mistérios acerca das ciências da vida. Nesta tese, o papel da bioinformática foi crucial para uma varredura inicial na busca de alvos moleculares para o desenvolvimento de novas drogas contra enfermidades causadas por parasitas. É importante ressaltar que somente uma análise *in silico* não é capaz de definir uma molécula como sendo um bom alvo metabólico, sendo necessária uma análise experimental *in vitro* e *in vivo*. Este compreende a demonstração que o alvo é essencial para o parasita. É necessário demonstrar que este alvo pode ser inibido ("drugability") e que isto bloqueará o crescimento ou idealmente matará o parasito sem toxicidade importante para o ser humano ("índice de seletividade"). Entretanto, esta abordagem computacional preliminar vem tomando espaço à medida que contribui para a diminuição de duas variáveis importantes nessa busca, tempo e dinheiro e aporta novos olhares e abordagens para esta temática.

*Anotação de sequências*

Com a análise comparativa de sequências que codificam proteínas em diversos genomas, é possível perceber que uma função não pode ser inferida para uma grande parte destes produtos gênicos, às vezes por causa da grande distância evolutiva (Koppensteiner *et al*., 2000). Partindo dessa pressuposição, a ferramenta AnEnPi (Artigo 1) possibilita a anotação funcional de proteínas que possuem atividade enzimática através da comparação com sequências similares em bancos de dados, verificando se descendem de um mesmo ancestral (homólogas), o que indicaria uma mesma funcionalidade. Parte-se da premissa segundo a qual sequências que possuem uma história evolutiva semelhante tendem a compartilhar funções similares (Lee *et al.*, 2007). Assim, a opção de "anotar" um produto gênico é feita usando o banco de dados KEGG, levando em consideração o alinhamento (executado com BLAST ou HMMER) entre a sequência de entrada (*query sequence*) e as sequências depositadas no banco (*subject sequence*). A homologia é inferida baseada no *e-value* do alinhamento. Este parâmetro é importante já que o valor obtido corresponde à probabilidade de se obter, com outra sequência aleatória de mesmo tamanho e composição de letras, outro alinhamento com pontuação igual ou superior. Assim, quanto mais próximo a zero, mais confiável é a consulta. Nas

análises feitas com alguns dos modelos escolhidos (Artigos 2, 3 e 4), valores de *e-value* bastante restritivos foram eleitos baseado em pontos de corte comumente estabelecidos em projetos genoma de diversos organismos. Por exemplo, na inferência de função para *T. cruzi* no Artigo 4 foram adotados pontos de corte distintos (*e-value* variando de $10e^{-20}$ a $10e^{-80}$), já que o valor exato depende de outros fatores, como conservação da família de proteínas, entre outros. Neste mesmo exemplo, o uso de um *e-value* bastante restritivo ($10e^{-80}$) confere um alto grau de confiabilidade para as análises seguintes de comparação dos análogos funcionais entre o genoma do parasita e de humano. Outros parâmetros obtidos do alinhamento entre as sequências de entrada e as sequências do banco de dados, como cobertura e pontuação do alinhamento normalizada (*bitscore*), podem ser bastante informativas na inferência de função e devem ser posteriormente incluídas na ferramenta.

Para o desenvolvimento da ferramenta AnEnPi, e o uso subseqüente em anotação, a base de dados inicial escolhida foi o KEGG, um importante banco de dados curado com informações funcionais relativas a vias metabólicas bem estabelecidas. O KEGG é composto por dados oriundos de projetos genomas individuais, produtos gênicos e suas funções, mas um diferencial apresentado por este banco é a integração de informações bioquímicas e genéticas, sendo um dos bancos de dados amplamente utilizados na consulta de mapas de vias bioquímicas. Foi possível obter um número considerável de “EC numbers” que representam funções enzimáticas para as reações envolvidas nas vias metabólicas de referência armazenadas no banco. Cada mapa de referência representa uma via metabólica, e é constituído de uma rede de enzimas (representadas por seu “EC number” correspondente) interconectadas. Uma vez identificados os genes que codificam enzimas em um genoma, os “EC numbers” são devidamente atribuídos, e assim, as vias organismo-específicas podem ser reconstruídas computacionalmente, correlacionando esses genes com seus produtos (enzimas) nas vias de referência depositadas no KEGG (Kanehisa & Goto, 2000). Existem “EC numbers” cuja função em uma via metabólica já foi comprovada experimentalmente, mas que não possuem sequência representativa nos genomas depositados no KEGG. Devido a este fato, muitas vias metabólicas podem não ser inteiramente caracterizadas pela falta de informação de alguns passos nelas presente e, portanto, a reconstrução metabólica de um organismo utilizando os mapas de referência pode ter casos de

falso-negativos pela ausência de informação nos bancos devido a um baixo número de organismos que contenham tais informações.

Um aspecto importante a ser ressaltado é que o KEGG utiliza a classificação de “EC numbers” proposta pela IUBMB. Existe controvérsia neste tipo de classificação hierárquica, a qual está relacionada com a reação catalisada pela enzima, mas não leva em conta outros aspectos importantes, como diferenças estruturais e história evolutiva (Hattori *et al.*, 2003). Outras classificações que abordam aspectos não inclusos nos ECs têm sido propostas, porém ainda não existe uma classificação ideal englobando todas as famílias protéicas. Deste modo, adotou-se para este trabalho a mesma classificação utilizada em bancos de dados relacionados com vias metabólicas.

*Enzimas isofuncionais não-homólogas: identificação e evolução*

Assume-se que enzimas que catalisam a mesma reação tipicamente terão significativas semelhanças entre suas estruturas primárias e terciárias. Para a maioria das enzimas, isto é provavelmente verdadeiro. No entanto, um grande número de enzimas possui duas ou mais formas com pequena ou nenhuma similaridade de sequência, mas demonstrando compartilharem a mesma função. Estas isoformas alternativas podem ser associadas com diferentes linhagens filogenéticas, podem ter diferentes mecanismos catalíticos e pequena semelhança estrutural. Tais enzimas podem ter surgido através de acontecimentos evolutivos independentes, ao invés de terem descendido de uma enzima ancestral comum, e apropriadamente são referidas como isoformas não-homólogas (enzimas análogas), ao contrário de homólogas. A comparação de sequências, por si só não pode provar que duas sequências têm ligação de um ponto de vista evolutivo; sua origem comum só pode ser inferida a partir da conservação de estrutura da proteína mesmo depois que a conservação da estrutura primária seja divergente. A possibilidade de uma origem comum só pode ser rejeitada quando uma candidata a enzima análoga tem diferente dobras tridimensionais (Fitch, 1970; Galperin *et al.*, 1998; Hegyi & Gerstein, 1999; Morett *et al.*, 2003; George *et al.*, 2004; Gherardini *et al.*, 2007; Omelchenko *et al.*, 2010).

Também é sabido que algumas atividades enzimáticas estão presentes em quase todos os organismos; estas reações de catálise são parecem ser centrais. Entretanto, o potencial do desenvolvimento de drogas atuando em atividades enzimáticas comuns ao homem e algum patógeno em particular tenha sido subestimado, pois diferentes enzimas com distintos mecanismos de catálise podem executar a mesma função, mas possuir diferenças estruturais em suas proteínas, as quais podem ser aproveitadas para o desenvolvimento de novas drogas (Congreve *et al*., 2005). Assim, diferenças na estrutura primária podem determinar diferenças estruturais menores que podem ser explorados como um primeiro critério para eleger a proteína como um alvo terapêutico (Karp *et al*., 1999; Kramer & Cohen, 2004). A comparação isolada de sequências não pode provar que duas proteínas não possuam ligação de um ponto de vista evolutivo; a origem comum das mesmas só pode ser descartada (ou inferida) a partir do exame do grau de conservação e similaridade de suas estruturas terciárias, mesmo depois da análise de suas estruturas primárias indicar um alto grau de divergência. No entanto, pouco foi feito para identificar e anotar a ocorrência de eventos de evolução convergente, no sentido deste trabalho. A ferramenta AnEnPi baseia-se neste princípio, utilizando valores de ponto de corte baseado na diferença estrutural proposta anteriormente por Galperin e colaboradores em 1998.

Uma observação importante é que análises prévias feitas por Galperin e colaboradores em 1998 chegaram a um ponto de corte baseado em análises estruturais de funções enzimáticas. Uma reavaliação destes parâmetros foi feita pelos mesmos autores (Omelchenko *et al*., 2010), e a maioria das análise foi validada. Vale a pena comentar que para outros tipos de proteínas (transportadores, estruturais, etc), o ponto de corte estipulado não é aceitável. Outras análises com as estruturas de tais proteínas teriam que ser refeitas para estabelecer um novo ponto de corte.

Domínios de proteínas vistos como unidades independentes formam blocos modulares que se dobram para formar a proteína (Das & Smith, 2000). Estruturas de classificação poderiam abordar tal estrutura. Porém, a definição da estrutura 3D não é um problema totalmente resolvido e que explica algumas das diferenças entre os atuais sistemas de classificação. Muitas proteínas são proteínas multi-domínio, e estas são mais comuns em eucariotos do que procariotos, sugerindo o papel da

evolução no surgimento e manutenção da estrutura de domínios na composição das proteínas, o que não é levado em consideração nos sistemas de classificação. Alguns domínios possuem similaridade alta entre suas sequências e são evolutivamente relacionados, enquanto outros tem similaridade distante, mas com topologia similar, não deixando clara a evidência de ancestralidade comum (Meier *et al.*, 2007). Uma análise com os domínios das proteínas, verificando sua posição na estrutura, assim como as interações com outras regiões podem vir a esclarecer a história evolutiva das funções enzimáticas e melhorar as análises de analogia funcional.

Outro aspecto importante diz respeito às enzimas multiméricas, ou seja, aquelas que possuem mais de uma cadeia protéica em sua estrutura quaternária. A análise buscando por analogia entre proteínas multiméricas com cadeias idênticas (homo-multiméricas) não interfere tanto nos resultados, já que cadeias iguais não seriam falsamente alocadas em grupos distintos após o agrupamento. Já as hetero-multiméricas, por apresentarem cadeias distintas, podem resultar em falsos casos de analogia quando houver diferença nas estruturas primárias que ultrapasse o valor de corte estipulado no agrupamento, separando as cadeias em grupos diferentes. Isto ocorre porque, apesar das diferentes cadeias poderem ser, em vários casos, oriundas de genes distintos, nas bases de dados é comum que todas “herdem” a atividade da enzima como um todo, ocasionando falsos positivos quando da busca por analogia. Ainda é difícil estabelecer uma metodologia para resolver tal problema. Tentativas mal sucedidas utilizando a nomenclatura foram feitas utilizando diversas fontes curadas, como BRENDA, ENZYME e SWISS-PROT. Entretanto o problema é ainda mais complexo, já que para uma mesma atividade enzimática determinada por um único “EC number”, podem existir proteínas representantes com diferentes quantidades de subunidades (dímero, trímero, tetrâmero, etc) dependendo do organismo analisado.

Alguns trabalhos já identificavam a presença de enzimas isofuncionais não-homólogas em uma fração substancial de reações bioquímicas (Galperin *et al.*, 1998; Hegyi & Gerstein, 1999; Morett *et al.*, 2003; George *et al.*, 2004; Gherardini *et al.*, 2007; Omelchenko *et al.*, 2010). Porém, o conhecimento do papel da convergência na evolução dessas proteínas ainda é limitado. Especula se que pelo menos algumas destas enzimas foram recrutadas a partir de famílias de enzimas

ativas com substratos relacionados e flexíveis para acomodar mudanças na especificidade do substrato. Sugere-se também que isso ocorreu de forma relativamente recente (Omelchenko *et al*., 2010).

A distribuição filogenética de enzimas análogas não tem um padrão estabelecido. As formas distintas podem estar distribuídas uniformemente em diferentes Reinos e também ter formas distintas em um mesmo organismo (analogia intragenômica); alguns casos possuem uma distribuição bem definida, onde cada forma está presente em um Reino diferente, com pouca ou nenhuma sobreposição; ou até de forma não uniforme, com uma forma presente em quase todos os organismos (Gherardini *et al.*, 2007; Omelchenko *et al*., 2010).

Em relação ao tamanho do genoma e ao ecossistema do organismo, foi observado que microorganismos com genomas pequenos são capazes de codificar uma forma simples da enzima, enquanto que os organismos com genoma maior frequentemente carregam genes para duas ou mais isoformas para a maioria das enzimas análogas analisadas. Além disso, organismos que vivem em situação de estresse tenderiam a conservar diferentes formas para uma mesma atividade. Estas observações parecem refletir na versatilidade metabólica de alguns organismos, principalmente aquelas que habitam ambientes complexos. Nesses organismos, a redundância parcial e especialização de enzimas parecem crescer mais rápido que aquela associada ao tamanho do genoma (Omelchenko *et al*., 2010).

Durante a evolução convergente, a função de proteínas como produtos de genes não-homólogos se dá independentemente resultando na realização da mesma (ou similar) reação global, agindo em substratos iguais ou parecidos. Este tipo de evolução convergente já havia sido descrito 40 anos atrás por Wright e colaboradores (Wright *et al.*, 1969) para relatar a estrutura cristalina da subtilisina. Mas o mecanismo para tal evento ainda é desconhecido. Assim como o surgimento e evolução das vias metabólicas, várias hipóteses podem ser traçadas para o surgimento de enzimas isofuncionais não-homólogas; entretanto, não existem evidências suficientemente fortes para corroborá-las. A proposição de uma teoria baseada em transferência lateral esbarra na teoria do surgimento *de novo* de uma determinada função. Além disso, não se tem dados satisfatórios para garantir que houve uma neofuncionalização ou uma desfuncionalização.

*Reconstrução de vias metabólicas*

A qualidade da reconstrução de vias metabólicas depende em grande parte da qualidade inicial da designação das atividades enzimáticas. A falta de uma montagem completa das sequências do genoma de um organismo e a presença de muitas proteínas com funções desconhecidas tem dificultado a visão global das vias metabólicas de muitos parasitas (Galperin *et al.*, 1998). Além disso, identificar novos alvos para o desenvolvimento de medicamentos tem provado ser mais complexo do que previsto e revelou grandes lacunas no nosso entendimento das vias metabólicas e sua regulação integrada. Através da utilização da ferramenta AnEnPi, foi possível obter avanços na reconstrução do metabolismo de aminoácidos de *T. cruzi* (Artigo 2). Algumas vias catabólicas envolvidas nestes e outros processos participam de alguma maneira do metabolismo energético, através da utilização de aminoácidos como fonte de energia. Além disso, os aminoácidos também estão envolvidos no processo de diferenciação de formas replicativas para as infecciosas não-replicativas. Resíduos de prolina fazem parte do ciclo de diferenciação intracelular no interior do hospedeiro vertebrado (Sylvester e Krassner, 1976; Silber et al., 2005). Juntos, estes dados sugerem que as enzimas que participam no metabolismo dos aminoácidos podem ser bons candidatos para o desenvolvimento de novas drogas contra o parasita. Alvos potenciais identificados neste metabolismo, como a arginina quinase (EC: 2.7.3.3), a ATP-L-glutamato-5-fosfotransferase (EC: 2.7.2.11), a pirrolina-5-carboxilato sintetase (EC: 1.2.1.41) e outras, permitiram uma reconstrução metabólica mais precisa, mostrando a importância de métodos mais sensíveis para a atribuição da função. Algumas das funções identificadas neste trabalho já haviam sido reconhecidas pelo consórcio de anotação do genoma do parasita e também pelo KEGG, mostrando que a metodologia utilizada foi capaz de identificar corretamente tais funções. Este foi o caso da maioria das identificações, entretanto, alguns casos escaparam da detecção (3 e 11 casos usando *e-value* de $e^{-20}$ e $e^{-80}$ como ponto de corte respectivamente). Uma explicação para tal fato é difícil de ser encontrada já que não é muita clara qual a metodologia empregada para anotação feita pelo consórcio e pelo KEGG. Entretanto, neste mesmo trabalho, usando um ponto de corte bastante conservador de acordo com a literatura da área (*e-value* $e^{-80}$), o AnEnPi foi capaz de encontrar atividades enzimáticas ausentes no KEGG e no consórcio (por exemplo, o EC: 2.7.2.11). A identificação desta enzima

(ATP-L-glutamato 5 fosfotransferase) no parasita era esperada uma vez que o passo metabólico subsequente da via é seguido por uma enzima (sintetase pirrolina-5-carboxilato - EC: 1.2.1.41), que já tinha sido identificado anteriormente pelo KEGG. O reconhecimento da ATP-L-glutamato 5 fosfotransferase permitiu uma reconstrução metabólica mais precisa, mostrando a importância de métodos mais sensíveis para a atribuição da função. Além da reconstrução do metabolismo de aminoácido de *T. cruzi*, foi possível observar a existência de enzimas análogas, corroborando a idéia de que tais enzimas encontram-se distribuídas em diversas vias metabólicas (Galperin *et al.,* 1998; Morett *et al*., 2003).

Uma vez reconstruído o metabolismo de aminoácidos de *T. cruzi* utilizando o AnEnPi, optou-se por comparar uma análise global do metabolismo geral deste parasita com dados publicados na literatura (Artigo 3). Uma revisão do que já havia sido identificado, juntamente com as análises baseadas na abordagem de busca por análogos levou a uma comparação de atividades enzimáticas presentes em diversas vias metabólicas que possuíam isoformas não-homólogas em humanos e que poderiam ser consideradas como bons alvos terapêuticos. Sendo assim, foi possível prever um número considerável de funções identificadas para *T. cruzi* e melhorar a reconstrução de metabolismos de grande importância, como glicólise, ciclo de Krebs e metabolismo lipídico. Na literatura já era observada a grande importância desses metabolismos para a manutenção do parasita. A corroboração com nossos dados só levou à indicação de que moléculas que fazem parte destes metabolismos devem ser analisadas mais profundamente.

Alguns dados de estrutura foram analisados a partir de enzimas de um mesmo grupo de um determinado “EC number” que possuem estrutura resolvida no PDB. Um alinhamento estrutural dá indicação da semelhança estrutural em um grupo de proteínas. Não foi possível fazer uma análise global para todos os grupos gerados, já que não haviam muitas estruturas resolvidas no PDB para as sequências protéicas destes clusters.

A comparação de enzimas análogas em organismos distintos nos dá a perspectiva de encontrar bons alvos moleculares que possam vir a ser usados como modelos para o desenvolvimento de novas drogas contra diversas enfermidades. A análise comparativa entre as isoformas de *T. cruzi e H. sapiens* mostrou que a

integração de outra ferramenta (MHOLline) para predizer modelos tridimensionais de atividades enzimáticas através de modelagem comparativa, pode agilizar o processo de triagem inicial nesta busca por alvos moleculares (Artigo 4). Para que as enzimas análogas encontradas entre dois organismos sejam usadas como alvos para possíveis drogas é necessária uma série de experimentos adicionais para validar esta informação. Uma boa proposta que inclui a análise *in silico*, é a modelagem molecular *ab initio* das proteínas que não puderam ter modelos resolvidos por modelagem comparativa, quando a proteína é composta por um número pequeno de resíduos de aminoácidos, e posterior *docking* contra um banco de fármacos conhecidos com os modelos de proteínas relevantes para a sobrevivência do parasita. A modelagem também enriquece as análises de análogos em geral, uma vez que o número de estruturas tridimensionais ainda não é suficiente para uma análise global. Neste trabalho, a partir de 19.607 seqüências de proteínas preditas do genoma do *T. cruzi* (CL-Brener) como dados iniciais, foi possível obter modelos construídos para 3286 seqüências (26,6% do total). Destes, 1164 modelos (35,4%) foram classificados, segundo os critérios do MHOLline, como qualidade de “média para boa" a ‘muito alta”, adequando-se aos projetos de desenho racional de drogas baseado em estrutura. Um resultado importante deste trabalho foi a identificação e construção de modelos 3D de três sequências protéicas classificadas como análogos e 38 sequências específicas do parasita em relação às sequências de humanos. Estas 41 sequências podem ser consideradas como interessantes alvos moleculares para o desenvolvimento de drogas contra a doença de Chagas. É importante ressaltar que dentre estas proteínas, foi possível observar algumas proteínas que já estão sendo estudadas como alvos de drogas(Cruzipaína e tripanotiona-disulfide redutase). Das análogas, foi possível identificar duas atividades enzimáticas: 2,4-dienoyl-CoA redutase (EC 1.3.1.34) e lipase triacilglicerol (EC 3.1.1.3), envolvidas no metabolismo de lipídios, importante para a manutenção do parasita.

A busca pela presença de isoformas não-homólogas dentro do genoma de um mesmo organismo, o que denominamos de analogia intragenômica, pode levar a informações sobre a finalidade da existência destas formas em um mesmo organismo. As informações que temos atualmente não podem ser conclusivas, mas dados esparsos nos levam a especular que sua presença em um mesmo genoma eucariótico poderia estar relacionada com uma possível expressão diferencial ao longo do ciclo de vida, ou sua expressão em diferentes locais (dentro da própria

célula ou em diferentes tecidos), ou ainda em diferentes estágios do seu desenvolvimento. No caso dos procariotos, a presença de distintas enzimas, com prováveis diferenças entre suas afinidades (pelo substrato específico), na utilização de cofatores, entre suas cinéticas, estabilidade, etc, pode oferecer vantagens importantes na sobrevivência em determinadas situações. Estas informações se tornarão mais claras com a incorporação de dados experimentais como, por exemplo, dados oriundos da transcriptômica e proteômica e por uma vasta busca bibliográfica.

*Trabalhos em andamento, perspectivas futuras, dificuldades encontradas, contribuições deste trabalho, suporte financeiro*

Este trabalho se tornará mais completo com a finalização de algumas análises, como o estudo da existência de analogias de uma maneira global para todos os organismos; análise do comportamento destes análogos nas vias metabólicas, buscando o entendimento da evolução destas vias e as possíveis rotas alternativas ainda não identificadas; o desenvolvimento de um banco de dados com as informações obtidas na clusterização, anotação e busca de enzimas análogas com todos os organismos presentes no KEGG; e a disponibilização de todos os dados em uma interface Web.

O AnEnDB está sendo desenvolvido como um banco de dados relacional para armazenar os dados oriundos das análises do AnEnPi, incluindo os dados de enzimas análogas, comparações de sequências, dados de outros bancos de dados, taxonomia e outros. O sistema gerenciador de banco de dados é o ORACLE. Este banco de dados será disponibilizado na internet com uma interface programada em HTML/CGI/Perl ou localmente com Perl scripts.

Consultas previamente estabelecidas poderão ser acessadas, como por exemplo, quais enzimas reanotadas de uma determinada espécie em uma via específica estão ou não presentes em uma outra espécie, possibilidade de mudar os parâmetros da análise, assim como a possibilidade de consultas complexas usando linha de comando. Outro aspecto importante deste banco de dados é a representação dos resultados. Com o *output* serão gerados gráficos (utilizando bibliotecas públicas), tabelas (compatíveis com Excel e Matlab), mapas comparativos coloridos de vias metabólicas contendo código de cores diferenciado.

Um diagrama de entidade-relacionamento (DER) foi proposto baseado em um pré-existente, concebido para o projeto ProteinWorldDB (http://157.86.176.108/ProteinWorldDB/index.php) e está esquematizado na figura 11.

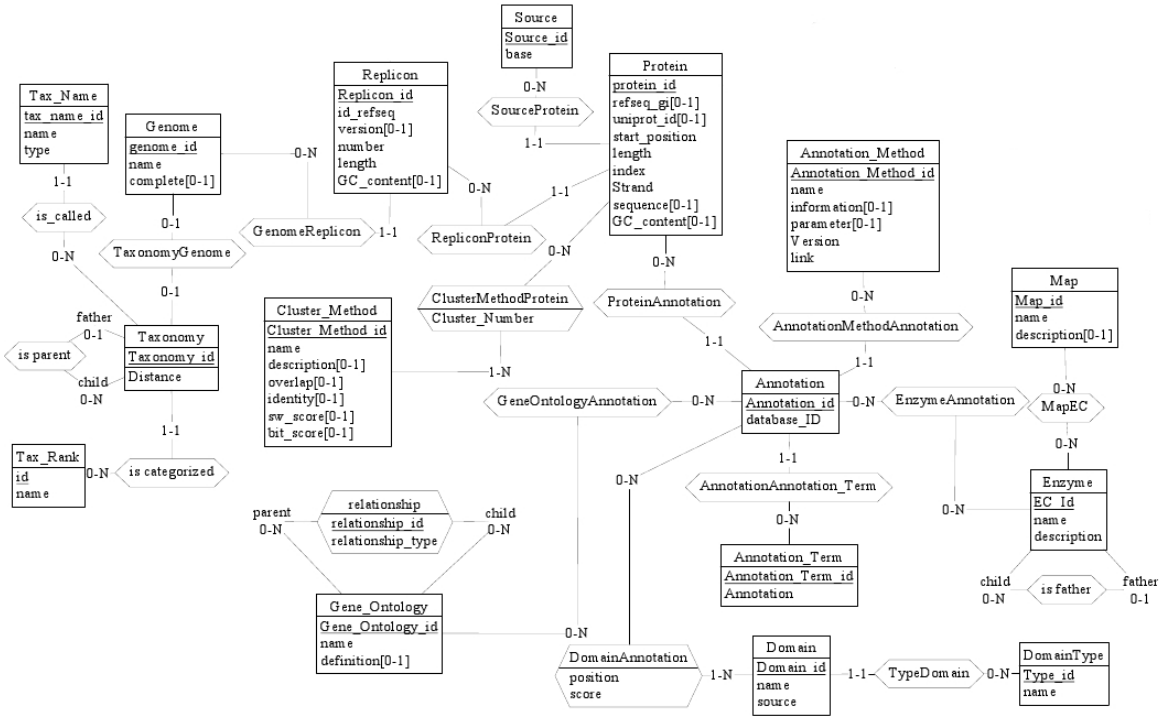



**Figura 11:** Diagrama de relacionamento preliminar do banco de dados AnEnDB, baseado no modelo conceitual do ProteinWorldDB (http://157.86.176.108/ProteinWorldDB/).

Com a implementação deste banco de dados a consulta à informação contida nos dados provenientes do AnEnPi se tornará mais rápida possibilitando uma maior eficiência na análise destes dados. Uma interface preliminar deste banco é apresentada na figura 12.

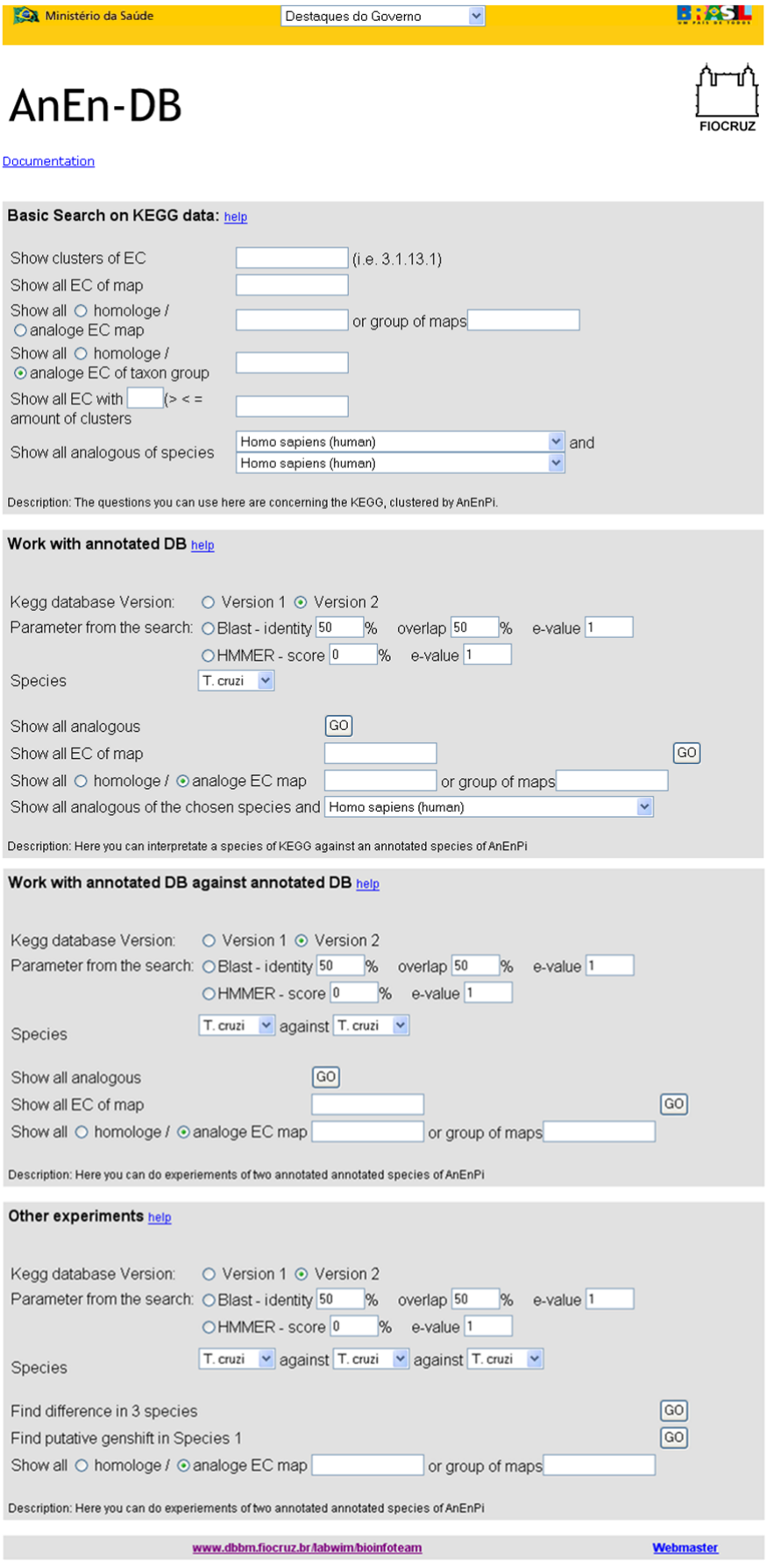



**Figura 12:** Interface preliminar do banco de dados AnEnDB.

No decorrer deste trabalho, me deparei com algumas dificuldades. Algumas delas foram superadas, como no caso da ferramenta gráfica utilizada para a geração dos mapas metabólicos. A princípio, a idéia era construir um *script* que pudesse "desenhar" os mapas de forma a representar os resultados obtidos nas análises do AnEnPi, utilizando os mapas de referência do KEGG. Entretanto, o próprio banco de dados do KEGG disponibilizou uma ferramenta, a *Color Objects in KEGG Pathways* (http://www.genome.jp/kegg/tool/color_pathway.html) que executa este mesmo trabalho, diminuindo o tempo de espera do resultado pelo usuário, agilizando o processo de análise da reconstrução de vias metabólicas.

Uma dificuldade que não pôde ser superada foi a quantidade limitada de dados do banco usado como fonte, o KEGG. A utilização de um número maior de dados, que poderia dar a falsa impressão de ser mais completo, como o GenBank (http://www.ncbi.nlm.nih.gov/genbank/) por exemplo, poderia levar a informações errôneas por não ser tão curado. Porém, existe a possibilidade da inclusão de novos bancos de dados de vias metabólicas e de reações enzimáticas, assim como o cruzamento com informações contidas em outros bancos de dados referentes a enzimas, como BRENDA (http://*www.brenda-enzymes.org/)*, ENZYME (http://expasy.org/enzyme/), METACYC (http://metacyc.org/), REACTOME (http://www.reactome.org/), entre outros.

Uma metodologia baseada em matrizes de distância foi proposta para fazer as análises de variabilidade das vias metabólicas, porém, a falta de informação suficiente não gerou resultados consistentes. É necessária a inclusão de novos organismos para este tipo de abordagem, o que vem sendo discutido por nosso grupo. Além disso, uma metodologia alternativa a utilizada neste estudo, não baseada na similaridade entre estruturas primárias para a inferência de função, ainda está sendo formulada para uma análise diferenciada dos mesmos dados.

No futuro, pretendemos examinar a variabilidade das vias metabólicas, levando em conta a existência de eventos de analogia, almejando a obtenção de conhecimento sobre a frequência e distribuição desses eventos, usando como modelo os genomas procarióticos e eucarióticos completamente sequenciados e disponíveis publicamente.

Outros trabalhos se beneficiaram das idéias e ferramentas utilizadas e desenvolvidas nesta tese. Entre eles podemos citar:

- “ESTs from seeds to assist the breeding of *Jatropha curcas.*”, em fase de finalização, em colaboração com o Dr. Nicolas Carels (CDTS/FIOCRUZ). Este trabalho visa à reconstrução metabólica de sementes de pinho-manso, utilizando ESTs (*Expressed Sequence Tags*), concentrando-se nas vias de ácidos graxos, energia e produtos secundários. A relevância deste trabalho se dá pelo fato destas sementes apresentarem substâncias oleosas que podem ser usadas na produção de biodiesel, combustível alternativo de grande importância para o país.

- “Reconstrução *in silico* das vias de processamento da informação genética nos TriTryps (*Trypanosoma cruzi*, *Trypanosoma brucei* e *Leishmania major*) – busca por análogos funcionais.”, dissertação de mestrado apresentada pela aluna Monete Rajão Gomes, no programa de pós-graduação em Biologia Computacional e de Sistemas do Instituto Oswaldo Cruz, da Fundação Oswaldo Cruz (IOC/FIOCRUZ). Este trabalho teve como objetivo a reconstrução das vias de processamento de informação genética nos Tritryps (os parasitos *Leishmania major*, *Trypanosoma brucei* e *Trypanosoma cruzi*), com ênfase na busca de análogos funcionais nas enzimas pertencentes a essas vias.

-“Genômica Comparativa de Procariotos: Análise da variabilidade em Funções Enzimáticas.”, trabalho de tese de doutorado apresentado pelo aluno Marcos Paulo Catanho de Souza, no programa de pós-graduação em Biologia Celular e Molecular, Fundação Oswaldo Cruz (IOC/FIOCRUZ). A colaboração exercida neste estudo se refere a identificação de análogos funcionais em procariotos presentes nas vias glicolítica e de gliconeogênese.

- “Identificação, clonagem, expressão heteróloga e localização da ribose 5-fosfato isomerase de *Leishmania major.”* Monografia de final de curso de graduação apresentada pelo aluno Marcos Gustavo Araujo Schwarz, para obtenção do Bacharelado em Genética pela Universidade Federal do Rio de Janeiro (UFRJ).

- “Identificação molecular, clonagem e expressão da fumarato hidratase de Leishmania major*.”* Monografia de final de curso de graduação apresentada pela

aluna Renata Cristina Menegussi Pereira, para obtenção do Bacharelado em Farmácia pela Universidade Estácio de Sá.

- “Caracterização molecular e bioquímica de enzimas da via de síntese de esteróides de *Trypanosoma cruzi*.”, trabalho de dissertação de mestrado que será defendido pela aluna Renata Aloise no Instituto de Biofísica Carlos Chagas Filho, da Universidade Federal do Rio de Janeiro (IBCCF/UFRJ).

O objetivo dos três últimos trabalhos foi fazer uma validação experimental de alvos moleculares propostos a partir da abordagem baseada em enzimas análogas, exposta neste trabalho.

A realização desta tese contou com o apoio financeiro de diversas fontes, listadas a seguir:

- Bolsa de doutorado do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo nº 142632/2006-6;

- Bolsa de doutorado-sanduíche da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), dentro do programa de Cooperação Internacional da CAPES com a Universidad de La Republica (UDELAR), no Uruguai, como bolsista do Programa de Doutorado no País com Estágio no Exterior (PDEE), com processo BEX nº 2941090;

- Programa de Desenvolvimento Tecnológico em Insumos para Saúde (PDTIS), através da Plataforma de Bioinformática – RJ (RPT04A).

- Projeto “Ecologia genômica: genes competindo por um nicho metabólico.”, do Programa Estratégico de Apoio à Pesquisa em Saúde (PAPES), através do Programa PAPES-V, 2008;

- Projeto “Identificação *in silico*, caracterização molecular e bioquímica de enzimas envolvidas com o metabolismo lipídico em *Leishmania braziliensis* e *Trypanosoma cruzi*: busca de novos alvos terapêuticos.”, do Edital Universal - MCT/CNPq - nº 15/2007 (processo no 486019/2007-1);

- Projeto “Estratégias racionais para a identificação de alvos terapêuticos e o desenvolvimento de uma quimioterapia.”, do programa DECIT conforme Edital MCT-CNPq / MS-SCTIE-DECIT - nº 25/2006 (processo nº 41.0544/2006-0);

- Projeto “Inovação em medicamentos e terapêutica visando profilaxia e tratamento de pacientes chagásicos agudos e crônicos.”, do programa DECIT conforme Edital MCT- CNPq / MS-SCTIE-DECIT - Nº 25/2006 (processo nº 40.9078/2006-9);

- Projeto “The Genome Comparison Project: Improving protein functional annotation in databases.”, que faz parte do World Community Grid™ (http://www.worldcommunitygrid.org/) da IBM®;

- Projeto “Análise Comparativa de Genomas Procarióticos.”, do programa CAPES-UDELAR, conforme Edital – CGCI – nº 029/2007;

- Projeto “Estratégias racionais para a identificação de alvos terapêuticos e o desenvolvimento de uma quimioterapia antiprotozoários patogênico.”, do programa PP-SUS/FAPERJ 2010, com o Dr Wanderley de Souza do Instituto de Biofísica da Universidade Federal do Rio de Janeiro (UFRJ).

# *Conclusões*

Com base nos resultados obtidos neste trabalho, podemos concluir que:

- ✓ A metodologia desenvolvida no AnEnPi mostrou-se eficiente na identificação e anotação de enzimas análogas. A estrutura modular permite ao usuário a utilização em diferentes contextos. A apresentação dos resultados é bastante amigável, com código de cores que permitem a representação de atributos biológicos, para uma melhor visualização das vias metabólicas, facilitando a interpretação dos resultados.
- ✓ Os pontos de corte (*cut-offs*) utilizados na inferência de função foram capazes de identificar um grande número de atividades enzimáticas, mesmo quando aplicado um valor bastante restritivo (*e-value* $e^{-80}$). A inclusão de outros parâmetros, além do *e-value,* podem refinar tais inferências.
- ✓ A ausência de algumas atividades nos mapas não significa necessariamente que esta atividade não esteja presente no organismo; isto pode ser explicado pela falta de sequências representativas da atividade em questão no banco de dados inicial.
- ✓ Assim como existem diversas atividades enzimáticas conhecidas sem um gene associado, existem diversos genes sem função conhecida, necessitando de mais estudos para a identificação destas atividades e posterior reconstrução metabólica
- ✓ Problemas com a montagem e anotação de sequências de genomas e a quantidade limitada de dados no banco pode dificultar a análise de busca por análogos.
- ✓ Enzimas multiméricas podem levar a falsos-positivos na identificação de análogas, uma vez que cadeias distintas podem se alocar em grupos diferentes (devido a forma como as anotações originais foram adicionadas aos registros das sequências).
- ✓ A abordagem computacional descrita neste estudo para a identificação das funções enzimáticas e da analogia, aplicada no AnEnPi e usadas nos tripanossomatídeos como modelos, podem ser utilizadas para outros organismos e mostrou que a identificação de novos genes ajudará a obter uma visão mais realista do metabolismo em geral.

- ✓ A classificação enzimática poderia ser melhorada com a inclusão de um quinto nível hierárquico no “EC number”, refletindo sua origem; ou a implementação de novas abordagens que levassem em conta características estruturais.
- ✓ A detecção de analogia mostrou-se interessante não só do ponto de vista evolutivo, mas também como uma abordagem alternativa na busca de enzimas que podem ser candidatos para o desenvolvimento de novas drogas.
- ✓ A combinação das ferramentas MHOLline e AnEnPi mostrou-se importante para inferência de função de proteínas, construção de modelos estruturais e análise comparativa entre dois organismos para identificar proteínas análogas, homólogas e específicas que podem ser consideradas como candidatas em uma triagem inicial para o desenvolvimento de novas drogas. Foram obtidos modelos para 397 genes codificadores de enzimas de *T. cruzi*, representando ~6% das proteínas totais.
- ✓ A estratégia utilizada suporta o conceito de que a análise estrutural, juntamente com a análise funcional de proteínas, pode ser uma metodologia computacional interessante para detectar potenciais alvos para o desenho racional de drogas baseado em estrutura.
- ✓ As análogas funcionais entre *T. cruzi* e *H. sapiens* 2,4-dienoyl-CoA reductase (EC 1.3.1.34) e triacylglycerol lipase (EC 3.1.1.3), foram identificadas como potenciais alvos moleculares com estrutura tridimensional resolvida. Ambas são de grande importância no metabolismo do parasita.

# *Referências Bibliográficas*

-Acosta H, Dubourdieu M, Quinõnes W, *et al.* Pyruvate phosphate dikinase and pyrophosphate metabolism in the glycosome of *Trypanosoma cruzi* epimastigotes. Comp Biochem and Physiol B. 2004; 138: 347-356.

-Adams CP, Brantner VV. Estimating The Cost Of New Drug Development: Is It Really $802 Million? Health Tracking. 2006; 25(2):420-428.

-Adroher FJ, Osuna A, Lupiáñez JA. Differential energetic metabolism during *Trypanosoma cruzi* differentiation. II. Hexokinase, phosphofructokinase and pyruvate kinase. Mol Cell Biochem. 1990; 94: 71-82.

-Affranchino JL, De Tarlovsky MN, Stoppani AO. Respiratory control in mitochondria from *Trypanosoma cruzi*. Mol Biochem Parasitol. 1985; 16: 289-298.

-Aggarwal K, Lee KH. Functional genomics and proteomics as a foundation for systems biology. Brief Funct Genomic Proteomic. 2003; 2(3): 175-184.

-Aho AV, Hopcroft JE, Ullman JD. The Design and Analysis of Computer Algorithms. Addison-Wesley 1974.

-Alphey MS, Yu W, Byers E, Li D, Hunter WN. Structure and Reactivity of Human Mitochondrial 2,4-Dienoyl-CoA Reductase: Enzyme-Ligand Interactions in a Distinctive Short-Chain Reductase Active Site. The Journal of Biological Chemistry. 2005; 280(4):3068-3077.

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

-Altschul SF, Madden TL, Schaffer AA, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997; 25: 3389–3402.

-Alves-Ferreira M, Guimarães ACR, Capriles PVSZ, *et al.* A new approach for potential drug target discovery through in silico metabolic pathway analysis using *Trypanosoma cruzi* genome information. Mem Inst Oswaldo Cruz. 2009; 104(8):1100-1110.

-Anderson LE, Pacold I. Chloroplast and Cytoplasmic Enzymes: Three Distinct Isoenzymes Associated with the Reductive Pentose Phosphate Cycle. Plant Physiology. 1970; 45:583-585.

-Ashburner M, Ball CA, Blake JA, *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000; 25(1): 25-29.

-Atwood JA 3rd, Weatherly DB, Minning TA, *et al.* The *Trypanosoma cruzi* proteome. Science. 2005; 309: 473-476.

-Bairoch A, Boeckmann B, Ferro S, Gasteiger E. Swiss-Prot: juggling between evolution and stability. Briengs in Bioinformatics. 2004; 5:39-55.

-Bard JB, Rhee SY. Ontologies in biology: design, applications and future challenges. Nat Rev Genet. 2004; 5(3):213-22.

-Barrett AJ, Canter CR, Liebecq C, *et al.* Enzyme Nomenclature 1992: Recommendations of the NCIUBMB on the Nomenclature and Classification of Enzymes. San Diego: Academic Press, 1992. 862p.

-Barrett MP, Tetaud E, Seyfang A, *et al.* Trypanosome glucose transporters. Mol Biochem Parasitol. 1998; 91: 195-205.

-Barycki JJ, O'Brien LK, Strauss AW, Banaszak LJ. Sequestration of the Active Site by Interdomain Shifting. The Journal of Biological Chemistry. 2000; 275(35):27186-27196.

-Beach DH, Goad LJ, Holz GG Jr. Effects of ketoconazole on sterol biosynthesis by *Trypanosoma cruzi* epimastigotes. Biochem Biophys Res Commun. 1986; 136: 851-856.

-Beard CB, Pye G, Steurer FJ, Rodriguez R, Campman R, Peterson AT, Ramsey J, Wirtz RA, Robinson LE. Chagas Disease in a Domestic Transmission Cycle in Southern Texas, USA. Emerging Infectious Diseases 2003, 9:103-105.

-Bennett TP, Frieden E. Modern Topics in Biochemistry. London: Macmillan, 1969. p. 43-45.

-Berman HM, Westbrook J, Feng Z, *et al.* The Protein Data Bank. Nucleic Acids

Res. 2000; 28(1): 235-242.

-Berriman M, Ghedin E, Hertz-Fowler C, *et al.* The genome of the African trypanosome *Trypanosoma brucei*. Science. 2005; 309: 416-422.

-Beswick TC, Willert EK, Phillips MA. Mechanisms of allosteric regulation of *Trypanosoma cruzi* S-adenosylmethionine decarboxylase. Biochemistry. 2006; 45: 7797-7807.

-Bilate AM, Cunha-Neto E. Chagas disease cardiomyopathy: current concepts of an old disease. Rev Inst Med Trop Sao Paulo. 2008; 50: 67-74.

-Bohager T. Enzymes: What the Experts Know! - Your Journey to Health and Longevity Starts Here. Prescott: One World Press, 2006. 292p.

-Bork P, Bairoch A. Go hunting in sequence databases but watch out for the traps. Trends Genet. 1996; 12(10): 425-427.

-Brak K, Doyle PS, McKerrow JH, Ellman JA. Identification of a new class of nonpeptidic inhibitors of cruzain. J Am Chem Soc. 2008; 130: 6404-6410.

-Bringaud F, Baltz T. Differential regulation of two distinct families of glucose transporter genes in *Trypanosoma brucei*. Mol Cell Biol. 1993; 13: 1146-1154.

-Bringaud F, Rivière L, Coustou V. Energy metabolism of trypanosomatids: adaptation to available carbon sources. Mol Biochem Parasitol. 2006; 149: 1-9.

-Buhaescu I, Izzedine H. Mevalonate pathway: a review of clinical and therapeutical implications. Clin Biochem. 2007; 40: 575-584.

-Cáceres AJ, Portillo R, Acosta H, *et al*. Molecular and biochemical characterization of hexokinase from *Trypanosoma cruzi*. Mol Biochem Parasitol. 2003; 126: 251-262.

-Canepa GE, Bouvier LA, Urias U, *et al.* Aspartate transport and metabolism in the protozoan parasite *Trypanosoma cruzi*. FEMS Microbiol Lett. 2005; 247: 65-71.

-Cannata JJ, Cazzulo JJ. The aerobic fermentation of glucose by *Trypanosoma*

*cruzi*. Comp Biochem Physiol B. 1984; 79: 297-308.

-Catanho M, Mascarenhas D, Degrave W, de Miranda AB: BioParser: A tool for processing of sequence similarity analysis reports. Applied Bioinformatics. 2006; 5:49-53.

-Cavalcanti MC, Targino R, Baião F, *et al.* Managing structural genomic workflows using web services. Data & Knowledge Engineering. 2005; 53: 45-74.

-Cavasotto CN, Phatak SS. Homology modeling in drug discovery: current trends and applications. Drug Discovery Today. 2009; 4(13-14):676-683.

-Cavazzuti A, Paglietti G, Hunter WN, et al. Discovery of potente pteridine reductase inhibitors to guide antiparasite drug development. Proc Natl Acad Sci USA. 2008; 105: 1448-1453.

-Cazzulo JJ, Cazzulo Franke MC, Franke de Cazzulo BM. On the regulatory properties of the pyruvate kinase from Trypanosoma cruzi epimastigotes. FEMS Microbiol Lett. 1989; 50: 259-263.

-Cazzulo JJ. Intermediate metabolism in Trypanosoma cruzi. J Bioenerg Biomembr. 1994; 26: 157-165.

-Chambers JW, Fowler ML, Morris MT, Morris JC. The antitrypanosomal agent lonidamine inhibits *Trypanosoma brucei* hexokinase 1. Mol Biochem Parasitol. 2008; 158: 202-207.

-Cherkasov A, Ho Sui SJ, Brunham RC, Jones SJ. Structural characterization of genomes by large scale sequence-structure threading: application of reliability analysis in structural genomics. BMC Bioinformatics. 2004; 5:101.

-Colebatch G, Trevaskis B, Udvardi M. Functional Genomics: Tools of the Trade. New Phytologist. 2002; 153(1); 27-36.

-Concepción JL, Adjé CA, Quiñones W, *et al.* The expression and intracellular distribution of phosphoglycerate kinase isoenzymes in *Trypanosoma cruzi*. Mol Biochem Parasitol. 2001; 118: 111-121.

-Congreve M, Murray CW, Blundell TL. Structural biology and drug discovery.

Drug Discov Today. 2005; 10(13): 895-907.

-Cordeiro AT, Cáceres AJ, Vertommen D, *et al.* The crystal structure of *Trypanosoma cruzi* glucokinase reveals features determining oligomerization and anomer specificity of hexose-phosphorylating enzymes. J Mol Biol. 2007; 372: 1215-1226.

-Cormen TT, Leiserson CE, Rivest RL. Introduction to Algorithms 2nd edition. Mit Press; 2001.

-Coura JR, Castro SL. A critical review on Chagas disease chemotherapy. Mem Inst Oswaldo Cruz. 2002; 97: 3-24.

-Coura JR. Chagas disease: what is known and what is needed - A background article. Memórias do Instituto Oswaldo Cruz. 2007; 102(I):113-122.

-Craven MW, Shavlik JW. Machine Learning Approaches to Gene Recognition. IEEE Expert. 1994; 9(2): 2-10.

-Croft SL, Seifert K, Duchêne M. Antiprotozoal activities of phospholipid analogues. Mol Biochem Parasitol. 2003; 126: 165-172.

-Das S, Smith TF. Identifying nature's protein Lego set. Adv Protein Chem. 2000; 54: 159–83

-de Boiso JF, Stoppani AO. The mechanism of acetate and pyruvate oxidation by *Trypanosoma cruzi*. J Protozool. 1973;20: 673-678.

-de Souza W. From the cell biology to the development of new chemotherapeutic approaches against trypanosomatids: dreams and reality. Kinetoplastid Biol Dis. 2002; 1: 3.

-de Souza W, Carreiro IP, Miranda K, Silva NL. Two special organelles found in *Trypanosoma cruzi*. An Acad Bras Cienc. 2000; 72(3):421-32.

-de Souza W. Microscopy and cytochemistry of the biogenesis of the parasitophorous vacuole. Histochem Cell Biol. 2005;123: 1-18.

-de Souza W, Campanati L, Attias M. Strategies and results of field emission scanning electron microscopy (FE-SEM) in the study of parasitic protozoa.

Micron. 2008; 39: 77-87.

-Dias JC, Machado EM, Fernandes AL, Vinhaes MC. General situation and perspectives of chagas disease in Northeastern Region, Brazil. Cadernos deSaúde Púublica. 2000; 16(2):13-34

-Dias JC. Chagas disease: successes and challenges. Cad Saúde Publica. 2006; 22: 2020-2021.

-Dixon H, Ginger CD, Williamson J. The lipid metabolism of blood and culture forms of *Trypanosoma lewisi* and *Trypanosoma rhodesiense*. Comp Biochem Physiol B. 1971; 39: 247-266.

-Dixon H, Ginger CD, Williamson J. Trypanosome sterols and their metabolic origins. Comp Biochem Physiol B. 1972; 41: 1-18.

-Docampo R, de Boiso JF, Stoppani AO. Tricarboxylic acid cycle operation at the kinetoplast-mitochondrion complex of *Trypanosoma cruzi*. Biochim Biophys Acta. 1978; 502: 466-476.

-Docampo R, de Souza W, Miranda K, et al . Acidocalcisomes - conserved from bacteria to man. Nat Rev Microbiol. 2005; 3: 251-261.

-Doolittle RF. Convergent evolution: the need to be explicit. Trends Biochem. Sci. 1994; 19: 15–18.

-Doyle PS, Zhou YM, Engel JC, McKerrow JH. A cysteine protease inhibitor cures Chagas' disease in an immunodeficientmouse model of infection. Antimicrob Agents Chemother. 2007; 51: 3932-3939.

-Dufresne A, Salanoubat M, Partensky F, *et al.* Genome sequence of the cyanobacterium Prochlorococcus marinus ss120, a nearly minimal oxyphototrophic genome. PNAS. 2003; 100(17):7-17.

-Durbin R, Eddy SR, Krogh A, Mitcison G. Biological sequence analysis: probabilistic models of protein and nucleic acids. Cambridge Uiversity Press; 1998.

-Duschak VG, Cazzulo JJ. Subcellular localization of glutamate dehydrogenases

and alanine aminotransferase in epimastigotes of *Trypanosoma cruzi*. FEMS Microbiol Lett. 1991; 67: 131-135.

-El-On J. Current status and perspectives of the immunotherapy of leishmaniasis. Isr Med Assoc J. 2009; 11(10): 623-628.

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

-El-Sayed NM, Myler PJ, Blandin G, Berriman M. Comparative Genomics of Trypanosomatid Parasitic Protozoa. Science 2005, 309(5733):404-409.

-Espinosa A, Clark D, Stanley SL Jr. Entamoeba histolytica alcohol dehydrogenase 2 (EhADH2) as a target for anti-amoebic agents. J Antimicrob Chemother. 2004; 54: 56-59.

-Eswar N, Webb B, Marti-Renom MA, *et al.* Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics. 2006; Capítulo 5: Unidade 5.6.

-Fan H, Mark AE. Refinement of homology-based protein structures by molecular dynamics simulation techniques. Protein Sci. 2004; 13(1):211-20.

-Fani R, Fondi M. Origin and evolution of metabolic pathways. Phys Life Rev. 2009.

-Fields S, Kohara Y, Lockhart DJ. Functional genomics. Proc Natl Acad Sci U S A. 1999; 96(16): 8825-8826.

-Fischer HP: Towards quantitative biology: Integration of biological information to elucidate Disease pathways and to guide drug discovery. Biotechnology Annual Review. 2005; 11:1-68.

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

-Fitch WM. Homology a personal view on some of the problems. Trends Genet. 2000; 16 (5): 227–231.

-Francke C, Siezen RJ, Teusink B: Reconstructing the metabolic network of a bacterium from its genome. Trends Microbiol. 2005; 13(11): 550-558.

-Freitas RF, Prokopczyk IM, Zottis A, *et al.* Discovery of novel *Trypanosoma cruzi* glyceraldehyde-3-phosphate dehydrogenase inhibitors. Bioorganic & Medicinal Chemistry. 2009; 17(6):2476-2482.

-Fricker SP, Mosi RM, Cameron BR, *et al.* Metal compounds for the treatment of parasitic diseases. J Inorg Biochem. 2008; 102: 1839-1845.

-Friedberg I. Automated protein function prediction--the genomic challenge. Brief Bioinform. 2006;7(3):225-42

-Frishman D. Protein annotation at genomic scale: the current status. Chem Rev. 2007;107(8):3448-66.

-Galarreta BC, Sifuentes R, Carrillo AK, *et al.* The use of natural product scaffolds as leads in the search for trypanothione reductase inhibitors. Bioorg Med Chem. 2008; 16: 6689-6695.

-Galperin MY, Koonin EV. Sources of systematic error in functional annotation of genomes: domain rearrangement, non-orthologous gene displacement and operons disruption. In Silico Biol. 1998; 1(1): 55-67.

-Galperin MY, Walker DR, Koonin EV. Analogous enzymes: independent inventions in enzyme evolution. Genome Res. 1998; 8(8):779-90.

-Galperin MY, Koonin EV: Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. Genetica. 1999; 106(1–2):159-170.

-Garnier J. Protein structure prediction. Biochimie. 1990; 72(8):513-24

-Gelb MH, Hol WGJ. Drugs to Combat Tropical Protozoan Parasites. Science. 2002; 297(19):343-344.

-George RA, Spriggs RV, Thornton JM, *et al.* SCOPEC: a database of protein catalytic domains. Bioinformatics. 2004; 20(l)1:i130-i136.

-Gerlt JA, Babbitt PC. Mechanistically diverse enzyme superfamilies: the

importance of chemistry in the evolution of catalysis. Curr Opin Chem Biol. 1998; 2: 607–612.

-Gherardini PF, Wass MN, Helmer-Citterich M, Sternberg MJ. Convergent evolution of enzyme active sites is not a rare phenomenon. J Mol Biol. 2007; 372(3): 817-845.

-Gilbert JV, Plaut AG, Wright A. Analysis of the immunoglobulin A protease gene of Streptococcus sanguis. Infect Immun. 1991; 59(1):7-17.

-Gogarten JP, Olendzenski L. Orthologs, paralogs and genome comparisons. Curr Opin Genet Dev. 1999; 9(6): 630-636.

-Goldsmith-Fischman S, Honig B. Structural genomics: computational methods for structure analysis. Protein Sci. 2003; 12(9): 1813-1821.

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

-Granick S. Speculations on the origins and evolution of photosynthesis. Ann NY Acad Sci. 1957; 69(2): 292–308.

-Guex N, Peitsch MC. SWISS-MODEL and Swiss-Pdb Viewer: An environment for comparative protein modeling. Electrophoresis. 1997; 18:2714.

-Guimarães ACR, Otto TD, Alves-Ferreira M, *et al. In silico* reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi*. Genetics and Molecular Research. 2008; 7(3):872-882.

-Haemers T, Wiesner J, Giemann D, *et al.* Synthesis of b- and c-oxa isosteres of fosmidomycin and FR900098 as antimalarial candidates. Bio Med Chem. 2008; 16: 3361-3371.

-Hall B, Zuzel T. Evolution of a new enzymatic function by recombination within a gene. Proc Natl Acad Sci USA. 1980; 77: 3529–3533.

-Hannaert V, Albert MA, Rigden DJ, *et al.* Kinetic characterization, structure modelling studies and crystallization of *Trypanosoma brucei* enolase. Eur J

Biochem. 2003; 270: 3205-3213.

-Harrill AH, Rusyn I. Systems biology and functional genomics approaches for the identification of cellular responses to drug toxicity. Expert Opin Drug Metab Toxicol. 2008; 4(11): 1379-1389.

-Harrow B, Mazur A. Textbook of Biochemistry. Philadelphia: Saunders, 1958. 109p.

-Hattori M, Okuno Y, Goto S, Kanehisa M. Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways. J Am Chem Soc. 2003; 125(39): 11853-11865.

-Hatzimanikatis V. Nonlinear metabolic control analysis. Metab Eng. 1999;1(1):75-87.

-Heby O, Persson L, Rentala M. Targeting the polyamine biosynthetic enzymes: a promising approach to therapy of African sleeping sickness, Chagas' disease and leishmaniasis. Amino Acids. 2007; 33: 359-366.

-Hegyi H, Gerstein M. The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. J. Mol. Biol. 1999; 288: 147–164.

-Hieter P, Boguski M. Functional genomics: it's all how you read it. Science. 1997: 24; 278(5338): 601-602.

-Hillisch A, Pineda LF, Hilgenfeld R: Utility of homology models in the drug discovery process. Drug Discovery Today. 2004, 9(15):659-669.

-Holum J. Elements of General and Biological Chemistry. 2nd edition. New York: Wiley, 1968. 377p.

-Hopkins AL, Groom CR. The druggable genome. Nature Reviews. 2002, 1:727-730.

-Horowitz NH. On the evolution of biochemical syntheses. Proc Natl Acad Sci USA. 1945; 31: 153–157.

-Horowitz NH. The evolution of biochemical syntheses – retrospect and prospect.

Evolving genes and proteins. New York: Academic Press, 1965: 15–23.

-Hubbard PA, Liang X, Schulz H, Kim JJP: The Crystal Structure and Reaction Mechanism of E. coli 2, 4 - Dienoyl CoA Reductase. Journal of Biological Chemistry. 2003; 278(39):37553-37560.

-Hudock MP, Sanz-Rodriguez CE, Song Y, *et al.* Inhibition of *Trypanosoma cruzi* hexokinase by bisphosphonates. J Med Chem. 2006; 49: 215-223.

-Humphrey W, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. Journal of Molecular Graphics .1996; 14:33-38.

-Hurtado-Guerrrero R, Peña-Díaz J, Montalvetti A, *et al.* Kinetic properties and inhibition of *Trypanosoma cruzi* 3-hydroxy-3-methylglutaryl CoA reductase. FEBS Letters. 2002; 510: 141-144.

-Igoillo-Esteve M, Maugeri D, Stern AL, *et al.* The pentose phosphate pathway in *Trypanosoma cruzi*: a potential target for the chemotherapy of Chagas disease. An Acad Bras Cienc. 2007; 79: 649-663.

-Ivens AC, Peacock CS, Worthey EA, *et al.* The genome of the Kinetoplastid parasite, *Leishmania major*. Science. 2005; 309(5733):436-342.

-Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. Journal of molecular biology. 1961; 3: 318–56.

-Jensen RA. Enzyme recruitment in evolution of new function. Annu Rev Microbiol. 1976; 30: 409–425.

-Jiménez-Jiménez J, Ledesma A, Zaragoza P, *et al.* Fatty acid activation of the uncoupling proteins requires the presence of the central matrix loop from UCP1. Biochim Biophys Acta. 2006; 1757: 1292-1296.

-Jones A, Hunt E, Wastling JM, *et al.* An object model and database for functional genomics. Bioinformatics. 2004; 20(10): 1583-1590.

-Jorgensen WL. The many roles of computation in drug discovery. Science. 2004; 303(5665):1813-8

-Juan SM, Cazzulo JJ, Segura EC. The pyruvate kinase of *Trypanosoma cruzi*.

Acta Physiol Lat Am. 1976; 26: 424-426.

-Jurado LA, Machín I, Urbina JA. *Trypanosoma cruzi* phosphoenolpyruvate carboxykinase (ATP-dependent): transition metal ion requirement for activity and sulfhydryl group reactivity. Biochim Biophys Acta. 1996; 1292: 188-196.

-Kandpal R, Saviola B, Felton J. The era of 'omics unlimited. Biotechniques. 2009; 46(5):351-2, 354-5.

-Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000; 28(1): 27-30.

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

-Kanehisa M, Bork P. Bioinformatics in the post-sequence era. Nature Genetics. 2003; 33:305-310.

-Kanehisa M, Goto S, Kawashima S, *et al.* The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004; 32(Database issue):D277-80.

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

-Karp PD, Krummenacker M, Paley S, Wagg J. Integrated pathway-genome databases and their role in drug discovery. Trends Biotechnol. 1999; 17(7): 275-281.

-Karp PD. Call for an enzyme genomics initiative. Genome Biol 2004; 5(8):401.

-Ke T, Klibanov AM. On enzymatic activity in organic solvents as a function of enzyme history. Biotechnol Bioeng. 1998; 57(6): 746-750.

-Kerleguer A, Massard S, Janus G, Joussemet M. Chagas disease: screening tests evaluation in a blood military center, prevalence in the French Army. Pathologie Biologie. 2007; 55:534-538.

-Kirchhoff LV, Paredes P, Lomeli-Guerrero A, *et al.* Transfusion-associated Chagas disease (American trypanosomiasis) in Mexico: implications for transfusion medicine in the United States. Transfusion. 2006; 46(2):298-304.

-Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? Nature

Reviews. 2004; 3:711-715.

-Koonin EV. Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet. 2005; 39: 309-338.

-Koppensteiner WA, Lackner P, Wiederstein M, Sippl MJ. Characterization of novel proteins based on known protein structures. J Mol Biol. 2000; 296(4):1139-52.

-Korn ED, Greenblatt CL. Synthesis of alpha-linoleic acid by *Leishmania enriettii*. Science. 1963; 142: 1301-1303.

-Korn ED, Von Brand T, Tobie EJ. The sterols of *Trypanosoma cruzi* and Crithidia fasciculata. Comp Biochem Physiol. 1969; 30: 601-610.

-Kotera M, Okuno Y, Hattori M, *et al.* Computational assignment of the EC numbers for genomic-scale analysis of enzymatic reactions. J Am Chem Soc. 2004;126(50):16487-16498.

-Kramer R, Cohen D. Functional genomics to new drug targets. Nature Reviews Drug Discovery. 2004; 3(11):965-972.

-Krauth-Siegel RL, Inhoff O. Parasite-specific trypanothione reductase as a drug target molecule. Parasitol Res. 2003; 90: S77-S85.

-Kukor JJ, Olsen RH. Catechol 2,3-dioxygenases functional in oxygen-limited (hypoxic) environments. Appl Environ Microbiol. 1996; 62(5):583-585.

-La D, Sutch B, Livesay DR. Predicting protein functional sites with phylogenetic motifs. Proteins. 2005; 58(2):309-20.

-Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst. 1993; 26: 283-291.

-Laskowski RA, Watson JD, Thornton JM. From protein structure to biochemical function? J Struct Funct Genomics. 2003;4(2-3):167-77.

-Lazcano A, Miller SL. The origin and early evolution of life: prebiotic chemistry,

the pre-RNA world, and time. Cell. 1996; 85(6): 793-798.

-Lee SH, Stephens JL, Paul KS, Englund PT. Fatty acid synthesis by elongases in trypanosomes. Cell. 2006; 126: 691-699.

-Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure. Nature Reviews Molecular Cell Biology. 2007; 8:995-1005.

-Leiby DA, Herron RM, Read EJ, *et al. Trypanosoma cruzi* in Los Angeles and Miami blood donors: impact of evolving donor demographics on seroprevalence an d implications for transfusion transmission. Transfusion. 2002; 42(5):549-555.

-Liendo A, Visbal G, Piras MM, Piras R, Urbina JA. Sterol composition and biosynthesis in *Trypanosoma cruzi* amastigotes. Mol Biochem Parasitol. 1999; 104: 81-91.

-Liñares GE, Ravaschino EL, Rodriguez JB. Progresses in the field of drug design to combat tropical protozoan parasitic diseases. Curr Med Chem. 2006; 13: 335-360.

-Lindsay MA. Target discovery. Nat Rev Drug Discov. 2003; 2(10):831-8

-Liò P, Brilli M, Fani R. Phylogenetics and computational biology of multigene families. Berlin: Springer; 2007. p. 191–205.

-Lorente SO, Jimenez CJ, Gros L, *et al.* Preparation of transition-state analogues of sterol 24-methyl transferase as potential anti-parasitics. Bioorg Med Chem. 2005; 13: 5435-5453.

-Lustbader JW, Cirilli M, Lin C, *et al.* ABAD Directly Links A to Mitochondrial Toxicity in Alzheimer's Disease. Science. 2004; 304:448-452.

-Ma H, Zeng AP. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 2003;19(2):270-7.

-Magaraci F, Jimenez CJ, Rodrigues C, *et al.* Azasterols as inhibitors of sterol 24-methyltransferase in *Leishmania* species and *Trypanosoma cruzi*. J Med Chem. 2003; 46: 4714-4727.

-Mandal S, Moudgil M, Mandal SK. Rational drug design. Eur J Pharmacol. 2009;625(1-3):90-100.

-Marin-Neto JA, Rassi A Jr, Morillo CA, *et al.* Rationale and design of a randomized placebo-controlled trial assessing the effects of etiologic treatment in Chagas' cardiomyopathy: the BENznidazole Evaluation for Interrupting Trypanosomiasis (BENEFIT). Am Heart J. 2008; 156: 37-43.

-Marsden RL, Lee D, Maibaum M, *et al.* Comprehensive genome analysis of 203 genomes provides structural genomics with new insights into protein family space. Nucleic Acids Res. 2006; 34(3):1066-80.

-Martinek R. Practical Clinical Enzymology. J. Am. Med. Tech. 1969; 31: 162.

-Marti-Renom MA, Madhusudhan MS, Fiser A, *et al.* Reliability of assessment of protein structure prediction methods. Structure. 2002; 10(3): 435-440.

-Martyn DC, Jones DC, Fairlamb AH, Clardy J. High-throughput screening affords novel and selective trypanothione reductase inhibitors with anti-trypanosomal activity. Bioorg Med Chem Lett. 2007; 17: 1280-1283.

-Maugeri DA, Cazzulo JJ. The pentose phosphate pathway in *Trypanosoma cruzi*. FEMS Microbiol Lett. 2004; 234: 117-123.

-Maya JD, Cassels BK, Iturriaga-Vásquez P, *et al.* Mode of action of natural and synthetic drugs against *Trypanosoma cruzi* and their interaction with the mammalian host. Comp Biochem Physiol A Mol Integr Physiol. 2007; 146: 601-620.

-Meier S, Jensen PR, David CN, *et al.* Continuous molecular evolution of protein-domain structures by single amino acid changes. Curr Biol. 2007; 17(2):173-8.

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

-Michels PA, Hannaert V, Bringaud F. Metabolic aspects of glycosomes in trypanosomatidae - new data and views. Parasitol Today. 2000; 16: 482-489.

-Mielniczki-Pereira AA, Chiavegatto CM, López JA, *et al. Trypanosoma cruzi*

strains, Tulahuen 2 and Y, besides the difference in resistance to oxidative stress, display differential glucose-6-phosphate and 6-phosphogluconate dehydrogenases activities. Acta Tropica. 2007; 101: 54-60.

-Milei J, Guerri-Guttenberg RA, Grana DR, Storino R: Prognostic impact of Chagas disease in the United States. American Heart Journal. 2008, 157:22-29.

-Miller SL. A production of amino acids under possible primitive earth conditions. Science. 1953; 117(3046): 528-529.

-Montalvetti A, Peña-Díaz J, Hurtado R, *et al.* Characterization and regulation of *Leishmania major* 3-hydroxy-3-methylglutaryl-CoA reductase. Biochem J. 2000; 349: 27-34.

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

-Moriya Y, Itoh M, Okuda S, *et al.* KAAS: an automatic genome annotation and Pathway reconstruction server. Nucleic Acids Res. 2007; 35:W182-W185.

-Moyersoen J, Choe J, Fan E, *et al.* Biogenesis of peroxisomes and glycosomes: trypanosomatid glycosome assembly is a promising new drug target. FEMS Microbiol Rev. 2004; 28: 603-643.

-Müller CE. Prodrug approaches for enhancing the bioavailability of drugs with low solubility. Chem Biodivers. 2009; 6(11):2071-83

-Murzin AG. Structural classification of proteins: new superfamilies. Curr Opin Struct Biol. 1996; 6(3): 386-394.

-Nelson DL, Cox MM. Lehninger - Principles of Biochemistry. 4ed. New York: W. H. Freeman & Company, 2004. 1100p.

-Nowicki C, Cazzulo JJ. Aromatic amino acid catabolism in trypanosomatids. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 2007.

-Ohta T. Evolution of gene families. Gene. 2000; 259(1–2): 45–52.

-Olivares-Illana V, Pérez-Montfort R, López-Calahorra F, *et al.* Structural differences in triosephosphate isomerase from different species and discovery of

a multitrypanosomatid inhibitor. Biochemistry. 2006; 45: 2556-2560.

-Olivares-Illana V, Rodríguez-Romero A, Becker I, *et al.* Perturbation of the dimer interface of triosephosphate isomerase and its effect on *Trypanosoma cruzi*. PLoS Negl Trop Dis. 2007; 1: e1.

-Omelchenko MV, Galperin MY, Wolf YI, Koonin EV. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution.Biol Direct. 2010; 5:31.

-Opperdoes FR. The trypanosomatidae: amazing organisms. J Bioenerg Biomembr. 1994; 26(2): 145-146.

-Otto TD, Catanho M, Degrave W, de Miranda AB. The PDTIS bioinformatics platform: from sequence to function. RECIIs. 2007; 1(2):286-294.

-Otto T, Guimarães A, Degrave W, Miranda A. AnEnPi: identification and annotation of analogous enzymes. BMC Bioinformatics. 2008; 9:544.

-Ouzounis CA, Valencia A. Early bioinformatics: the birth of a discipline--a personal view. Bioinformatics. 2003; 19(17): 2176-2190.

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

-Papin JA, Price ND, Wiback SJ, *et al.* Towards quantitative biology: Integration of biological information to elucidate disease pathways and to guide drug discovery. Biotechnology Annual Review. 2003; 11:1-68.

-Paulino M, Iribarne F, Dubin M, *et al.* The chemotherapy of Chagas' disease: an overview. Mini Rev Med Chem. 2005; 5: 499-519.

-Peitsch MC. About the use of protein models. Bioinformatics. 2002; 18:934-938.

-Peña-Díaz J, Montalvetti A, Camacho A, *et al.* A soluble 3-hydroxy-3-methylglutaryl-CoA reductase in the protozoan *Trypanosoma cruzi*. Biochem J. 1997; 324: 619-626.

-Pereira CA, Alonso GD, Paveto MC, *et al.* L-arginine uptake and L-

phosphoarginine synthesis in *Trypanosoma cruzi*. J Eukaryot Microbiol. 1999; 46: 566-570.

-Pérez-Montfort R, Garza-Ramos G, Alcántara GH, *et al.* Derivatization of the interface cysteine of triosephosphate isomerase from *Trypanosoma brucei* and *Trypanosoma cruzi* as probe of the interrelationship between the catalytic sites and the dimer interface. Biochemistry. 1999; 38: 4114-4120.

-Perez-Pineiro R, Burgos A, Jones DC, *et al.* Development of a novel virtual screening cascade protocol to identify potential trypanothione reductase inhibitors. J Med Chem. 2009; 52: 1670-1680.

-Pitkänen E, Rousu J, Ukkonen E. Computational methods for metabolic reconstruction. Curr Opin Biotechnol. 2010; 21(1): 70-77.

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

-Pourshafie M, Morand S, Virion A, *et al.* Cloning of S-adenosyl-L-methionine: C-24-Delta-sterolmethyltransferase (ERG6) from *Leishmania donovani* and characterization of mRNAs in wild-type and amphotericin B-Resistant promastigotes. Antimicrob Agents Chemother. 2004; 48: 2409-2414.

-Raamsdonk LM, Teusink B, Broadhurst D, *et al.* A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol. 2001;19(1):45-50.

-Racagni GE, Machado de Domenech EE. Characterization of *Trypanosoma cruzi* hexokinase. Mol Biochem Parasitol. 1983; 9: 181-188.

-Radwanska M. Emerging trends in the diagnosis of Human African Trypanosomiasis. Parasitology. 2010;1-10.

-Ramakrishnan C. In memoriam: Professor G.N. Ramachandran (1922-2001). Protein Sci. 2001;10(8):1689-91

-Rassi A Jr, Rassi A, Marin-Neto JA. Chagas disease. Lancet. 2010; 375(9723): 1388-1402.

-Reed JL, Patel TR, Chen KH, *et al.* Systems approach to refining genome annotation. Proc Natl Acad Sci U S A. 2006; 103(46):17480-4

-Reesink HW. European Strategies Against the Parasite Transfusion Risk. Transfusion Clinique et Biologique. 2005; 12:1-4.

-Reyes-Vivas H, Hernández-Alcantara G, López-Velazquez G, *et al.* Factors that control the reactivity of the interface cysteine of triosephosphate isomerase from *Trypanosoma brucei* and *Trypanosoma cruzi*. Biochemistry 2001; .40: 3134-3140.

-Reynolds KA, Holland KA. The mechanistic and evolutionary basis of stereospecificity for hydrogen Transfers in enzymecatalysed processes. Chemical Society Reviews. 1997; 26:337-343.

-Roberts CW, McLeod R, Rice DW, *et al.* Fatty acid and sterol metabolism: potential antimicrobial targets in apicomplexan and trypanosomatid parasitic protozoa. Mol Biochem Parasitol. 2003; 126: 129-142.

-Rocha MO, Teixeira MM, Ribeiro AL. An update on the management of Chagas cardiomyopathy. Expert Rev Anti Infect Ther. 2007; 5: 727-743.

-Rockey WM, Elcock AH. Progress toward virtual screening for drug side effects. Proteins. 2002; 48(4): 664-671.

-Rodríguez-Romero A, Hernández-Santoyo A, del Pozo Yauner L, *et al.* Structure and inactivation of triosephosphate isomerase from Entamoeba histolytica. J Mol Biol. 2002; 322: 669-675.

-Rogerson GW, Gutteridge WE. Catabolic metabolism in *Trypanosoma cruzi*. Int J Parasitol. 1980; 10: 131-135.

-Rost B. Protein structures sustain evolutionary drift. Fold Des. 1997; 2(3): S19-24.

-Sánchez R, Sali A. Evaluation of comparative protein structure modeling by MODELLER-3. PROTEINS: Structure, Function, and Genetics. 1997; 29(S1):50-58.

-Sanchez R, Pieper U, Melo F, *et al.* Protein structure modeling for structural genomics. Nat Struct Biol. 2000; 7 Suppl:986-90.

-Sanz-Rodriguez CE, Concepcion JL, Pekerar S, *et al.* Bisphosphonates as inhibitors of *Trypanosoma cruzi* hexokinase: kinetic and metabolic studies. J Biol Chem. 2007; 282: 12377-12387.

-Sattler R. Homology - a continuing challenge. Systematic Botany. 1984; 9 (4): 382–394.

-Schilling CH, Letscher D, Palsson BO. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. J Theor Biol. 2000; 203(3): 229-248.

-Schmunis GA. Epidemiology of Chagas disease in non-endemic countries: the role of international migration. Mem Inst Oswaldo Cruz. 2007; 102: 75-85.

-Selkov E Jr, Grechkin Y, Mikhailova N, Selkov E. MPW: the Metabolic Pathways Database. Nucleic Acids Res. 1998; 26(1): 43-45.

-Seo S, Lewin HA. Reconstruction of metabolic pathways for the cattle genome. BMC Syst Biol. 2009; 3: 33.

-Setúbal JC. A origem e o sentido da palavra bioinformática. Com Ciência – Revista Científica da SBPC. 2003. Disponível em: http://www.comciencia.br/reportagens/bioinformatica/bio10.shtml.

-Sgraja T, Smith TK, Hunter WN. Structure, substrate recognition and reactivity of *Leishmania major* mevalonate kinase. BMC Struct Biol 2007; 7: 20.

-Sharan R, Ulitsky I, Shamir R. Network-based prediction of protein function. Mol Syst Biol. 2007; 3: 88.

-Sigurdsson MI, Jamshidi N, Jonsson JJ, Palsson BO: Genome-scale network analysis of imprinted human metabolic genes. Epigenetics 2009; 4(1): 43-46.

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

-Silber AM, Rojas RL, Urias U, *et al.* Biochemical characterization of the glutamate transport in *Trypanosoma cruzi*. Int J Parasitol. 2006; 36: 157-163.

-Sjölander K. Phylogenomic inference of protein molecular function: advances and challenges. Bioinformatics. 2004; 20(2): 170-179.

-Skrabanek L, Saini HK, Bader GD, Enright AJ. Computational prediction of protein-protein interactions. Mol Biotechnol. 2008; 38(1):1-17.

-Snyder M, Gallagher JE. Systems biology from a yeast omics perspective. FEBS Lett. 2009; 583(24): 3895-3899.

-Soares MJ, de Souza W: Endocytosis of gold-labeled proteins and LDL by *Trypanosoma cruzi*. Parasitology Research. 1991; 77:461-468.

-Stein L. Genome annotation: from sequence to biology. Nat Rev Genet. 2001;2(7):493-503.

-Stephens JL, Lee SH, Paul KS, Englund PT. Mitochondrial fatty acid synthesis in *Trypanosoma brucei*. J Biol Chem. 2007; 282: 4427-4436.

-Stern AL, Burgos E, Salmon L, Cazzulo JJ. Ribose 5-phosphate isomerase type B from *Trypanosoma cruzi*: kinetic properties and site-directed mutagenesis reveal information about the reaction mechanism. Biochem J. 2007; 401: 279-285.

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

-Tarleton RL. Immune system recognition of *Trypanosoma cruzi*. Curr Opin Immunol. 2007; 19: 430-434.

-Taylor MB, Gutteridge WE. *Trypanosoma cruzi*: subcellular distribution of glycolytic and some related enzymes of epimastigotes. Expt Parasitol. 1987; 63: 84-97.

-Tetaud E, Bringaud F, Chabas S, *et al.* Characterization of glucose transport and cloning of a hexose transporter gene in *Trypanosoma cruzi*. Proc Natl Acad Sci USA. 1994; 91: 8278-8282.

-Thomas PD, Mi H, Lewis S. Ontology annotation: mapping genomic regions to biological function. Curr Opin Chem Biol. 2007; 11(1): 4-11.

-Thompson JD, Higgins DG, Gibson TJ: Clustal W. improving the sensitivity of progressive Multiple sequence alignment through sequence weighting, position-specific gap penalties and Weight matrix choice. Nucleic Acids Res. 1994; 22(22):4673-4680.

-Thompson JD, Gibson TJ, Plewniak F, *et al.* The ClustalX windows interface: exible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research. 1997; 24:4876-4882.

-Thornton JM. The future of bioinformatics. Trends Biotechnol. 1998; 16(1): 30-31.

-Tielens AG, Van Hellemond JJ. Differences in energy metabolism between trypanosomatidae. Parasitol Today. 1998; 14: 265-272.

-Todd MJ, Lorimer GH, Thirumalai D. Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. Proc Natl Acad Sci U S A. 1996; 93(9):4030-5

-Trapani S, Linss J, Goldenberg S, *et al.* Crystal structure of the dimeric phosphoenolpyruvate carboxykinase (PEPCK) from *Trypanosoma cruzi* at 2 A resolution. J Mol Biol. 2001; 313: 1059-1072.

-Tusnády GE, Simon I. Principles governing amino acid composition of integral membrane proteins: application to topology prediction. J Mol Biol. 1998; 283(2):489-506

-Urbina JA, Crespo A. Regulation of energy metabolism in *Trypanosoma* (Schizotrypanum) *cruzi* epimastigotes. I. Hexokinase and phosphofructokinase. Mol Biochem Parasitol. 1984; 11: 225-239.

-Urbina JA, Osorno CE, Rojas A. Inhibition of phosphoenolpyruvate carboxykinase from *Trypanosoma* (Schizotrypanum) *cruzi* epimastigotes by 3-mercaptopicolinic acid: in vitro and in vivo studies. Arch Biochem Biophys. 1990; 282: 91-99.

-Urbina JA, Machin I, Jurado L. The limitations of paradigms: studies on the

intermediary metabolism of *Trypanosoma cruzi*. Biol Res. 1993; 26: 81-88.

-Urbina JA. Lipid biosynthesis pathways as chemotherapeutic targets in kinetoplastid parasites. Parasitology. 1997; 114: S91-S99.

-Urbina JA, Docampo R. Specific chemotherapy of Chagas disease: controversies and advances. Trends Parasitol. 2003; 19: 495-501.

-Valas RE, Yang S, Bourne PE. Nothing about protein structure classification makes sense except in the light of evolution. Curr Opin Struct Biol. 2009; 19(3): 329-334.

-Valencia A. Automatic annotation of protein function. Curr Opin Struct Biol. 2005; 15(3):267-74.

-van Hellemond JJ, Opperdoes FR, Tielens AG. The extraordinary mitochondrion and unusual citric acid cycle in *Trypanosoma brucei*. Biochem Soc Trans. 2005; 33: 967-971.

-van Hellemond JJ, Tielens AG. Adaptations in the lipid metabolism of the protozoan parasite *Trypanosoma brucei*. FEBS Letters. 2006; 580(23):5552-5558.

-van Weelden SW, van Hellemond JJ, Opperdoes FR, Tielens AG. New functions for parts of the Krebs cycle in procyclic *Trypanosoma brucei*, a cycle not operating as a cycle. J Biol Chem. 2005; 280: 12451-12460.

-Vitkup D, Kharchenko P, Wagner A. Influence of metabolic network structure and function on enzyme evolution. Genome Biol. 2006; 7(5): R39.

-Von Brand T, Agosin M. The utilization of Krebs cycle intermediates by the culture forms of *Trypanosoma cruzi* and *Leishmania tropica*. J Infect Dis. 1955; 97: 274-279.

-Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature. 1953; 171(4356): 737-738.

-Watson JD, Laskowski RA, Thornton JM. Predicting protein function from sequence and structural data. Curr Opin Struct Biol. 2005; 15(3):275-84

-Westbrook J, Feng Z, Jain S, *et al.* The Protein Data Bank: unifying the archive. Nucleic Acids Research. 2002; 30:245-248.

-Wiemer EA, IJlst L, van Roy J, *et al.* Identification of 2-enoyl coenzyme A hydratase and NADP(+)- dependent 3-hydroxyacyl-CoA dehydrogenase activity in glycosomes of procyclic *Trypanosoma brucei*. Mol Biochem Parasitol. 1996; 82: 107-111.

-Wiesner J, Jomaa H. Isoprenoid biosynthesis of the apicoplast as drug target. Curr Drug Targets. 2007; 8: 3-13.

-Wilkinson SR, Taylor MC, Horn D, *et al.* A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. PNAS. 2008; 105(13):5022-5027.

-Williams DA, Lemke TL. Foye´s Principles of Medicinal Chemistry. 5 ed. Lippincott Williams & Wilkins; 2002.

-Woese C. The universal ancestor. Proc Natl Acad Sci U S A. 1998; 95(12): 6854-6859.

-Wright CS, Alden RA, Kraut J. Structure of subtilisin BPN' at 2.5 angström resolution. Nature. 1969; 221(5177):235-42.

-Wu J, Mao X, Cai T, *et al.* KOBAS server: a web-based platform for automated annotation And pathway identification. Nucleic Acids Res. 2006; 1:720-724.

-Xu EY, Schaefer WH, Xu Q. Metabolomics in pharmaceutical research and development: metabolites, mechanisms and pathways. Curr Opin Drug Discov Devel. 2009; 12(1): 40-52.

-Ycas M. On earlier states of the biochemical system. J Theor Biol 1974; 44: 145–160.

# *Anexos*

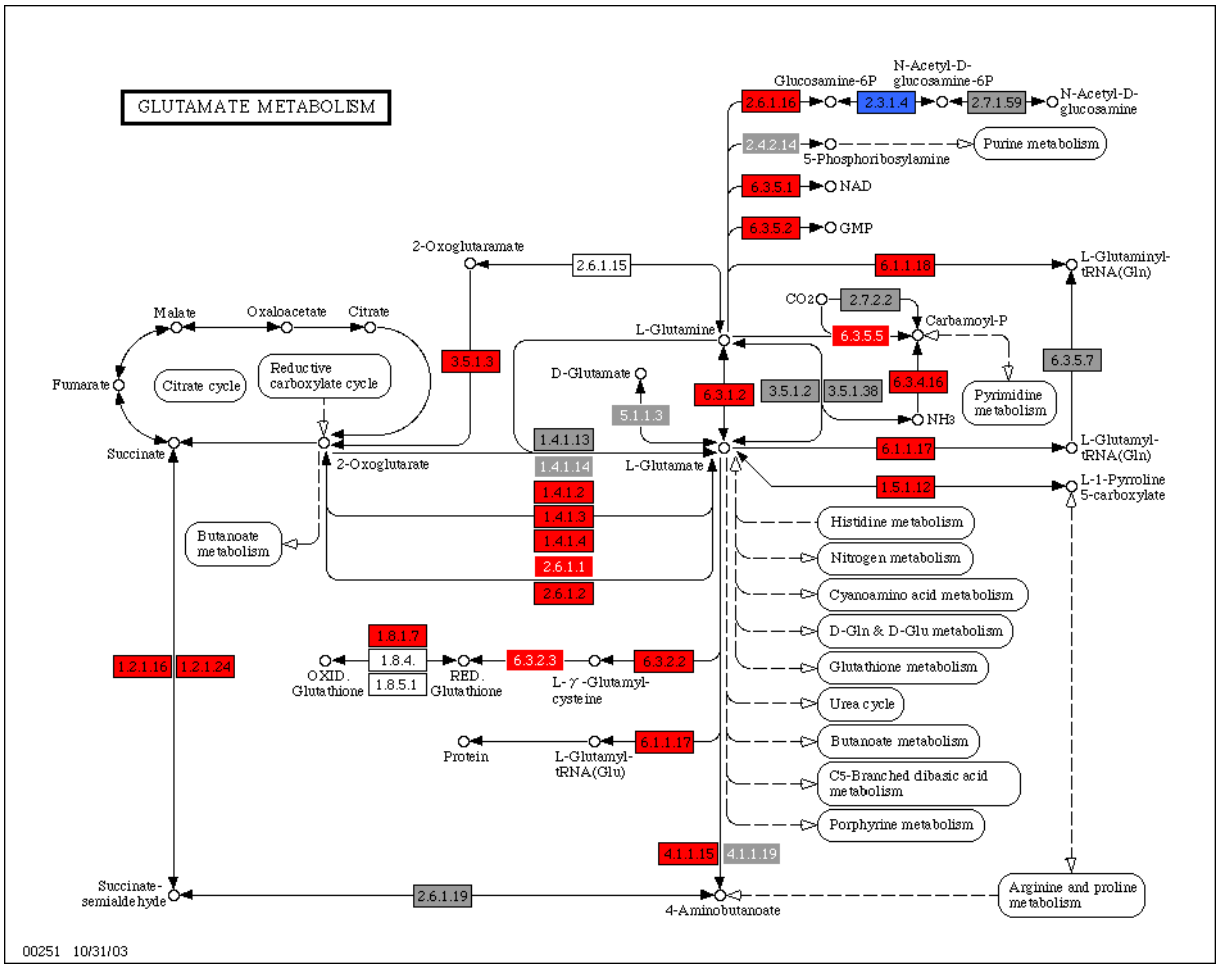
**Anexo I:** Material Suplementar referente ao artigo "*In silico* reconstruction of the amino acid metabolic pathways of *Trypanosoma cruzi*". Estes dados encontram-se também disponíveis em: http://www.dbbm.fiocruz.br/labwim/TcruziAA/.

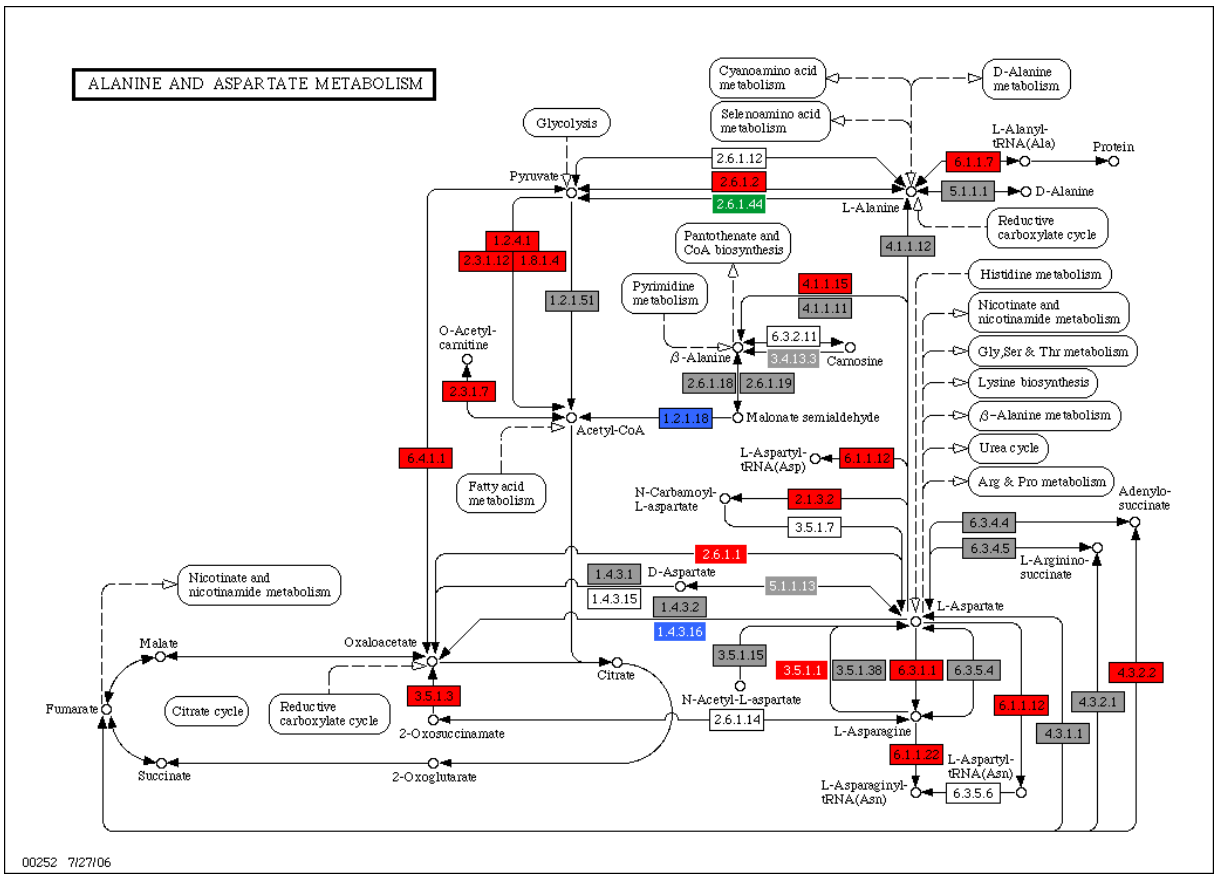
A) Lista dos mapas representativos do metabolismo de aminoácidos provenientes do KEGG, no qual foram feitas as reconstruções metabólicas para *T. cruzi,* utilizando o *pipeline* do AnEnPi.

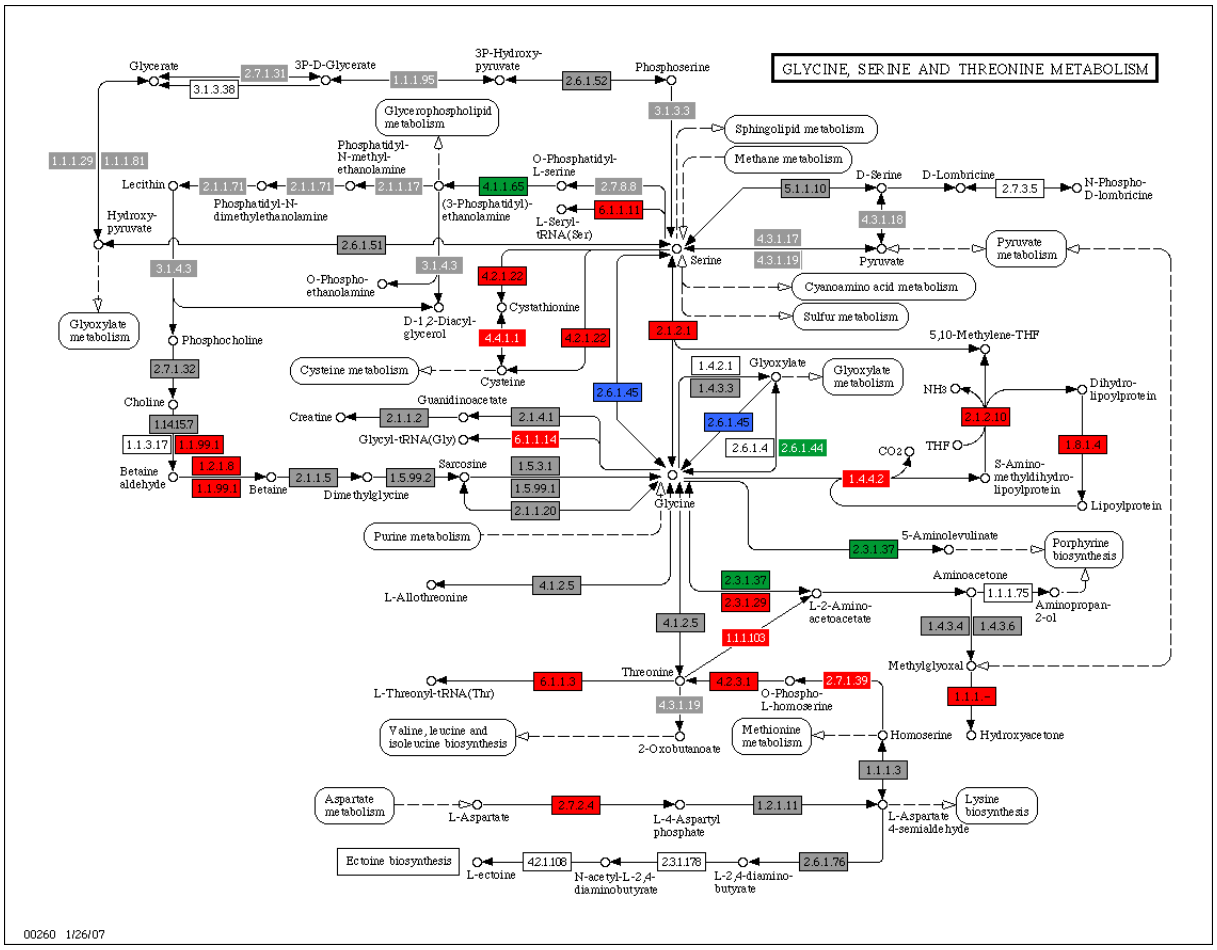
| Número do mapa no KEGG | Descrição do mapa no KEGG |
|---|---|
| 251 | Glutamate metabolism |
| 252 | Alanine and aspartate metabolism |
| 260 | Glycine, serine and threonine metabolism |
| 271 | Methionine metabolism |
| 272 | Cysteine metabolism |
| 280 | Valine, leucine and isoleucine degradation |
| 290 | Valine, leucine and isoleucine biosynthesis |
| 300 | Lysine biosynthesis |
| 310 | Lysine degradation |
| 330 | Arginine and proline metabolism |
| 340 | Histidine metabolism |
| 350 | Tyrosine metabolism |
| 360 | Phenylalanine metabolism |
| 380 | Tryptophan metabolism |
| 400 | Phenylalanine, tyrosine and tryptophan biosynthesis |
| 220 | Urea cycle and metabolism of amino groups |
| 410 | beta-Alanine metabolism |
| 430 | Taurine and hypotaurine metabolism |
| 440 | Aminophosphonate metabolism |
| 450 | Selenoamino acid metabolism |
| 460 | Cyanoamino acid metabolism |
| 471 | D-Glutamine and D-glutamate metabolism |
| 472 | D-Arginine and D-ornithine metabolism |
| 473 | D-Alanine metabolism |
| 480 | Glutathione metabolism |

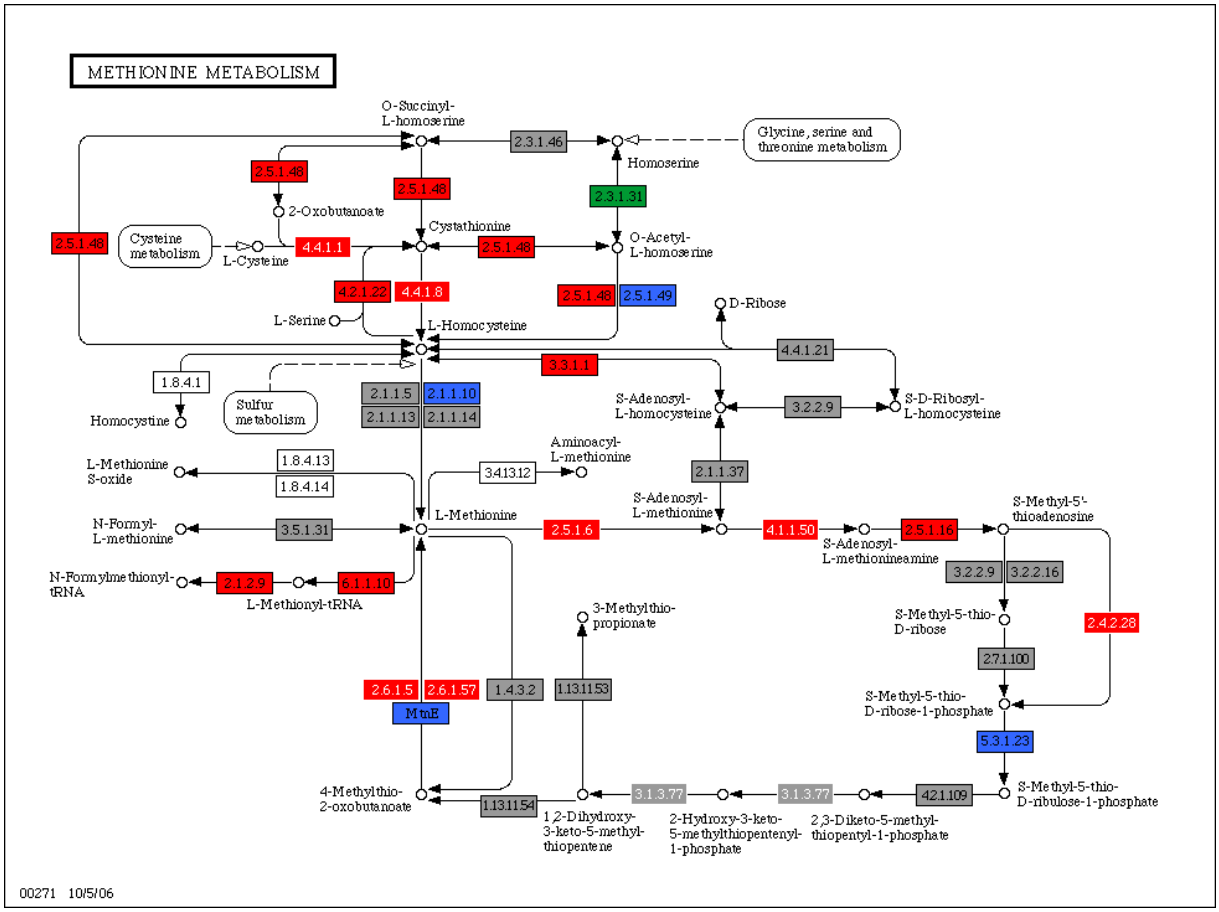
B) Mapas da reconstrução metabólica para os 25 mapas representativos do metabolismo de aminoácidos de *T. cruzi*.

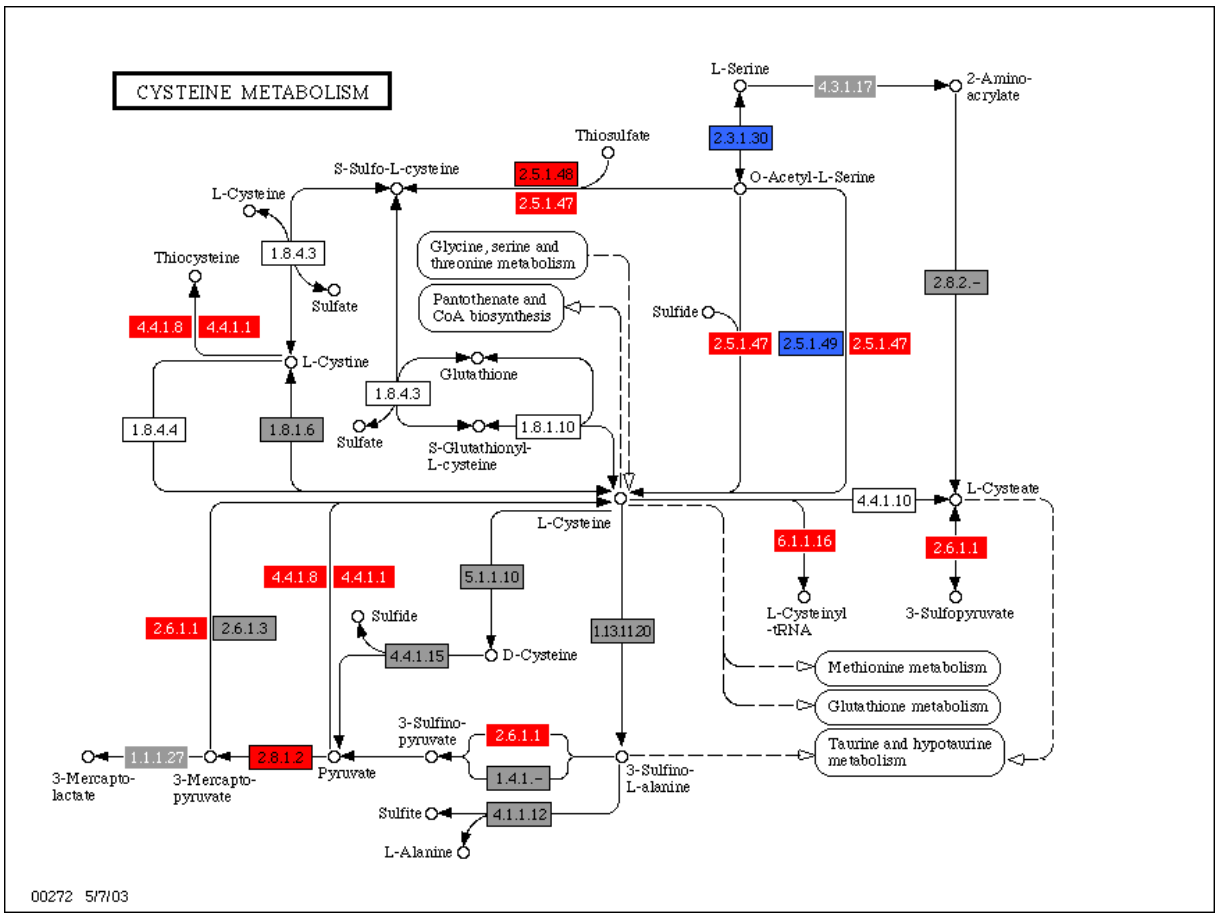
Legenda das figuras: Reconstrução metabólica das funções enzimáticas de *T.cruzi*, identificados pelo AnEnPi. As cores de fundo das caixas dos *EC numbers* representam: em branco, atividades enzimáticas sem representantes no KEGG; em cinza, não identificados em *T. cruzi*, mas com representantes no KEGG para outros organismos; coloridos, identificados em *T. cruzi* com o AnEnPi com diferentes *e-values* como ponto de corte valor, sendo $e^{-80}$ (vermelho), $e^{-60}$ (azul), e $e^{-40}$ (verde). Os algaritmos em branco representam atividades enzimáticas com possíveis casos de analogia.

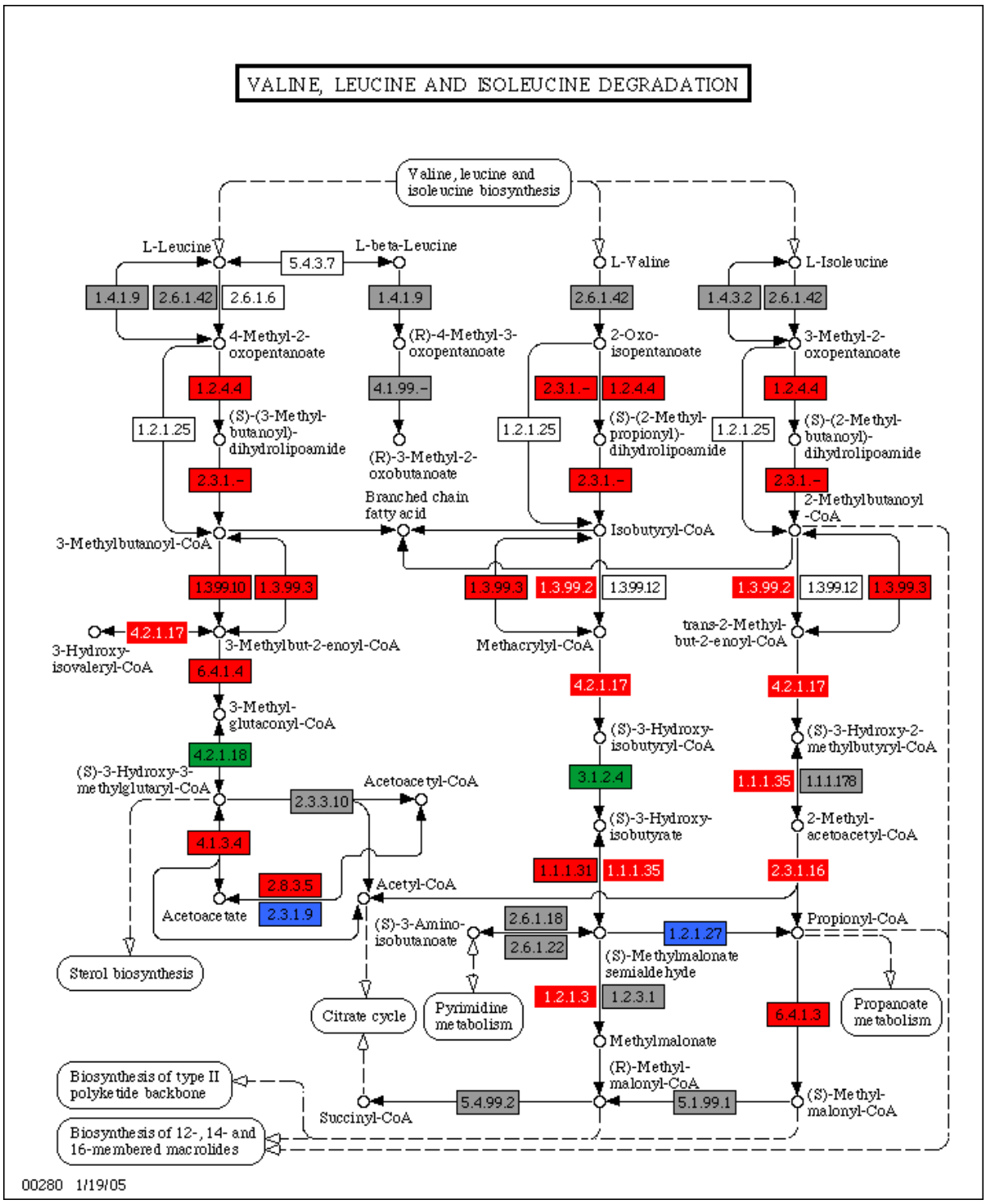


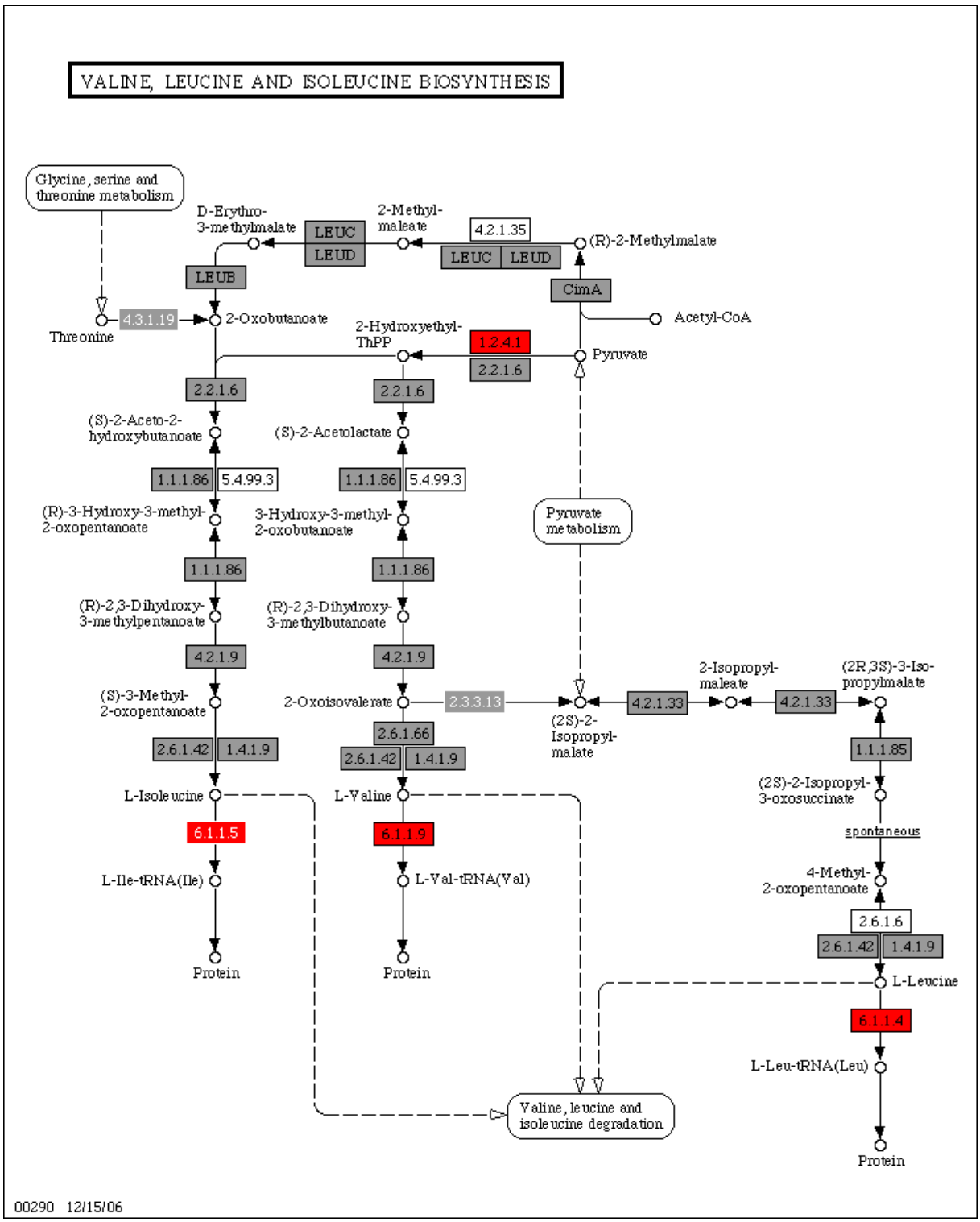

ALANINE AND ASPARTATE METABOLISM
Glycolysis
Pyruvate
Cyanoamino acid metabolism
Selenoamino acid metabolism
D-Alanine metabolism
L-Alanyl-tRNA(Ala)
Protein
L-Alanine
D-Alanine
Reductive carboxylate cycle
Pantothenate and CoA biosynthesis
Pyrimidine metabolism
Histidine metabolism
Nicotinate and nicotinamide metabolism
Gly,Ser & Thr metabolism
Lysine biosynthesis
β-Alanine metabolism
Urea cycle
Arg & Pro metabolism
O-Acetyl-carnitine
β-Alanine
Carnosine
Malonate semialdehyde
Acetyl-CoA
Fatty acid metabolism
L-Aspartyl-tRNA(Asp)
N-Carbamoyl-L-aspartate
Adenylo-succinate
L-Arginino-succinate
D-Aspartate
L-Aspartate
Nicotinate and nicotinamide metabolism
Malate
Oxaloacetate
Citrate
Fumarate
Citrate cycle
Reductive carboxylate cycle
2-Oxosuccinamate
N-Acetyl-L-aspartate
L-Asparagine
L-Aspartyl-tRNA(Asn)
L-Asparaginyl-tRNA(Asn)
Succinate
2-Oxoglutarate
00252 7/27/06

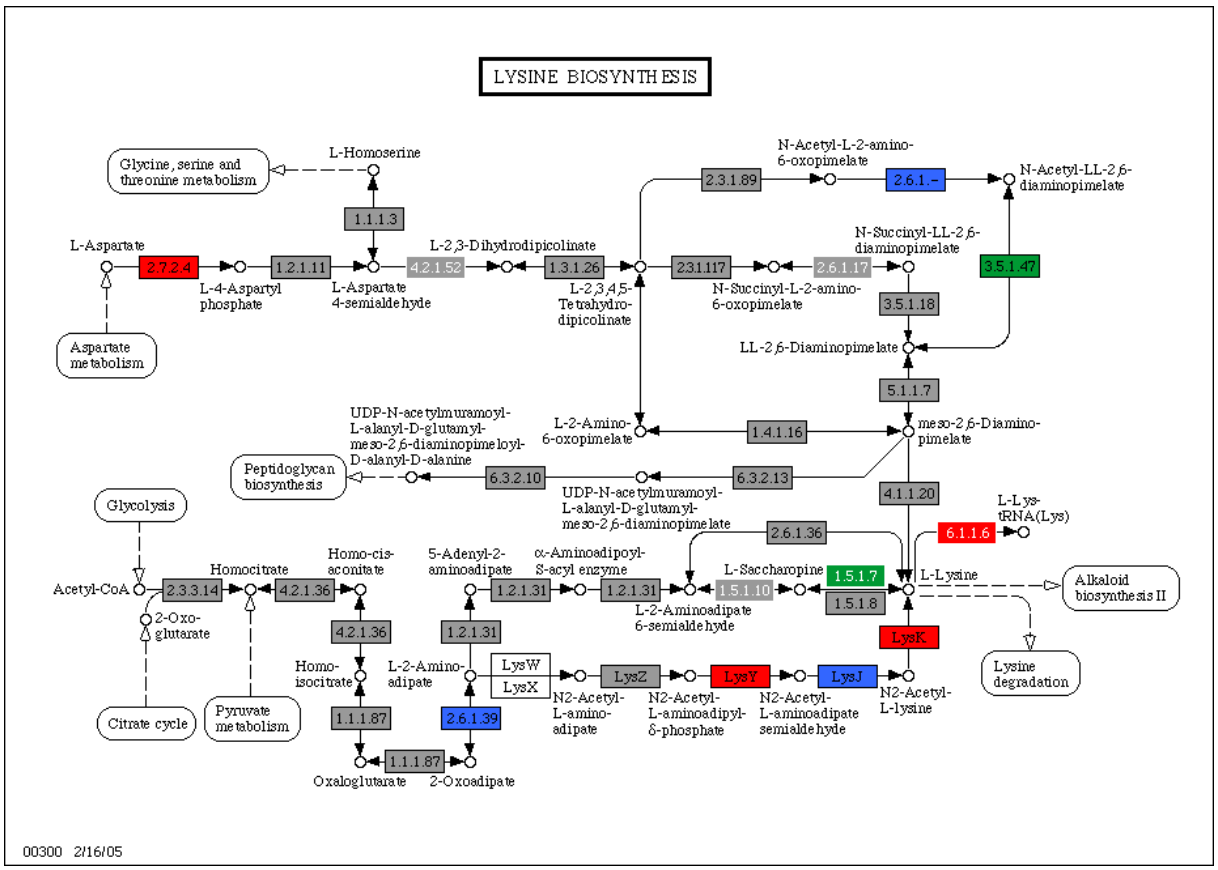

GLYCINE, SERINE AND THREONINE METABOLISM
Glycerate
3P-D-Glycerate
3P-Hydroxy-pyruvate
Phosphoserine
Glycerophospholipid metabolism
Sphingolipid metabolism
Methane metabolism
Phosphatidyl-N-methyl-ethanolamine
O-Phosphatidyl-L-serine
Lecithin
Phosphatidyl-N-dimethylethanolamine
(3-Phosphatidyl)-ethanolamine
L-Seryl-tRNA(Ser)
D-Serine
D-Lombricine
N-Phospho-D-lombricine
Hydroxy-pyruvate
Serine
Pyruvate
Pyruvate metabolism
O-Phospho-ethanolamine
Cystathionine
D-1,2-Diacyl-glycerol
Cyanoamino acid metabolism
Sulfur metabolism
Glyoxylate metabolism
Phosphocholine
Cysteine metabolism
Cysteine
5,10-Methylene-THF
Glyoxylate
Choline
Guanidinoacetate
Creatine
Glycyl-tRNA(Gly)
NH3
Dihydro-lipoylprotein
THF
CO2
S-Amino-methyldihydro-lipoylprotein
Lipoylprotein
Betaine aldehyde
Betaine
Dimethylglycine
Sarcosine
Glycine
Purine metabolism
5-Aminolevulinate
Porphyrine biosynthesis
L-Allothreonine
Aminoacetone
L-2-Amino-acetoacetate
Aminopropan-2-ol
Methylglyoxal
Threonine
L-Threonyl-tRNA(Thr)
O-Phospho-L-homoserine
Valine, leucine and isoleucine biosynthesis
2-Oxobutanoate
Methionine metabolism
Homoserine
Hydroxyacetone
Aspartate metabolism
L-Aspartate
L-4-Aspartyl phosphate
L-Aspartate 4-semialdehyde
Lysine biosynthesis
Ectoine biosynthesis
L-ectoine
N-acetyl-L-2,4-diaminobutyrate
L-2,4-diamino-butyrate
00260 1/26/07

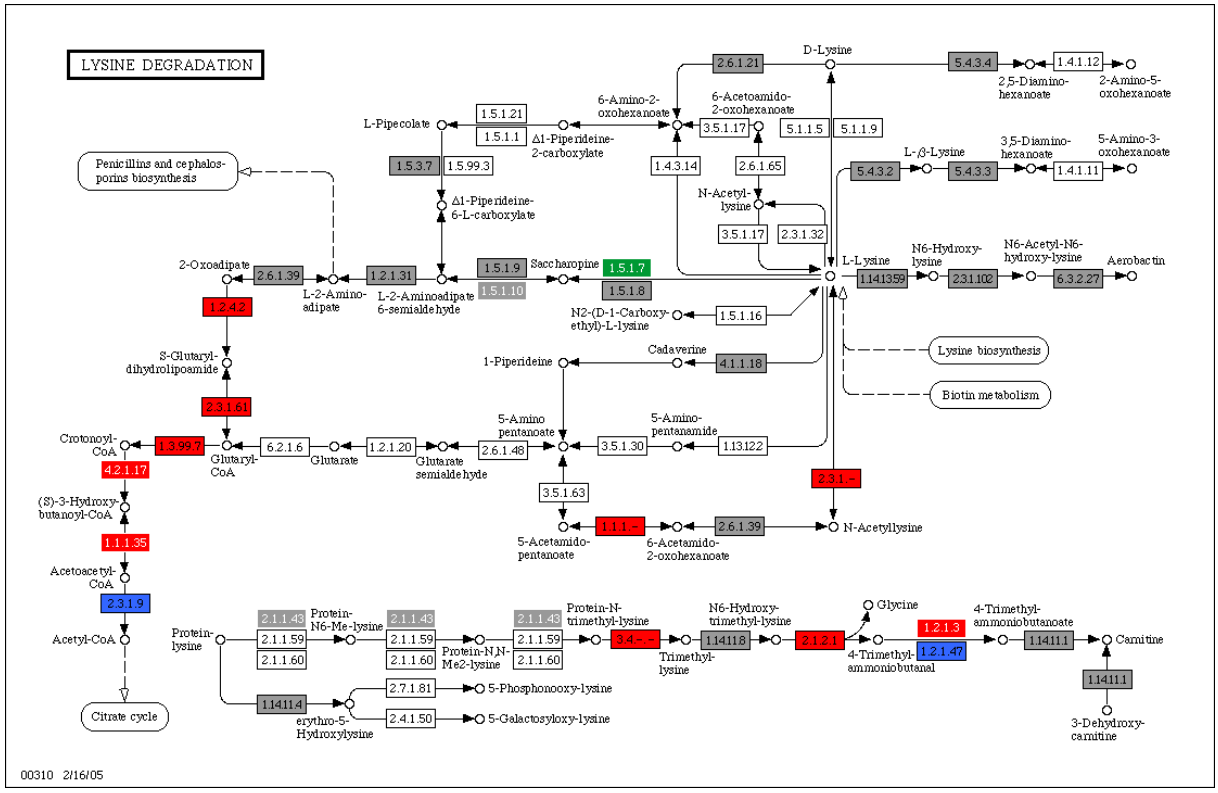

METHIONINE METABOLISM
O-Succinyl-L-homoserine
2.3.1.46
Glycine, serine and threonine metabolism
Homoserine
2.5.1.48
2.5.1.48
2.3.1.31
2-Oxobutanoate
Cystathionine
2.5.1.48
Cysteine metabolism
4.4.1.1
L-Cysteine
2.5.1.48
O-Acetyl-L-homoserine
4.2.1.22
4.4.1.8
2.5.1.48
2.5.1.49
D-Ribose
L-Serine
L-Homocysteine
4.4.1.21
3.3.1.1
1.8.4.1
Homocystine
Sulfur metabolism
2.1.1.5
2.1.1.10
2.1.1.13
2.1.1.14
S-Adenosyl-L-homocysteine
3.2.2.9
S-D-Ribosyl-L-homocysteine
Aminoacyl-L-methionine
L-Methionine S-oxide
1.8.4.13
1.8.4.14
3.4.13.12
2.1.1.37
S-Adenosyl-L-methionine
S-Methyl-5'-thioadenosine
N-Formyl-L-methionine
3.5.1.31
L-Methionine
2.5.1.6
4.1.1.50
2.5.1.16
S-Adenosyl-L-methionineamine
3.2.2.9
3.2.2.16
N-Formylmethionyl-tRNA
2.1.2.9
6.1.1.10
L-Methionyl-tRNA
3-Methylthio-propionate
S-Methyl-5-thio-D-ribose
2.4.2.28
2.7.1.100
2.6.1.5
2.6.1.57
MtnE
1.4.3.2
1.13.11.53
S-Methyl-5-thio-D-ribose-1-phosphate
5.3.1.23
4-Methylthio-2-oxobutanoate
1.13.11.54
3.1.3.77
3.1.3.77
4.2.1.109
S-Methyl-5-thio-D-ribulose-1-phosphate
1,2-Dihydroxy-3-keto-5-methyl-thiopentene
2-Hydroxy-3-keto-5-methylthiopentenyl-1-phosphate
2,3-Diketo-5-methyl-thiopentyl-1-phosphate
00271 10/5/06

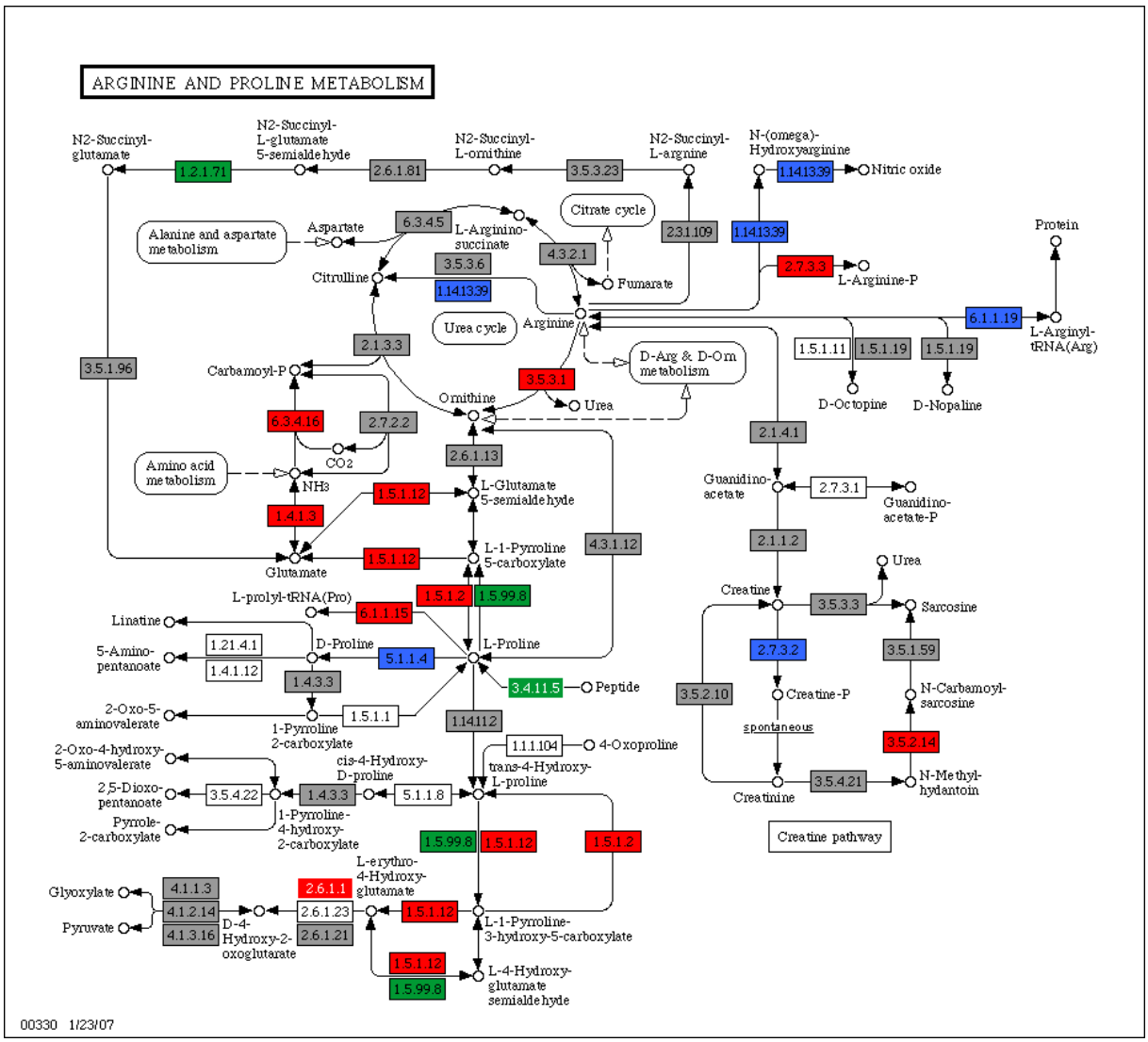

CYSTEINE METABOLISM
L-Serine
4.3.1.17
2-Amino-acrylate
Thiosulfate
2.3.1.30
S-Sulfo-L-cysteine
2.5.1.48
2.5.1.47
O-Acetyl-L-Serine
L-Cysteine
Thiocysteine
1.8.4.3
Sulfate
Glycine, serine and threonine metabolism
Pantothenate and CoA biosynthesis
2.8.2.-
Sulfide
4.4.1.8
4.4.1.1
2.5.1.47
2.5.1.49
2.5.1.47
L-Cystine
Glutathione
1.8.4.3
1.8.4.4
1.8.1.6
Sulfate
1.8.1.10
S-Glutathionyl-L-cysteine
4.4.1.10
L-Cysteate
L-Cysteine
6.1.1.16
2.6.1.1
4.4.1.8
4.4.1.1
5.1.1.10
L-Cysteinyl-tRNA
3-Sulfopyruvate
2.6.1.1
2.6.1.3
Sulfide
1.13.11.20
4.4.1.15
D-Cysteine
Methionine metabolism
Glutathione metabolism
3-Sulfino-pyruvate
2.6.1.1
Taurine and hypotaurine metabolism
1.1.1.27
2.8.1.2
1.4.1.-
3-Sulfino-L-alanine
3-Mercapto-lactate
3-Mercapto-pyruvate
Pyruvate
Sulfite
4.1.1.12
L-Alanine
00272 5/7/03

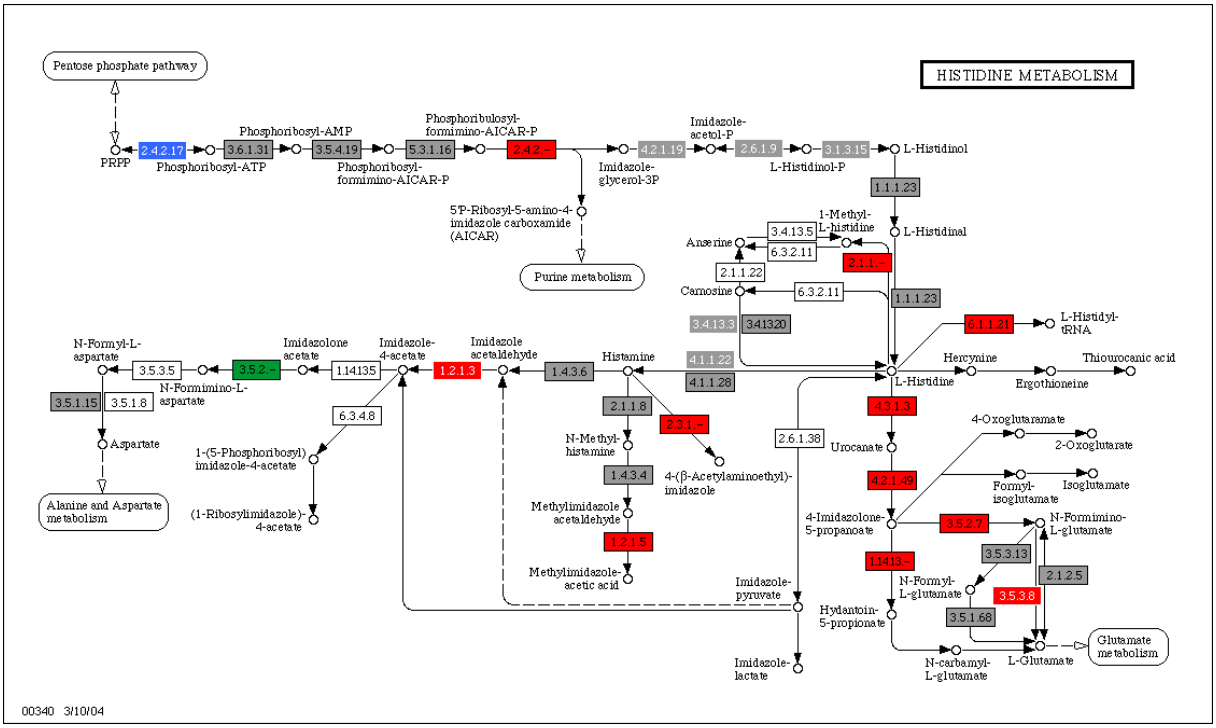

VALINE, LEUCINE AND ISOLEUCINE DEGRADATION
Valine, leucine and isoleucine biosynthesis
L-Leucine
5.4.3.7
L-beta-Leucine
L-Valine
L-Isoleucine
1.4.1.9
2.6.1.42
2.6.1.6
1.4.1.9
2.6.1.42
1.4.3.2
2.6.1.42
4-Methyl-2-oxopentanoate
(R)-4-Methyl-3-oxopentanoate
2-Oxo-isopentanoate
3-Methyl-2-oxopentanoate
1.2.4.4
4.1.99.-
2.3.1.-
1.2.4.4
1.2.4.4
1.2.1.25
(S)-(3-Methyl-butanoyl)-dihydrolipoamide
1.2.1.25
(S)-(2-Methyl-propionyl)-dihydrolipoamide
1.2.1.25
(S)-(2-Methyl-butanoyl)-dihydrolipoamide
2.3.1.-
(R)-3-Methyl-2-oxobutanoate
2.3.1.-
2.3.1.-
Branched chain fatty acid
2-Methylbutanoyl-CoA
3-Methylbutanoyl-CoA
Isobutyryl-CoA
1.3.99.10
1.3.99.3
1.3.99.3
1.3.99.2
1.3.99.12
1.3.99.2
1.3.99.12
1.3.99.3
4.2.1.17
3-Hydroxy-isovaleryl-CoA
3-Methylbut-2-enoyl-CoA
Methacrylyl-CoA
trans-2-Methyl-but-2-enoyl-CoA
6.4.1.4
4.2.1.17
4.2.1.17
3-Methyl-glutaconyl-CoA
(S)-3-Hydroxy-isobutyryl-CoA
(S)-3-Hydroxy-2-methylbutyryl-CoA
4.2.1.18
3.1.2.4
1.1.1.35
1.1.1.178
(S)-3-Hydroxy-3-methylglutaryl-CoA
2.3.3.10
Acetoacetyl-CoA
(S)-3-Hydroxy-isobutyrate
2-Methyl-acetoacetyl-CoA
4.1.3.4
1.1.1.31
1.1.1.35
2.3.1.16
2.8.3.5
Acetyl-CoA
Acetoacetate
2.3.1.9
(S)-3-Amino-isobutanoate
2.6.1.18
2.6.1.22
1.2.1.27
Propionyl-CoA
(S)-Methylmalonate semialdehyde
Sterol biosynthesis
Citrate cycle
Pyrimidine metabolism
1.2.1.3
1.2.3.1
6.4.1.3
Propanoate metabolism
Methylmalonate
(R)-Methyl-malonyl-CoA
Biosynthesis of type II polyketide backbone
5.4.99.2
5.1.99.1
(S)-Methyl-malonyl-CoA
Succinyl-CoA
Biosynthesis of 12-, 14- and 16-membered macrolides
00280 1/19/05

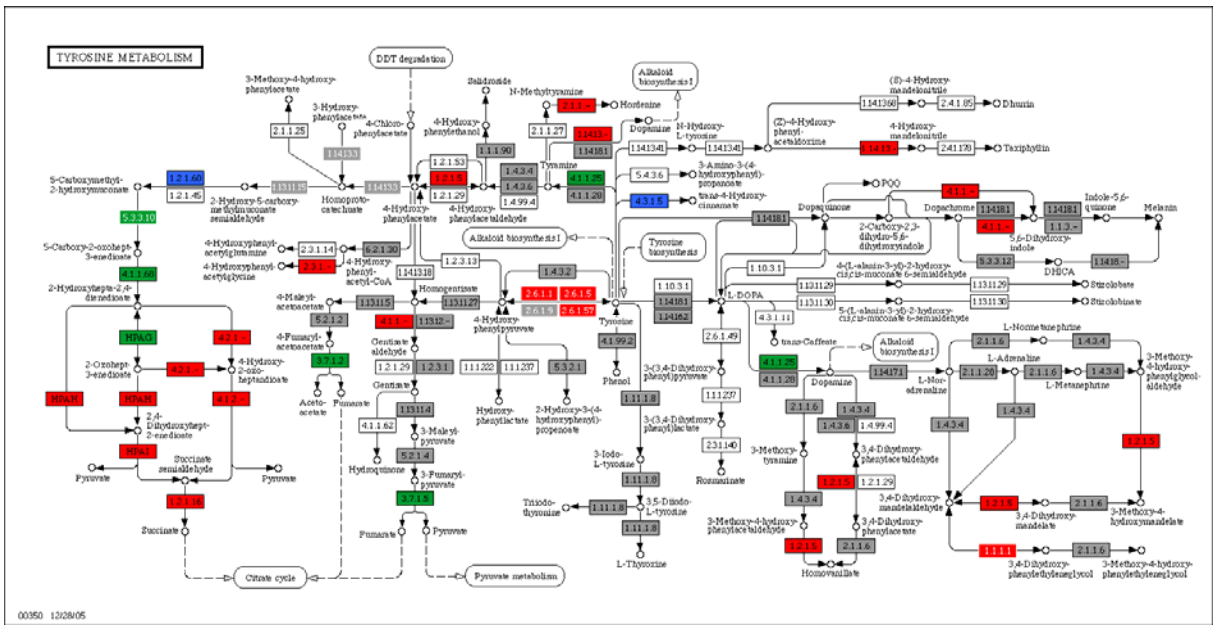

VALINE, LEUCINE AND ISOLEUCINE BIOSYNTHESIS
Glycine, serine and threonine metabolism
D-Erythro-3-methylmalate
LEUC
LEUD
2-Methyl-maleate
4.2.1.35
(R)-2-Methylmalate
LEUB
CimA
Acetyl-CoA
Threonine
4.3.1.19
2-Oxobutanoate
2-Hydroxyethyl-ThPP
1.2.4.1
2.2.1.6
Pyruvate
(S)-2-Aceto-2-hydroxybutanoate
(S)-2-Acetolactate
1.1.1.86
5.4.99.3
(R)-3-Hydroxy-3-methyl-2-oxopentanoate
3-Hydroxy-3-methyl-2-oxobutanoate
Pyruvate metabolism
(R)-2,3-Dihydroxy-3-methylpentanoate
(R)-2,3-Dihydroxy-3-methylbutanoate
4.2.1.9
(S)-3-Methyl-2-oxopentanoate
2-Oxoisovalerate
2.3.3.13
(2S)-2-Isopropyl-malate
2-Isopropyl-maleate
(2R,3S)-3-Iso-propylmalate
2.6.1.66
2.6.1.42
1.4.1.9
1.1.1.85
L-Isoleucine
L-Valine
(2S)-2-Isopropyl-3-oxosuccinate
6.1.1.5
6.1.1.9
spontaneous
L-Ile-tRNA(Ile)
L-Val-tRNA(Val)
4-Methyl-2-oxopentanoate
2.6.1.6
Protein
L-Leucine
6.1.1.4
L-Leu-tRNA(Leu)
Valine, leucine and isoleucine degradation
00290 12/15/06

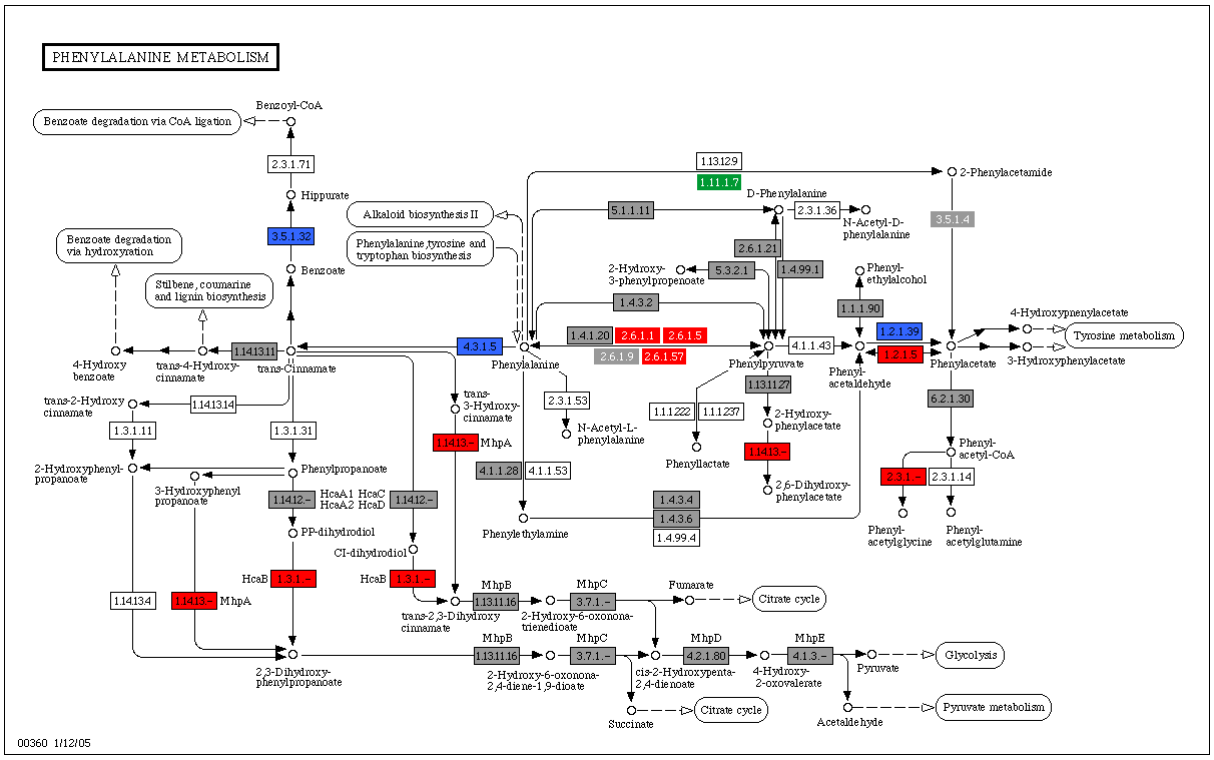

LYSINE BIOSYNTHESIS
Glycine, serine and threonine metabolism
L-Homoserine
1.1.1.3
L-Aspartate
2.7.2.4
L-4-Aspartyl phosphate
1.2.1.11
L-Aspartate 4-semialdehyde
4.2.1.52
L-2,3-Dihydrodipicolinate
1.3.1.26
L-2,3,4,5-Tetrahydro-dipicolinate
2.3.1.89
N-Acetyl-L-2-amino-6-oxopimelate
2.6.1.-
N-Acetyl-LL-2,6-diaminopimelate
2.3.1.117
N-Succinyl-L-2-amino-6-oxopimelate
2.6.1.17
N-Succinyl-LL-2,6-diaminopimelate
3.5.1.47
3.5.1.18
Aspartate metabolism
LL-2,6-Diaminopimelate
5.1.1.7
L-2-Amino-6-oxopimelate
1.4.1.16
meso-2,6-Diamino-pimelate
UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimeloyl-D-alanyl-D-alanine
Peptidoglycan biosynthesis
6.3.2.10
6.3.2.13
UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-meso-2,6-diaminopimelate
4.1.1.20
Glycolysis
2.6.1.36
6.1.1.6
L-Lys-tRNA(Lys)
Acetyl-CoA
Homocitrate
2.3.3.14
4.2.1.36
Homo-cis-aconitate
5-Adenyl-2-aminoadipate
1.2.1.31
α-Aminoadipoyl-S-acyl enzyme
1.2.1.31
L-Saccharopine
1.5.1.7
1.5.1.10
1.5.1.8
L-Lysine
Alkaloid biosynthesis II
2-Oxo-glutarate
L-2-Aminoadipate 6-semialdehyde
LysK
Lysine degradation
Homo-isocitrate
L-2-Amino-adipate
LysW
LysX
LysZ
LysY
LysJ
N2-Acetyl-L-amino-adipate
N2-Acetyl-L-aminoadipyl-δ-phosphate
N2-Acetyl-L-aminoadipate semialdehyde
N2-Acetyl-L-lysine
Citrate cycle
Pyruvate metabolism
1.1.1.87
2.6.1.39
1.1.1.87
Oxaloglutarate
2-Oxoadipate
00300 2/16/05

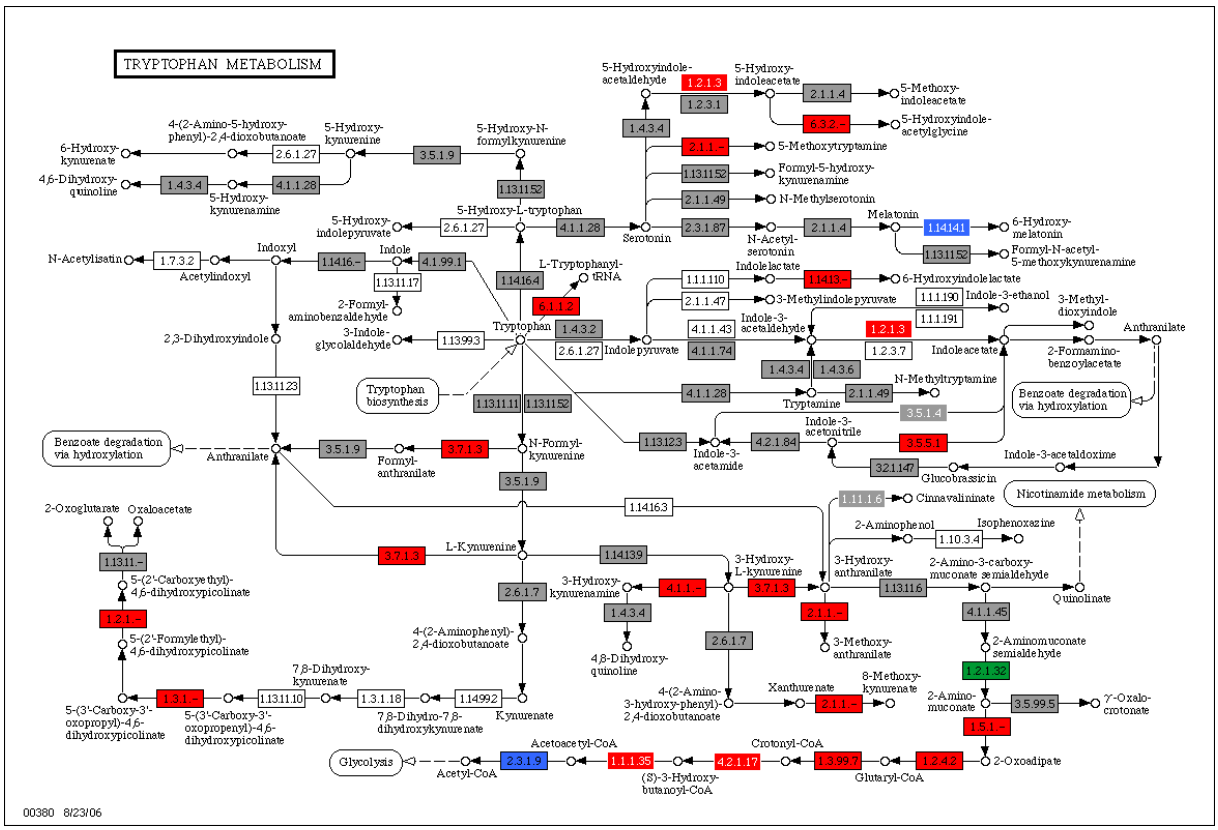

LYSINE DEGRADATION
D-Lysine
2.6.1.21
5.4.3.4
1.4.1.12
2,5-Diamino-hexanoate
2-Amino-5-oxohexanoate
6-Amino-2-oxohexanoate
6-Acetoamido-2-oxohexanoate
L-Pipecolate
1.5.1.21
1.5.1.1
Δ1-Piperideine-2-carboxylate
3.5.1.17
5.1.1.5
5.1.1.9
L-β-Lysine
3,5-Diamino-hexanoate
5-Amino-3-oxohexanoate
1.5.3.7
1.5.99.3
1.4.3.14
2.6.1.65
5.4.3.2
5.4.3.3
1.4.1.11
Penicillins and cephalosporins biosynthesis
Δ1-Piperideine-6-L-carboxylate
N-Acetyl-lysine
3.5.1.17
2.3.1.32
2-Oxoadipate
2.6.1.39
1.2.1.31
L-2-Amino-adipate
L-2-Aminoadipate 6-semialdehyde
1.5.1.9
1.5.1.10
Saccharopine
1.5.1.7
1.5.1.8
L-Lysine
1.14.13.59
N6-Hydroxy-lysine
2.3.1.102
N6-Acetyl-N6-hydroxy-lysine
6.3.2.27
Aerobactin
1.2.4.2
N2-(D-1-Carboxy-ethyl)-L-lysine
1.5.1.16
S-Glutaryl-dihydrolipoamide
Cadaverine
1-Piperideine
4.1.1.18
Lysine biosynthesis
2.3.1.61
Biotin metabolism
Crotonoyl-CoA
1.3.99.7
Glutaryl-CoA
6.2.1.6
Glutarate
1.2.1.20
Glutarate semialdehyde
2.6.1.48
5-Amino pentanoate
3.5.1.30
5-Amino-pentanamide
1.13.12.2
4.2.1.17
3.5.1.63
2.3.1.-
(S)-3-Hydroxy-butanoyl-CoA
1.1.1.35
5-Acetamido-pentanoate
1.1.1.-
6-Acetamido-2-oxohexanoate
2.6.1.39
N-Acetyllysine
Acetoacetyl-CoA
2.3.1.9
Acetyl-CoA
Protein-lysine
2.1.1.43
2.1.1.59
2.1.1.60
Protein-N6-Me-lysine
Protein-N,N-Me2-lysine
Protein-N-trimethyl-lysine
3.4.-.-
N6-Hydroxy-trimethyl-lysine
1.14.11.8
Trimethyl-lysine
2.1.2.1
Glycine
4-Trimethyl-ammoniobutanal
1.2.1.3
1.2.1.47
4-Trimethyl-ammoniobutanoate
1.14.11.1
Carnitine
1.14.11.1
3-Dehydroxy-carnitine
Citrate cycle
1.14.11.4
erythro-5-Hydroxylysine
2.7.1.81
5-Phosphonooxy-lysine
2.4.1.50
5-Galactosyloxy-lysine
00310 2/16/05

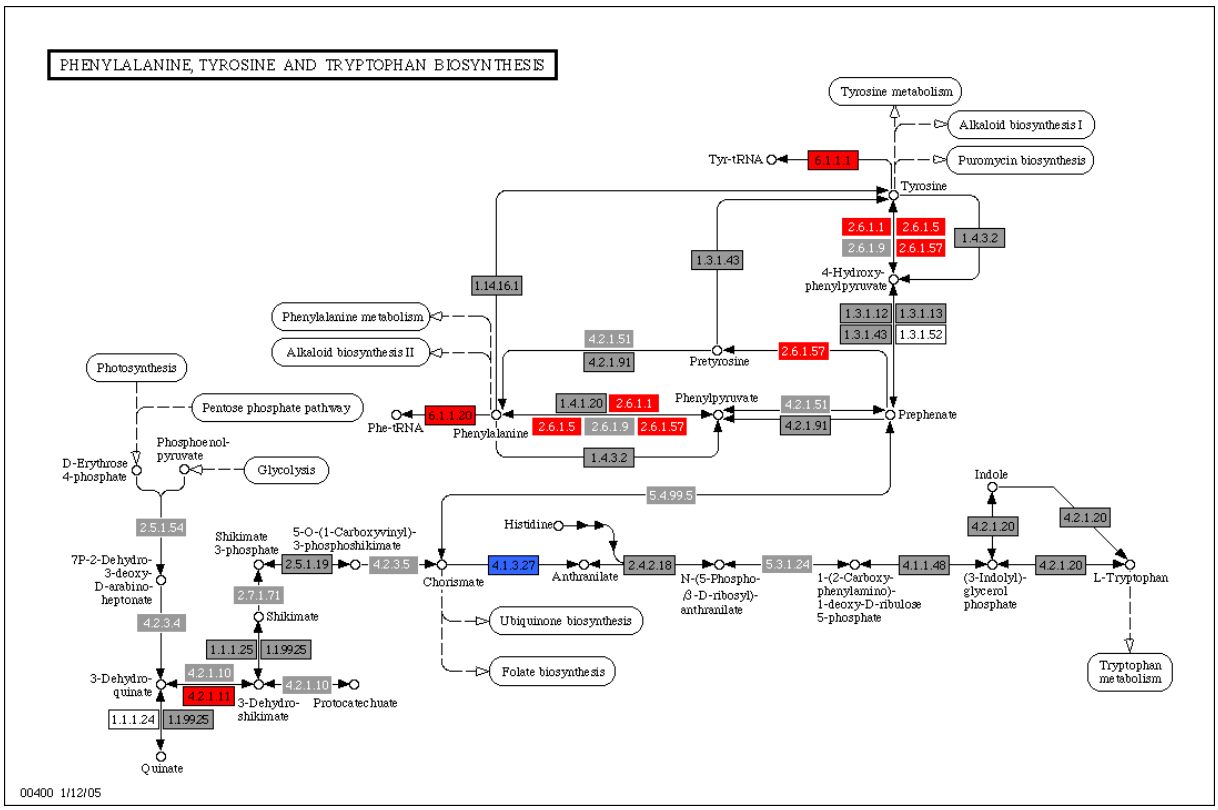

ARGININE AND PROLINE METABOLISM
N2-Succinyl-glutamate
N2-Succinyl-L-glutamate 5-semialdehyde
N2-Succinyl-L-ornithine
N2-Succinyl-L-arginine
N-(omega)-Hydroxyarginine
Nitric oxide
Alanine and aspartate metabolism
Aspartate
L-Arginino-succinate
Citrate cycle
Fumarate
Citrulline
Urea cycle
Arginine
Protein
L-Arginyl-tRNA(Arg)
L-Arginine-P
D-Arg & D-Orn metabolism
D-Octopine
D-Nopaline
Carbamoyl-P
Ornithine
Urea
CO2
Amino acid metabolism
NH3
L-Glutamate 5-semialdehyde
Glutamate
L-1-Pyrroline 5-carboxylate
L-prolyl-tRNA(Pro)
Linatine
5-Amino-pentanoate
D-Proline
L-Proline
Peptide
2-Oxo-5-aminovalerate
1-Pyrroline 2-carboxylate
4-Oxoproline
2-Oxo-4-hydroxy-5-aminovalerate
cis-4-Hydroxy-D-proline
trans-4-Hydroxy-L-proline
2,5-Dioxo-pentanoate
Pyrrole-2-carboxylate
1-Pyrroline-4-hydroxy-2-carboxylate
L-erythro-4-Hydroxy-glutamate
Glyoxylate
Pyruvate
D-4-Hydroxy-2-oxoglutarate
L-1-Pyrroline-3-hydroxy-5-carboxylate
L-4-Hydroxy-glutamate semialdehyde
Guanidino-acetate
Guanidino-acetate-P
Creatine
Sarcosine
Urea
N-Carbamoyl-sarcosine
Creatine-P
spontaneous
Creatinine
N-Methyl-hydantoin
Creatine pathway
00330 1/23/07

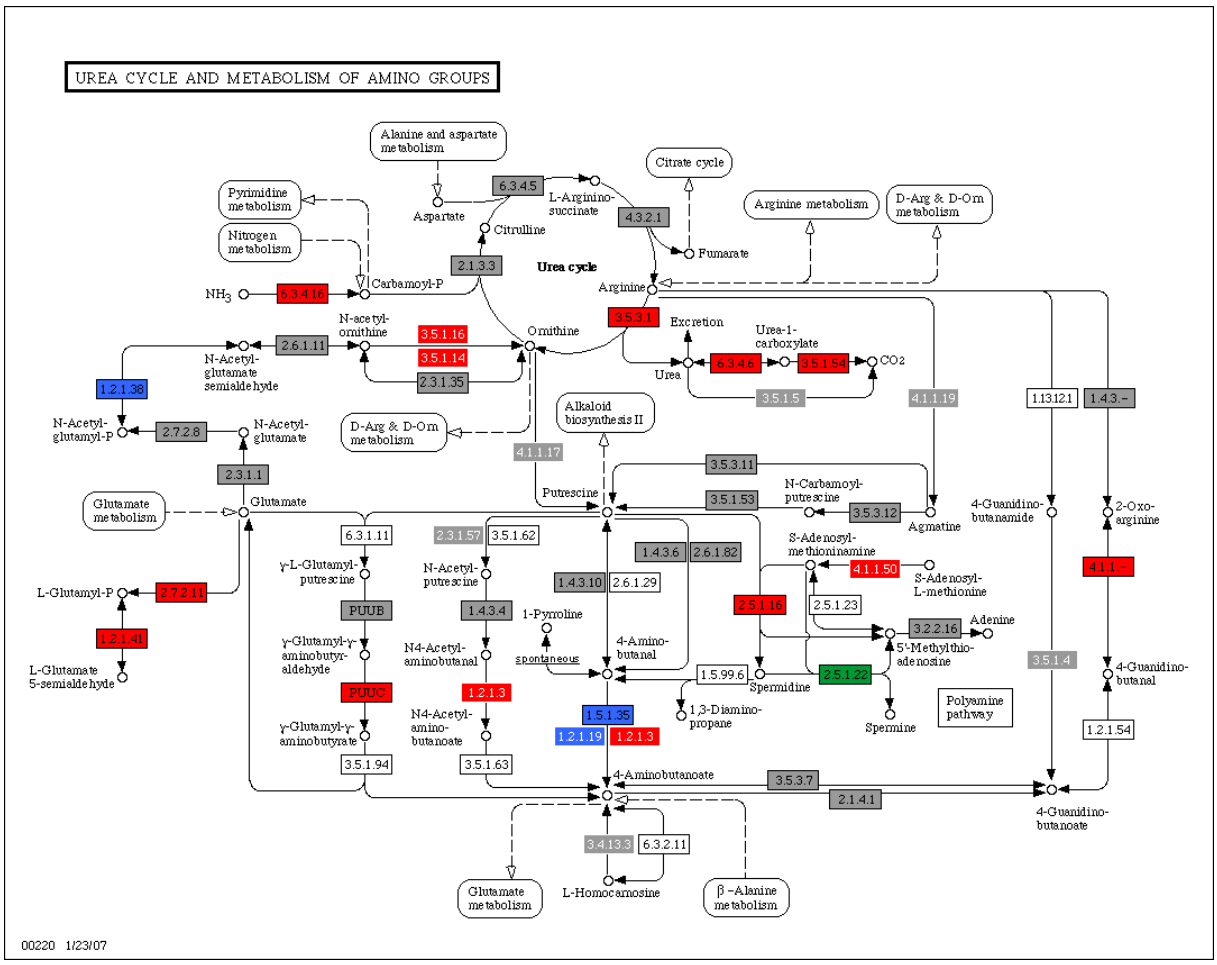

HISTIDINE METABOLISM
Pentose phosphate pathway
PRPP
Phosphoribosyl-ATP
Phosphoribosyl-AMP
Phosphoribosyl-formimino-AICAR-P
Imidazole-glycerol-3P
Imidazole-acetol-P
L-Histidinol-P
L-Histidinol
5'P-Ribosyl-5-amino-4-imidazole carboxamide (AICAR)
Purine metabolism
1-Methyl-L-histidine
L-Histidinal
Anserine
Carnosine
L-Histidyl-tRNA
N-Formyl-L-aspartate
Imidazolone acetate
Imidazole-4-acetate
Imidazole acetaldehyde
Histamine
Hercynine
Thiourocanic acid
L-Histidine
Ergothioneine
N-Formimino-L-aspartate
Aspartate
Alanine and Aspartate metabolism
1-(5-Phosphoribosyl)-imidazole-4-acetate
(1-Ribosylimidazole)-4-acetate
N-Methyl-histamine
4-(β-Acetylaminoethyl)-imidazole
Methylimidazole acetaldehyde
Methylimidazole-acetic acid
Urocanate
4-Oxoglutaramate
2-Oxoglutarate
Formyl-isoglutamate
Isoglutamate
4-Imidazolone-5-propanoate
N-Formimino-L-glutamate
N-Formyl-L-glutamate
Hydantoin-5-propionate
Imidazole-pyruvate
Imidazole-lactate
N-carbamyl-L-glutamate
L-Glutamate
Glutamate metabolism
00340 3/10/04

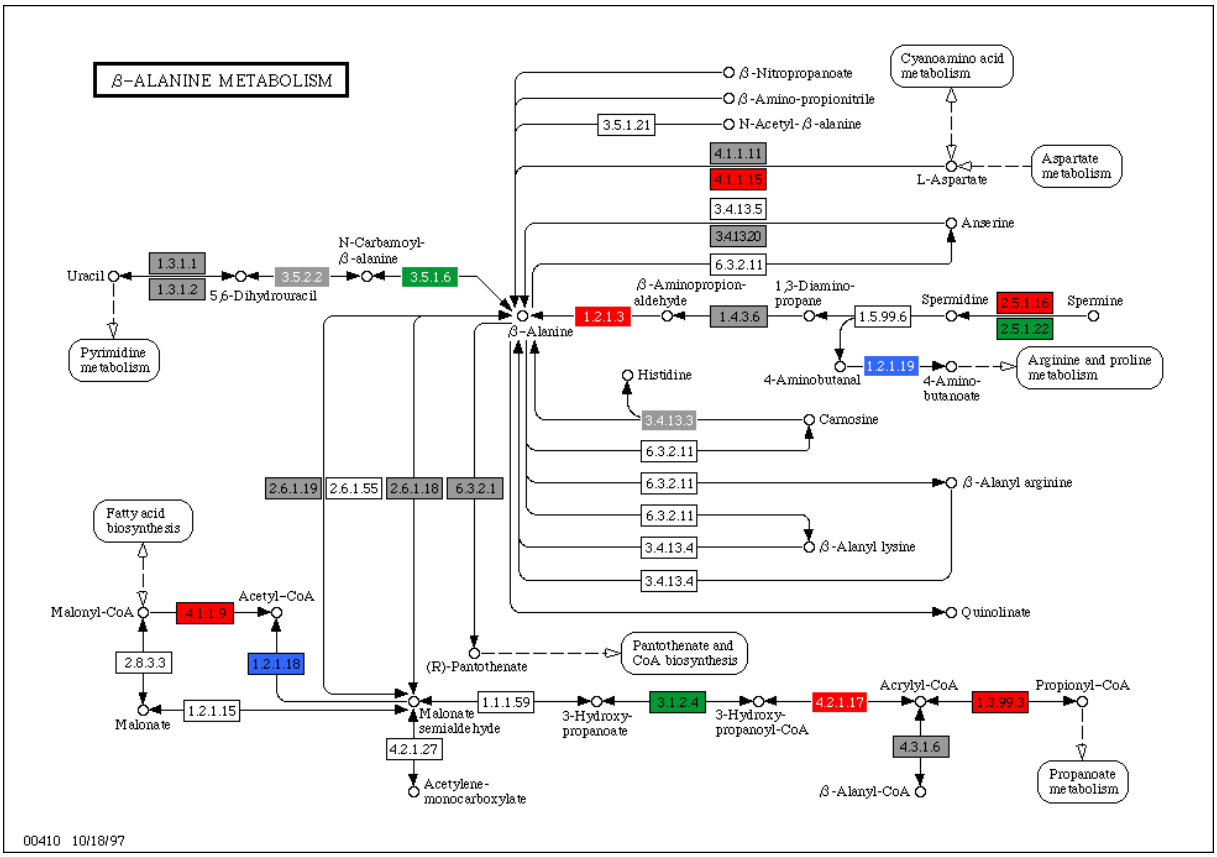

TYROSINE METABOLISM
DDT degradation
Alkaloid biosynthesis I
Tyrosine biosynthesis
Alkaloid biosynthesis II
Citrate cycle
Pyruvate metabolism
00350 12/28/05

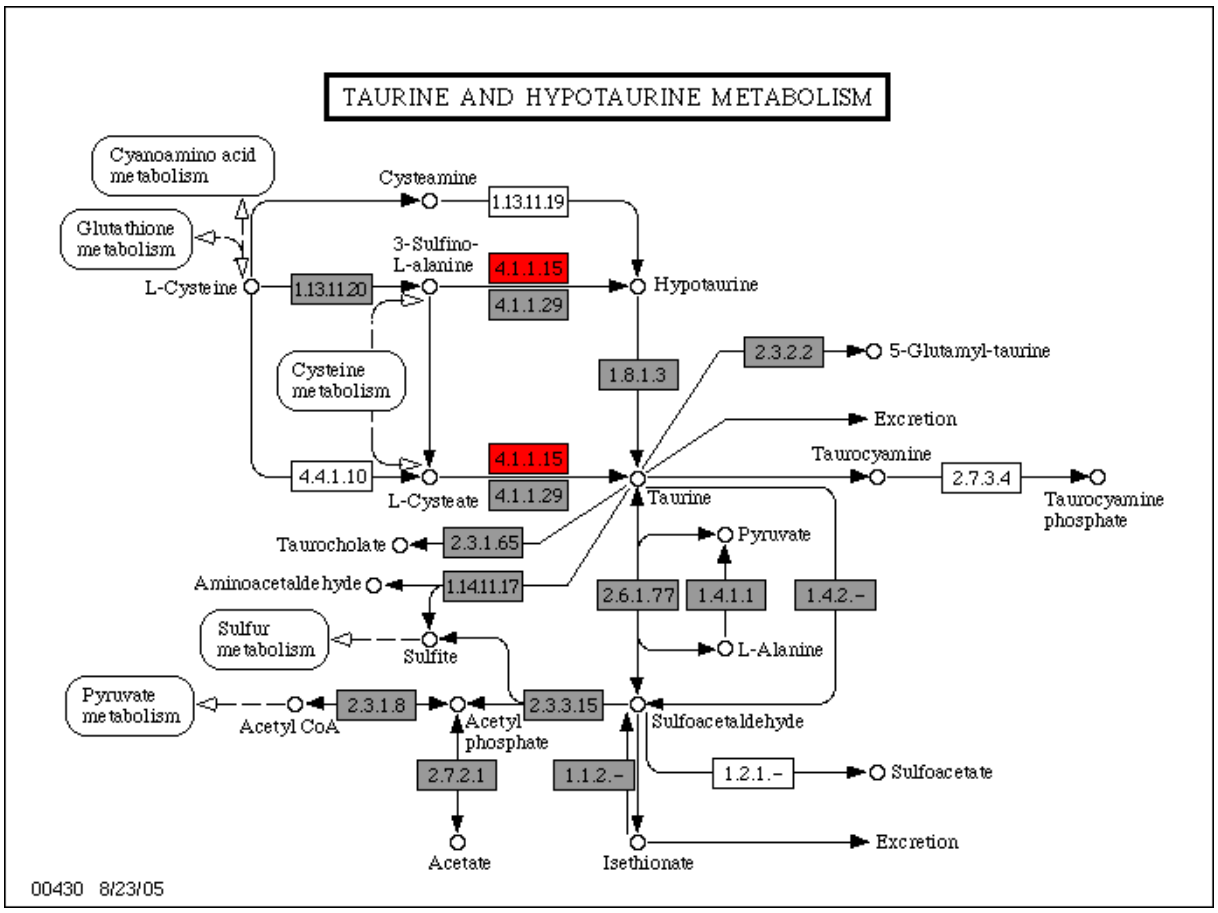

PHENYLALANINE METABOLISM
Benzoate degradation via CoA ligation
Benzoyl-CoA
Hippurate
Alkaloid biosynthesis II
Benzoate degradation via hydroxylation
Phenylalanine,tyrosine and tryptophan biosynthesis
Benzoate
Stilbene, coumarine and lignin biosynthesis
D-Phenylalanine
2-Phenylacetamide
N-Acetyl-D-phenylalanine
2-Hydroxy-3-phenylpropenoate
Phenyl-ethylalcohol
4-Hydroxyphnylacetate
Tyrosine metabolism
3-Hydroxyphenylacetate
4-Hydroxy benzoate
trans-4-Hydroxy-cinnamate
trans-Cinnamate
Phenylalanine
Phenylpyruvate
Phenyl-acetaldehyde
Phenylacetate
trans-2-Hydroxy cinnamate
trans-3-Hydroxy-cinnamate
N-Acetyl-L-phenylalanine
2-Hydroxy-phenylacetate
Phenyl-acetyl-CoA
2-Hydroxyphenyl-propanoate
Phenylpropanoate
3-Hydroxyphenyl propanoate
Phenyllactate
2,6-Dihydroxy-phenylacetate
PP-dihydrodiol
CI-dihydrodiol
Phenylethylamine
Phenyl-acetylglycine
Phenyl-acetylglutamine
MhpA
HcaB
trans-2,3-Dihydroxy cinnamate
2-Hydroxy-6-oxonona-trienedioate
Fumarate
Citrate cycle
MhpB
MhpC
MhpD
MhpE
2,3-Dihydroxy-phenylpropanoate
2-Hydroxy-6-oxonona-2,4-diene-1,9-dioate
cis-2-Hydroxypenta-2,4-dienoate
4-Hydroxy-2-oxovalerate
Pyruvate
Glycolysis
Succinate
Acetaldehyde
Pyruvate metabolism
00360 1/12/05

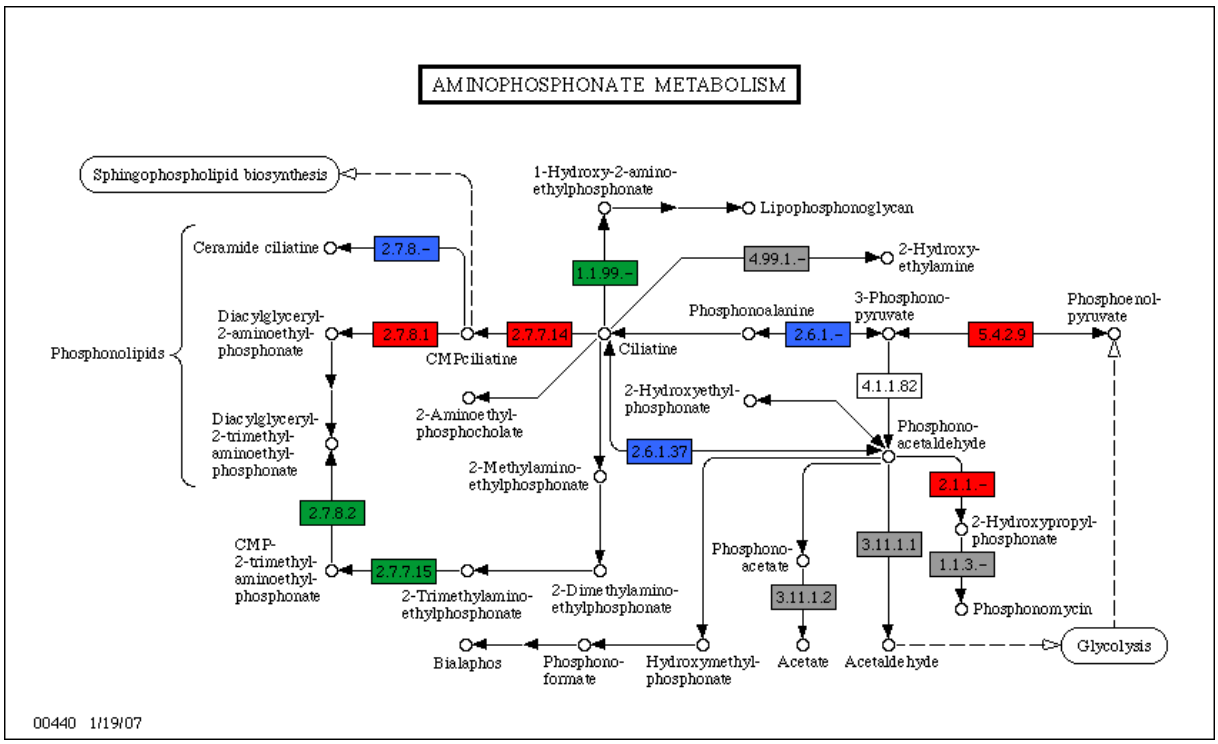

TRYPTOPHAN METABOLISM
5-Hydroxyindole-acetaldehyde
5-Hydroxy-indoleacetate
5-Methoxy-indoleacetate
5-Hydroxyindole-acetylglycine
4-(2-Amino-5-hydroxy-phenyl)-2,4-dioxobutanoate
5-Hydroxy-kynurenine
5-Hydroxy-N-formylkynurenine
6-Hydroxy-kynurenate
4,6-Dihydroxy-quinoline
5-Hydroxy-kynurenamine
5-Methoxytryptamine
Formyl-5-hydroxy-kynurenamine
N-Methylserotonin
5-Hydroxy-indolepyruvate
5-Hydroxy-L-tryptophan
Serotonin
N-Acetyl-serotonin
Melatonin
6-Hydroxy-melatonin
Formyl-N-acetyl-5-methoxykynurenamine
N-Acetylisatin
Acetylindoxyl
Indoxyl
Indole
L-Tryptophanyl-tRNA
Indolelactate
6-Hydroxyindolelactate
2-Formyl-aminobenzaldehyde
3-Indole-glycolaldehyde
Tryptophan
3-Methylindolepyruvate
Indole-3-ethanol
3-Methyl-dioxyindole
2,3-Dihydroxyindole
Indole-3-acetaldehyde
Indolepyruvate
Indoleacetate
Anthranilate
2-Formamino-benzoylacetate
Tryptophan biosynthesis
Tryptamine
N-Methyltryptamine
Benzoate degradation via hydroxylation
Indole-3-acetonitrile
Indole-3-acetamide
Indole-3-acetaldoxime
Glucobrassicin
Benzoate degradation via hydroxylation
Anthranilate
Formyl-anthranilate
N-Formyl-kynurenine
Nicotinamide metabolism
Cinnavalininate
2-Aminophenol
Isophenoxazine
2-Oxoglutarate
Oxaloacetate
L-Kynurenine
3-Hydroxy-L-kynurenine
3-Hydroxy-anthranilate
2-Amino-3-carboxy-muconate semialdehyde
Quinolinate
5-(2'-Carboxyethyl)-4,6-dihydroxypicolinate
3-Hydroxy-kynurenamine
5-(2'-Formylethyl)-4,6-dihydroxypicolinate
4-(2-Aminophenyl)-2,4-dioxobutanoate
4,8-Dihydroxy-quinoline
3-Methoxy-anthranilate
2-Aminomuconate semialdehyde
7,8-Dihydroxy-kynurenate
8-Methoxy-kynurenate
4-(2-Amino-3-hydroxy-phenyl)-2,4-dioxobutanoate
Xanthurenate
2-Amino-muconate
γ'-Oxalo-crotonate
5-(3'-Carboxy-3'-oxopropyl)-4,6-dihydroxypicolinate
5-(3'-Carboxy-3'-oxopropenyl)-4,6-dihydroxypicolinate
7,8-Dihydro-7,8-dihydroxykynurenate
Kynurenate
Acetoacetyl-CoA
Crotonyl-CoA
Glycolysis
Acetyl-CoA
(S)-3-Hydroxy-butanoyl-CoA
Glutaryl-CoA
2-Oxoadipate
00380 8/23/06

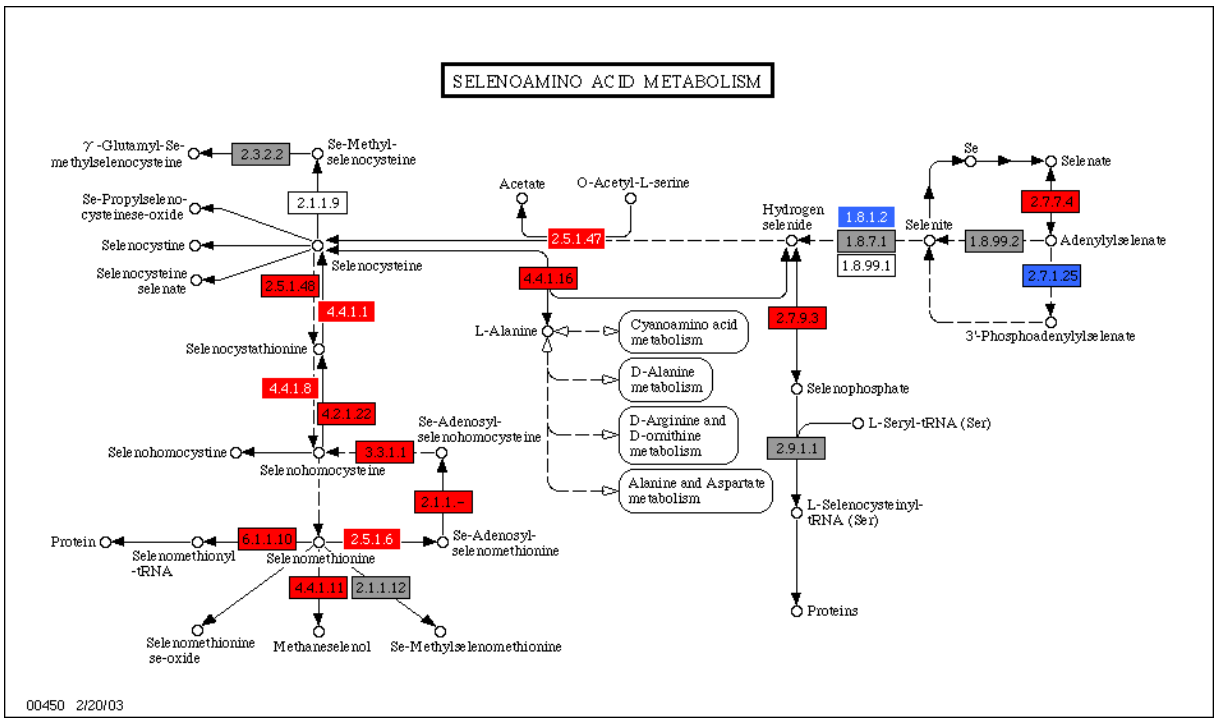

PHENYLALANINE, TYROSINE AND TRYPTOPHAN BIOSYNTHESIS
Tyrosine metabolism
Alkaloid biosynthesis I
Puromycin biosynthesis
Tyr-tRNA
Tyrosine
4-Hydroxy-phenylpyruvate
Phenylalanine metabolism
Alkaloid biosynthesis II
Photosynthesis
Pentose phosphate pathway
Pretyrosine
Phenylpyruvate
Prephenate
Phe-tRNA
Phenylalanine
Phosphoenol-pyruvate
D-Erythrose 4-phosphate
Glycolysis
Indole
Histidine
Shikimate 3-phosphate
5-O-(1-Carboxyvinyl)-3-phosphoshikimate
Chorismate
Anthranilate
N-(5-Phospho-β-D-ribosyl)-anthranilate
1-(2-Carboxy-phenylamino)-1-deoxy-D-ribulose 5-phosphate
(3-Indolyl)-glycerol phosphate
L-Tryptophan
7P-2-Dehydro-3-deoxy-D-arabino-heptonate
Shikimate
Ubiquinone biosynthesis
Folate biosynthesis
Tryptophan metabolism
3-Dehydro-quinate
3-Dehydro-shikimate
Protocatechuate
Quinate
00400 1/12/05

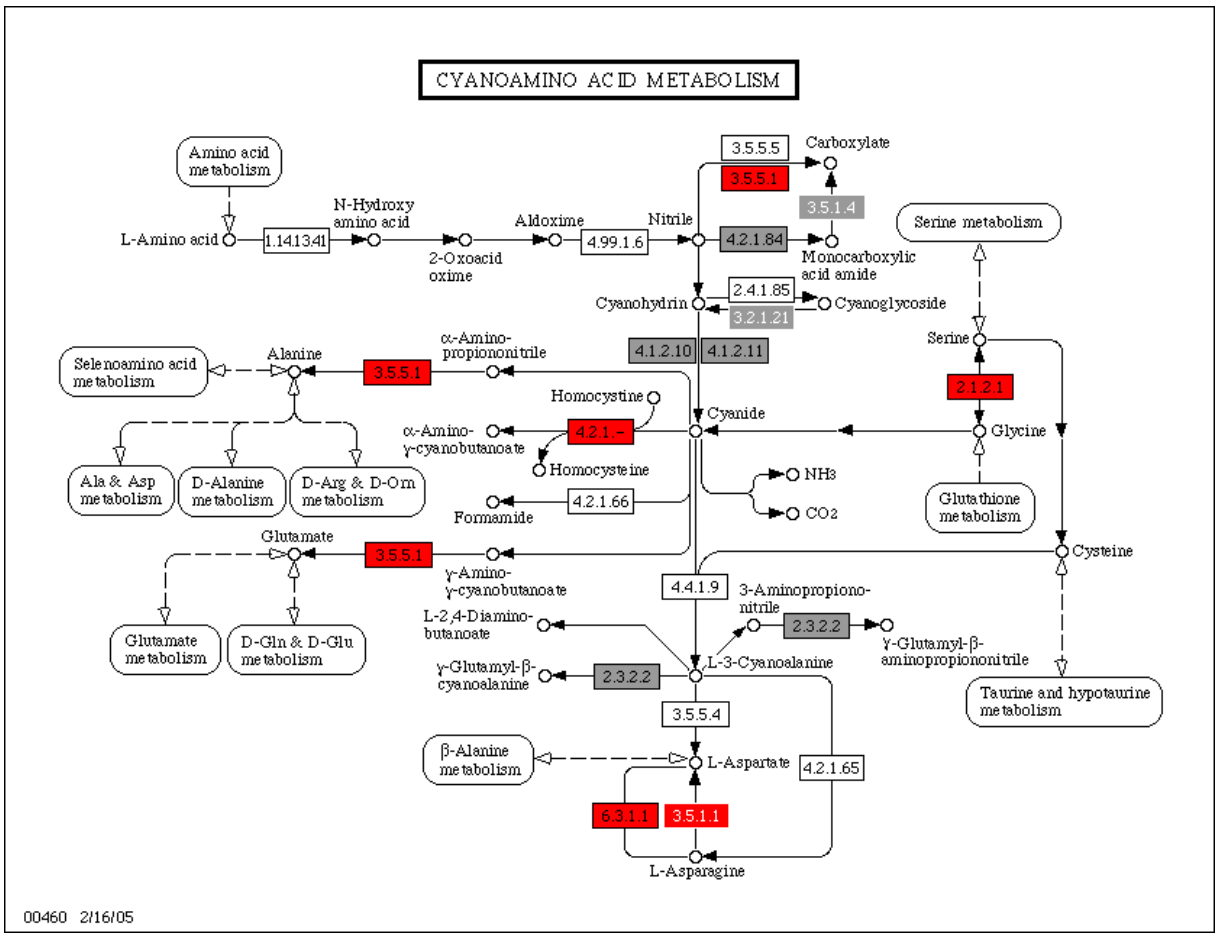

UREA CYCLE AND METABOLISM OF AMINO GROUPS
Alanine and aspartate metabolism
Pyrimidine metabolism
Nitrogen metabolism
Aspartate
Citrate cycle
Arginine metabolism
D-Arg & D-Orn metabolism
L-Arginino-succinate
Fumarate
Citrulline
Urea cycle
Carbamoyl-P
NH3
Arginine
N-acetyl-ornithine
Ornithine
Excretion
Urea-1-carboxylate
Urea
CO2
N-Acetyl-glutamate semialdehyde
N-Acetyl-glutamyl-P
N-Acetyl-glutamate
Alkaloid biosynthesis II
Glutamate metabolism
Glutamate
Putrescine
N-Carbamoyl-putrescine
Agmatine
4-Guanidino-butanamide
2-Oxo-arginine
γ-L-Glutamyl-putrescine
N-Acetyl-putrescine
S-Adenosyl-methioninamine
S-Adenosyl-L-methionine
L-Glutamyl-P
L-Glutamate 5-semialdehyde
PUUB
PUUC
1-Pyrroline
spontaneous
4-Amino-butanal
γ-Glutamyl-γ-aminobutyr-aldehyde
N4-Acetyl-aminobutanal
γ-Glutamyl-γ-aminobutyrate
N4-Acetyl-amino-butanoate
Spermidine
1,3-Diamino-propane
Spermine
5'-Methylthio-adenosine
Adenine
Polyamine pathway
4-Guanidino-butanal
4-Aminobutanoate
4-Guanidino-butanoate
Glutamate metabolism
L-Homocarnosine
β-Alanine metabolism
00220 1/23/07

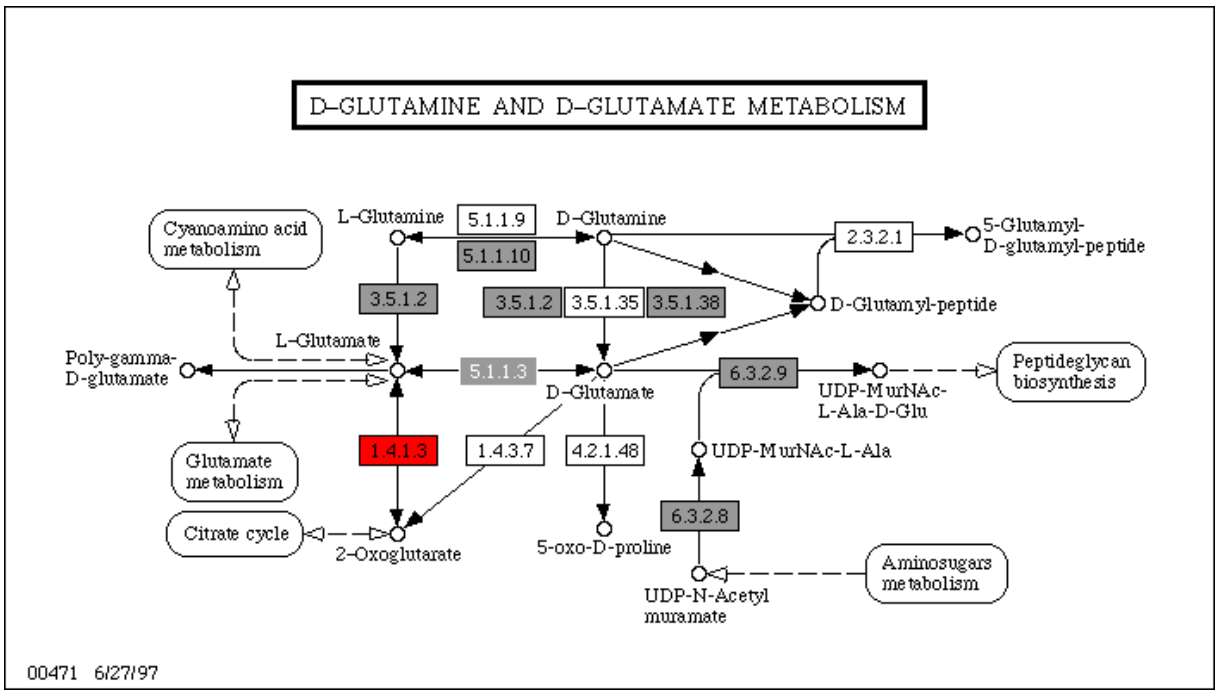

β-ALANINE METABOLISM
β-Nitropropanoate
β-Amino-propionitrile
N-Acetyl-β-alanine
Cyanoamino acid metabolism
L-Aspartate
Aspartate metabolism
Anserine
N-Carbamoyl-β-alanine
Uracil
5,6-Dihydrouracil
Pyrimidine metabolism
β-Aminopropion-aldehyde
1,3-Diamino-propane
Spermidine
Spermine
β-Alanine
4-Aminobutanal
4-Amino-butanoate
Arginine and proline metabolism
Histidine
Carnosine
β-Alanyl arginine
β-Alanyl lysine
Quinolinate
Fatty acid biosynthesis
Malonyl-CoA
Acetyl-CoA
(R)-Pantothenate
Pantothenate and CoA biosynthesis
Malonate
Malonate semialdehyde
3-Hydroxy-propanoate
3-Hydroxy-propanoyl-CoA
Acrylyl-CoA
Propionyl-CoA
Acetylene-monocarboxylate
β-Alanyl-CoA
Propanoate metabolism
00410 10/18/97

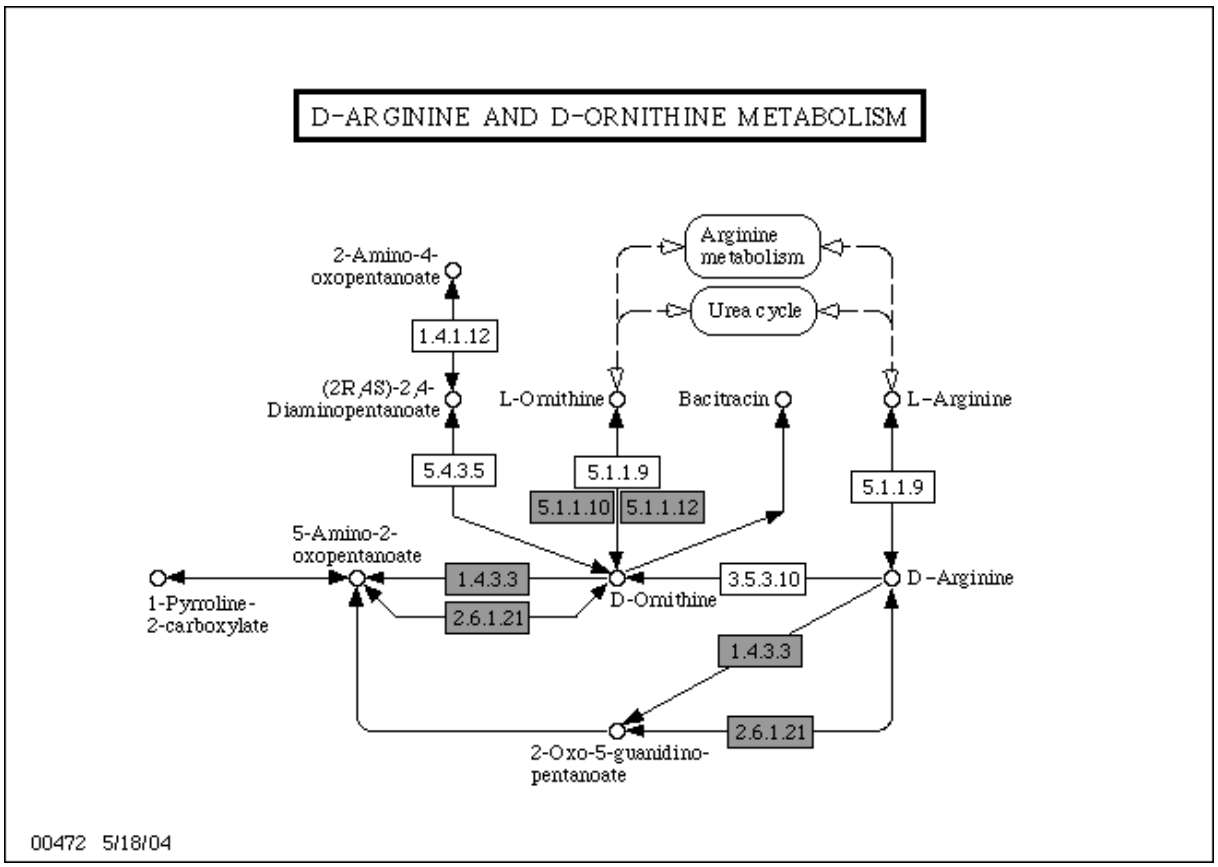

TAURINE AND HYPOTAURINE METABOLISM
Cyanoamino acid metabolism
Glutathione metabolism
Cysteine metabolism
Pyruvate metabolism
Sulfur metabolism
Cysteamine
1.13.11.19
3-Sulfino-L-alanine
4.1.1.15
4.1.1.29
L-Cysteine
1.13.11.20
Hypotaurine
1.8.1.3
2.3.2.2
5-Glutamyl-taurine
Excretion
Taurocyamine
2.7.3.4
Taurocyamine phosphate
4.4.1.10
L-Cysteate
Taurine
Taurocholate
2.3.1.65
Pyruvate
Aminoacetaldehyde
1.14.11.17
2.6.1.77
1.4.1.1
1.4.2.-
L-Alanine
Sulfite
2.3.1.8
Acetyl CoA
Acetyl phosphate
2.3.3.15
Sulfoacetaldehyde
2.7.2.1
1.1.2.-
1.2.1.-
Sulfoacetate
Acetate
Isethionate
Excretion
00430 8/23/05

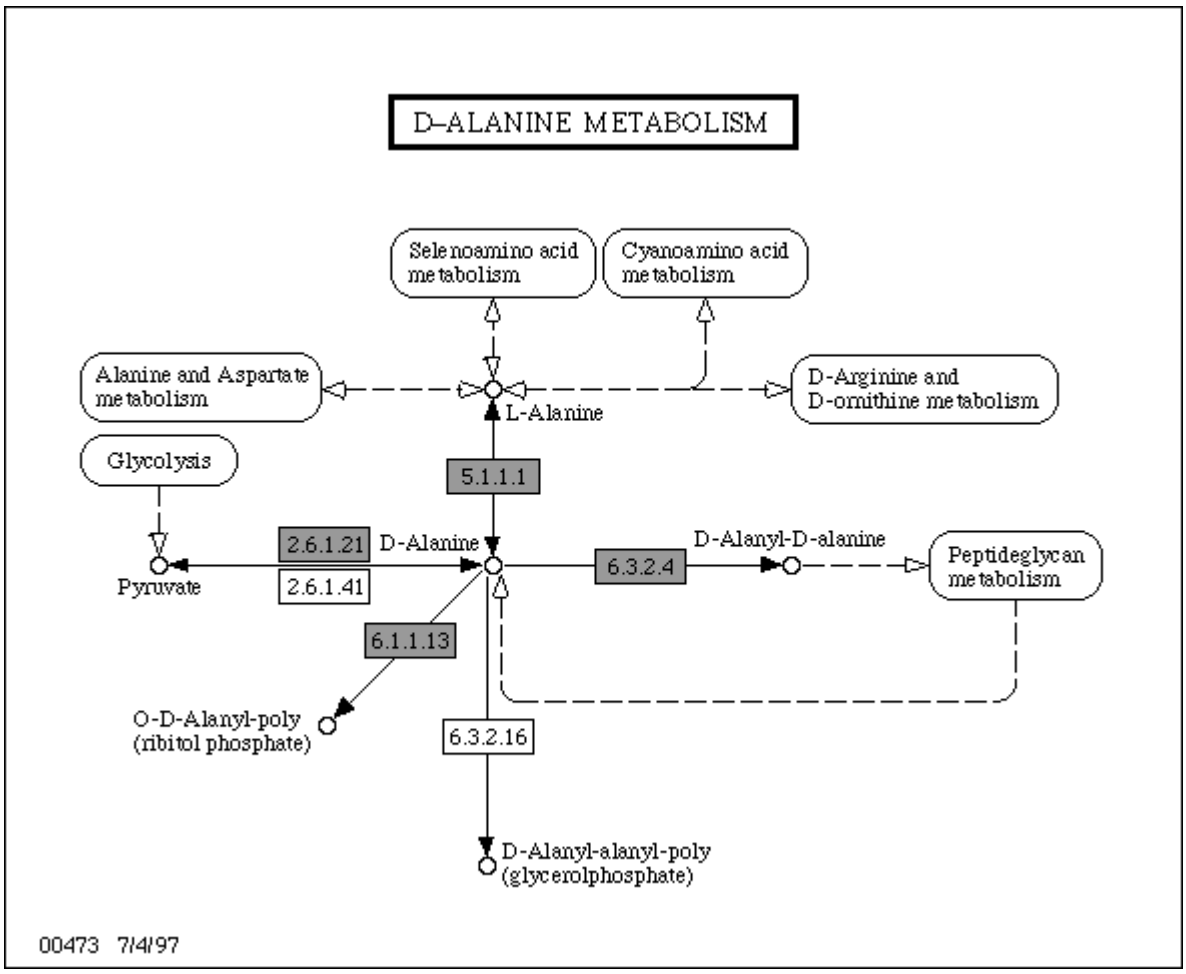

AMINOPHOSPHONATE METABOLISM
Sphingophospholipid biosynthesis
Phosphonolipids
1-Hydroxy-2-amino-ethylphosphonate
Lipophosphonoglycan
Ceramide ciliatine
2.7.8.-
1.1.99.-
4.99.1.-
2-Hydroxy-ethylamine
Diacylglyceryl-2-aminoethyl-phosphonate
2.7.8.1
2.7.7.14
CMPciliatine
Ciliatine
Phosphonoalanine
2.6.1.-
3-Phospho-pyruvate
5.4.2.9
Phosphoenol-pyruvate
2-Aminoethyl-phosphocholate
2-Hydroxyethyl-phosphonate
4.1.1.82
Diacylglyceryl-2-trimethyl-aminoethyl-phosphonate
2-Methylamino-ethylphosphonate
2.6.1.37
Phospho-acetaldehyde
2.1.1.-
2-Hydroxypropyl-phosphonate
2.7.8.2
CMP-2-trimethyl-aminoethyl-phosphonate
2.7.7.15
2-Trimethylamino-ethylphosphonate
2-Dimethylamino-ethylphosphonate
Phospho-acetate
3.11.1.1
1.1.3.-
3.11.1.2
Phosphonomycin
Bialaphos
Phospho-formate
Hydroxymethyl-phosphonate
Acetate
Acetaldehyde
Glycolysis
00440 1/19/07

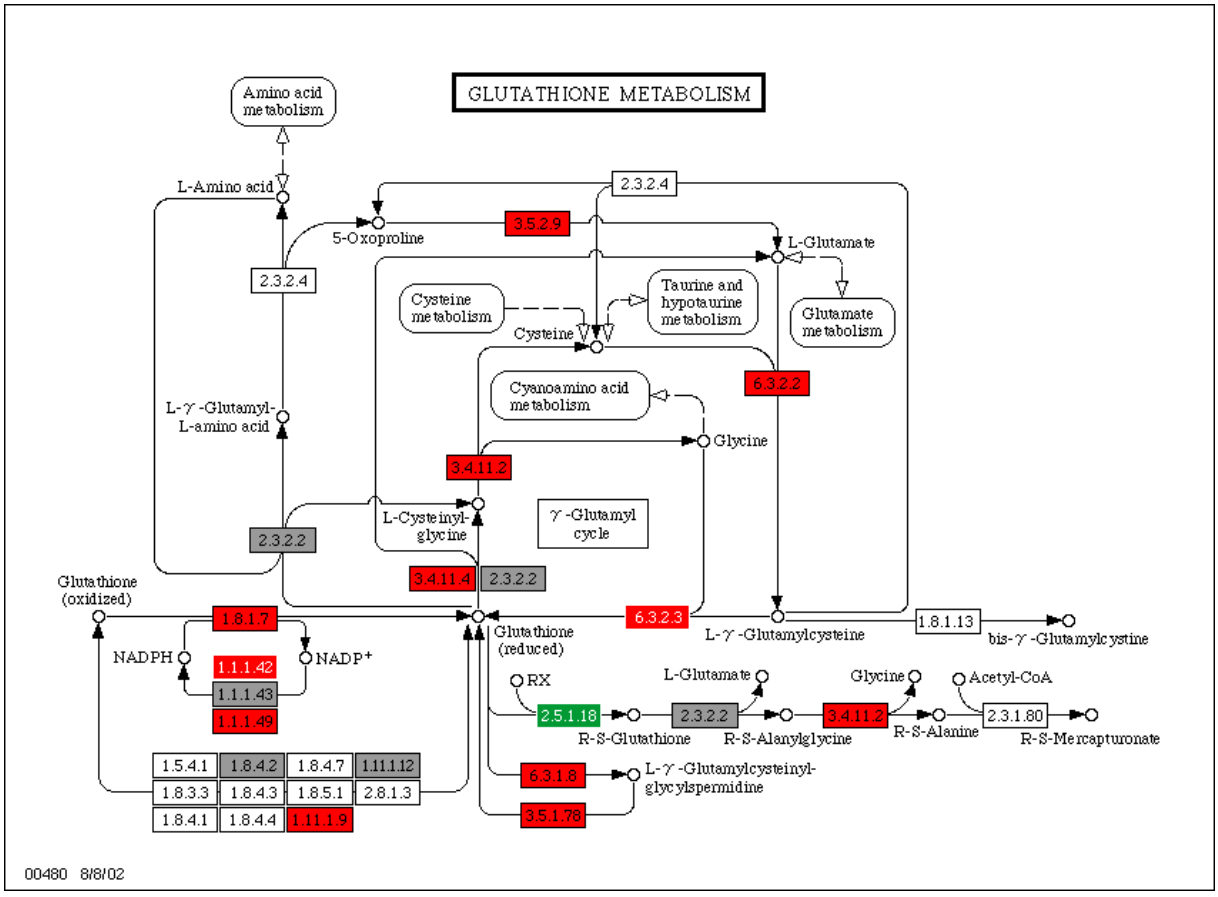

SELENOAMINO ACID METABOLISM
γ-Glutamyl-Se-methylselenocysteine
2.3.2.2
Se-Methyl-selenocysteine
2.1.1.9
Se-Propylseleno-cysteinese-oxide
Selenocystine
Selenocysteine selenate
Selenocysteine
2.5.1.47
Acetate
O-Acetyl-L-serine
Hydrogen selenide
1.8.1.2
1.8.7.1
1.8.99.1
Selenite
Se
Selenate
2.7.7.4
1.8.99.2
Adenylylselenate
2.7.1.25
3'-Phosphoadenylylselenate
2.5.1.48
4.4.1.1
4.4.1.16
Selenocystathionine
L-Alanine
Cyanoamino acid metabolism
D-Alanine metabolism
D-Arginine and D-ornithine metabolism
Alanine and Aspartate metabolism
2.7.9.3
Selenophosphate
L-Seryl-tRNA (Ser)
2.9.1.1
L-Selenocysteinyl-tRNA (Ser)
Proteins
4.4.1.8
4.2.1.22
Se-Adenosyl-selenohomocysteine
Selenohomocystine
3.3.1.1
Selenohomocysteine
2.1.1.-
Protein
6.1.1.10
Selenomethionyl -tRNA
2.5.1.6
Selenomethionine
Se-Adenosyl-selenomethionine
4.4.1.11
2.1.1.12
Selenomethionine se-oxide
Methaneselenol
Se-Methylselenomethionine
00450 2/20/03


CYANOAMINO ACID METABOLISM
Amino acid metabolism
L-Amino acid
1.14.13.41
N-Hydroxy amino acid
2-Oxoacid oxime
Aldoxime
4.99.1.6
Nitrile
4.2.1.84
3.5.5.5
3.5.5.1
Carboxylate
3.5.1.4
Monocarboxylic acid amide
Serine metabolism
Cyanohydrin
2.4.1.85
3.2.1.21
Cyanoglycoside
Serine
2.1.2.1
Glycine
Glutathione metabolism
Cysteine
4.1.2.10
4.1.2.11
α-Amino-propiononitrile
Alanine
Selenoamino acid metabolism
3.5.5.1
Homocystine
Cyanide
α-Amino-γ-cyanobutanoate
4.2.1.-
Homocysteine
NH3
CO2
Ala & Asp metabolism
D-Alanine metabolism
D-Arg & D-Orn metabolism
Formamide
4.2.1.66
Glutamate
3.5.5.1
γ-Amino-γ-cyanobutanoate
4.4.1.9
3-Aminopropiono-nitrile
2.3.2.2
γ-Glutamyl-β-aminopropiononitrile
L-2,4-Diamino-butanoate
γ-Glutamyl-β-cyanoalanine
2.3.2.2
L-3-Cyanoalanine
Taurine and hypotaurine metabolism
Glutamate metabolism
D-Gln & D-Glu metabolism
3.5.5.4
β-Alanine metabolism
L-Aspartate
4.2.1.65
6.3.1.1
3.5.1.1
L-Asparagine
00460 2/16/05


D-GLUTAMINE AND D-GLUTAMATE METABOLISM
Cyanoamino acid metabolism
L-Glutamine
5.1.1.9
D-Glutamine
5.1.1.10
2.3.2.1
5-Glutamyl-D-glutamyl-peptide
3.5.1.2
3.5.1.2
3.5.1.35
3.5.1.38
D-Glutamyl-peptide
L-Glutamate
Poly-gamma-D-glutamate
5.1.1.3
D-Glutamate
6.3.2.9
UDP-MurNAc-L-Ala-D-Glu
Peptideglycan biosynthesis
1.4.1.3
1.4.3.7
4.2.1.48
UDP-MurNAc-L-Ala
Glutamate metabolism
6.3.2.8
Citrate cycle
2-Oxoglutarate
5-oxo-D-proline
Aminosugars metabolism
UDP-N-Acetyl muramate
00471 6/27/97


D-ARGININE AND D-ORNITHINE METABOLISM
Arginine metabolism
Urea cycle
2-Amino-4-oxopentanoate
1.4.1.12
(2R,4S)-2,4-Diaminopentanoate
L-Ornithine
Bacitracin
L-Arginine
5.4.3.5
5.1.1.9
5.1.1.10
5.1.1.12
5.1.1.9
5-Amino-2-oxopentanoate
1.4.3.3
3.5.3.10
D-Arginine
1-Pyrroline-2-carboxylate
2.6.1.21
D-Ornithine
1.4.3.3
2.6.1.21
2-Oxo-5-guanidino-pentanoate
00472 5/18/04


D-ALANINE METABOLISM
Selenoamino acid metabolism
Cyanoamino acid metabolism
Alanine and Aspartate metabolism
L-Alanine
D-Arginine and D-ornithine metabolism
Glycolysis
5.1.1.1
2.6.1.21
D-Alanine
D-Alanyl-D-alanine
Peptideglycan metabolism
Pyruvate
2.6.1.41
6.3.2.4
6.1.1.13
O-D-Alanyl-poly (ribitol phosphate)
6.3.2.16
D-Alanyl-alanyl-poly (glycerolphosphate)
00473 7/4/97


GLUTATHIONE METABOLISM
Amino acid metabolism
L-Amino acid
2.3.2.4
3.5.2.9
5-Oxoproline
L-Glutamate
Cysteine metabolism
Taurine and hypotaurine metabolism
Glutamate metabolism
Cysteine
Cyanoamino acid metabolism
6.3.2.2
L-γ-Glutamyl-L-amino acid
Glycine
3.4.11.2
L-Cysteinyl-glycine
γ-Glutamyl cycle
2.3.2.2
3.4.11.4
Glutathione (oxidized)
1.8.1.7
6.3.2.3
L-γ-Glutamylcysteine
1.8.1.13
bis-γ-Glutamylcystine
NADPH
NADP+
1.1.1.42
1.1.1.43
1.1.1.49
Glutathione (reduced)
RX
L-Glutamate
Glycine
Acetyl-CoA
2.5.1.18
2.3.2.2
3.4.11.2
2.3.1.80
R-S-Glutathione
R-S-Alanylglycine
R-S-Alanine
R-S-Mercapturonate
1.5.4.1
1.8.4.2
1.8.4.7
1.11.1.12
1.8.3.3
1.8.4.3
1.8.5.1
2.8.1.3
1.8.4.1
1.8.4.4
1.11.1.9
6.3.1.8
L-γ-Glutamylcysteinyl-glycylspermidine
3.5.1.78
00480 8/8/02

C) Atividades enzimáticas identificadas a partir das proteínas preditas de *T.cruzi*, nos respectivos grupos (clusters), provenientes das análises feitas com o AnEnPi, tendo como ponto de corte um *e-value* de $e^{-20}$, com os identificadores das sequências (TCR HIT).

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 1.1.1.- | 1 | Tc00.1047053509715.90, Tc00.1047053510105.100 |
| 1.1.1.- | 11 | Tc00.1047053504057.40 |
| 1.1.1.- | 13 | Tc00.1047053506263.10, Tc00.1047053506263.20, Tc00.1047053506263.30, Tc00.1047053510099.120 |
| 1.1.1.- | 17 | Tc00.1047053506295.160, Tc00.1047053510091.80, Tc00.1047053510421.320, Tc00.1047053511469.100 |
| 1.1.1.- | 2 | Tc00.1047053506219.40, Tc00.1047053506567.70, Tc00.1047053506959.64, Tc00.1047053507049.60, Tc00.1047053508297.50, Tc00.1047053508805.80, Tc00.1047053509213.100, Tc00.1047053509767.170, Tc00.1047053510257.60, Tc00.1047053510997.10 |
| 1.1.1.- | 22 | Tc00.1047053507547.40, Tc00.1047053508441.70, Tc00.1047053508981.39, Tc00.1047053509701.10, Tc00.1047053509717.90, Tc00.1047053510105.240 |
| 1.1.1.- | 3 | Tc00.1047053505807.180, Tc00.1047053507017.40 |
| 1.1.1.- | 30 | Tc00.1047053507831.70 |
| 1.1.1.- | 32 | Tc00.1047053505183.20, Tc00.1047053505183.30, Tc00.1047053508647.270, Tc00.1047053508647.280 |
| 1.1.1.- | 4 | Tc00.1047053505183.120, Tc00.1047053511287.49, Tc00.1047053511627.120 |
| 1.1.1.- | 45 | Tc00.1047053507617.9, Tc00.1047053508461.80, Tc00.1047053509941.100 |
| 1.1.1.- | 5 | Tc00.1047053504151.60, Tc00.1047053509459.79 |
| 1.1.1.- | 6 | Tc00.1047053504339.19, Tc00.1047053504425.60, Tc00.1047053506485.80, Tc00.1047053508677.80, Tc00.1047053509331.210 |
| 1.1.1.- | 7 | Tc00.1047053506357.50, Tc00.1047053511277.60 |
| 1.1.1.1 | 1 | Tc00.1047053504339.19, Tc00.1047053504425.60, Tc00.1047053506485.80, Tc00.1047053508677.80, Tc00.1047053509331.210, Tc00.1047053510265.10, Tc00.1047053510311.90 |
| 1.1.1.1 | 2 | Tc00.1047053506357.50, Tc00.1047053511277.60 |
| 1.1.1.1 | 3 | Tc00.1047053505183.120, Tc00.1047053511287.49, Tc00.1047053511627.120 |
| 1.1.1.103 | 1 | Tc00.1047053504339.19, Tc00.1047053504425.60, Tc00.1047053508677.80, Tc00.1047053509331.210 |
| 1.1.1.103 | 2 | Tc00.1047053507923.10, Tc00.1047053511753.120 |
| 1.1.1.29 | 1 | Tc00.1047053506263.10, Tc00.1047053506263.30, Tc00.1047053510099.120 |
| 1.1.1.31 | 1 | Tc00.1047053505807.180, Tc00.1047053506727.100, Tc00.1047053506727.90, Tc00.1047053507017.40, Tc00.1047053511529.160, Tc00.1047053511529.170, Tc00.1047053511865.50, Tc00.1047053511871.40 |
| 1.1.1.35 | 1 | Tc00.1047053506567.70, Tc00.1047053507049.60, Tc00.1047053507801.90, Tc00.1047053511627.150 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 1.1.1.35 | 3 | Tc00.1047053506629.220, Tc00.1047053507001.80, Tc00.1047053508465.110, Tc00.1047053508827.40, Tc00.1047053508951.40, Tc00.1047053509153.120, Tc00.1047053509829.20, Tc00.1047053510121.40, Tc00.1047053510303.290 |
| 1.1.1.42 | 1 | Tc00.1047053506925.319, Tc00.1047053511575.60 |
| 1.1.1.49 | 1 | Tc00.1047053436535.19, Tc00.1047053506953.49, Tc00.1047053509287.50 |
| 1.1.1.85 | 3 | Tc00.1047053510817.60 |
| 1.1.1.95 | 1 | Tc00.1047053506263.10, Tc00.1047053506263.20, Tc00.1047053506263.30, Tc00.1047053510099.120 |
| 1.1.1.95 | 2 | Tc00.1047053508207.200, Tc00.1047053509509.60 |
| 1.1.2.- | 2 | Tc00.1047053504151.60 |
| 1.1.3.- | 2 | Tc00.1047053507047.150, Tc00.1047053509179.100 |
| 1.1.99.- | 1 | Tc00.1047053506851.20, Tc00.1047053511141.20, Tc00.1047053511229.20, Tc00.1047053511589.170 |
| 1.1.99.- | 8 | Tc00.1047053504151.60, Tc00.1047053509459.79 |
| 1.1.99.1 | 1 | Tc00.1047053506851.20, Tc00.1047053511141.20, Tc00.1047053511229.20, Tc00.1047053511589.170 |
| 1.11.1.12 | 1 | Tc00.1047053503899.110, Tc00.1047053503899.119, Tc00.1047053503899.130, Tc00.1047053507515.10, Tc00.1047053507515.4 |
| 1.11.1.6 | 2 | Tc00.1047053503745.30, Tc00.1047053506193.60 |
| 1.11.1.7 | 3 | Tc00.1047053487507.10, Tc00.1047053504839.28, Tc00.1047053504839.44, Tc00.1047053505983.9, Tc00.1047053507259.10, Tc00.1047053508649.5, Tc00.1047053509445.10, Tc00.1047053509499.14 |
| 1.11.1.7 | 4 | Tc00.1047053503899.110, Tc00.1047053503899.119, Tc00.1047053503899.130, Tc00.1047053507515.10 |
| 1.11.1.7 | 6 | Tc00.1047053503745.30, Tc00.1047053506193.60 |
| 1.11.1.9 | 1 | Tc00.1047053503899.110, Tc00.1047053503899.119, Tc00.1047053503899.130, Tc00.1047053507515.10, Tc00.1047053507515.4, Tc00.1047053511019.99, Tc00.1047053511543.60 |
| 1.13.11.- | 8 | Tc00.1047053508409.160, Tc00.1047053510141.10 |
| 1.14.13.- | 1 | Tc00.1047053508173.100 |
| 1.14.13.- | 2 | Tc00.1047053508153.170, Tc00.1047053508183.10 |
| 1.14.13.- | 7 | Tc00.1047053505941.30, Tc00.1047053506567.60, Tc00.1047053507049.50, Tc00.1047053507623.90, Tc00.1047053509117.19, Tc00.1047053509777.90, Tc00.1047053509789.10 |
| 1.14.13.39 | 1 | Tc00.1047053506585.70, Tc00.1047053506931.80, Tc00.1047053510657.170, Tc00.1047053510747.60, Tc00.1047053510877.120 |
| 1.14.14.1 | 1 | Tc00.1047053506297.260, Tc00.1047053506585.70, Tc00.1047053506931.80, Tc00.1047053506945.190, Tc00.1047053509231.10, Tc00.1047053509719.40, Tc00.1047053510101.50, Tc00.1047053510657.170, Tc00.1047053510747.60, Tc00.1047053510877.120 |
| 1.2.1.- | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053509351.10, Tc00.1047053510943.50, Tc00.1047053511523.100, Tc00.1047053511733.10, Tc00.1047053511751.120 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 1.2.1.- | 3 | Tc00.1047053503687.20, Tc00.1047053503723.90, Tc00.1047053506409.240, Tc00.1047053506885.413, Tc00.1047053506943.50, Tc00.1047053506943.60, Tc00.1047053508537.10, Tc00.1047053509065.60, Tc00.1047053509065.70, Tc00.1047053509717.80, Tc00.1047053510105.230, Tc00.1047053510187.60, Tc00.1047053511235.20, Tc00.1047053511461.14 |
| 1.2.1.16 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053510943.50, Tc00.1047053511523.100, Tc00.1047053511733.10, Tc00.1047053511751.120 |
| 1.2.1.18 | 1 | Tc00.1047053506155.60, Tc00.1047053507641.60 |
| 1.2.1.19 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053510943.50, Tc00.1047053511751.120 |
| 1.2.1.19 | 2 | Tc00.1047053509799.140, Tc00.1047053511037.20 |
| 1.2.1.24 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053510943.50, Tc00.1047053511523.100, Tc00.1047053511733.10, Tc00.1047053511751.120 |
| 1.2.1.27 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053510943.50, Tc00.1047053511751.120 |
| 1.2.1.3 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053509351.10, Tc00.1047053510943.50, Tc00.1047053511523.100, Tc00.1047053511733.10, Tc00.1047053511751.120 |
| 1.2.1.32 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053510943.50, Tc00.1047053511751.120 |
| 1.2.1.38 | 2 | Tc00.1047053506155.60, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053510943.50, Tc00.1047053511751.120 |
| 1.2.1.39 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053510943.50, Tc00.1047053511523.100, Tc00.1047053511733.10, Tc00.1047053511751.120 |
| 1.2.1.41 | 1 | Tc00.1047053509067.70, Tc00.1047053511023.10 |
| 1.2.1.47 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053510943.50, Tc00.1047053511751.120 |
| 1.2.1.5 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053510943.50, Tc00.1047053511751.120 |
| 1.2.1.51 | 1 | Tc00.1047053506585.70, Tc00.1047053510657.170, Tc00.1047053510747.60, Tc00.1047053510877.120 |
| 1.2.1.60 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509351.10, Tc00.1047053510943.50, Tc00.1047053511751.120 |
| 1.2.1.71 | 1 | Tc00.1047053506155.60, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053510943.50, Tc00.1047053511751.120 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 1.2.1.8 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509065.9, Tc00.1047053509351.10, Tc00.1047053510943.50, Tc00.1047053511523.100, Tc00.1047053511733.10, Tc00.1047053511751.120 |
| 1.2.4.1 | 1 | Tc00.1047053506295.160, Tc00.1047053506853.50, Tc00.1047053507831.70, Tc00.1047053508781.110, Tc00.1047053510091.80, Tc00.1047053510421.320, Tc00.1047053511261.160, Tc00.1047053511469.100 |
| 1.2.4.1 | 3 | Tc00.1047053508637.114, Tc00.1047053511911.159 |
| 1.2.4.2 | 1 | Tc00.1047053408345.20, Tc00.1047053503793.10, Tc00.1047053506337.70, Tc00.1047053510717.30, Tc00.1047053510797.10 |
| 1.2.4.2 | 4 | Tc00.1047053506853.50, Tc00.1047053507831.70, Tc00.1047053508781.110 |
| 1.2.4.4 | 1 | Tc00.1047053506295.160, Tc00.1047053506853.50, Tc00.1047053507831.70, Tc00.1047053508781.110, Tc00.1047053510091.80, Tc00.1047053510421.320, Tc00.1047053511261.160, Tc00.1047053511469.100 |
| 1.3.1.- | 2 | Tc00.1047053506567.70, Tc00.1047053507049.60 |
| 1.3.1.- | 6 | Tc00.1047053507641.280, Tc00.1047053507641.290, Tc00.1047053507641.300, Tc00.1047053509063.30, Tc00.1047053509157.279, Tc00.1047053510187.551, Tc00.1047053510725.99, Tc00.1047053511733.130, Tc00.1047053511735.10 |
| 1.3.1.- | 7 | Tc00.1047053507617.9, Tc00.1047053508461.80, Tc00.1047053509941.100, Tc00.1047053510303.210 |
| 1.3.99.10 | 1 | Tc00.1047053506367.10, Tc00.1047053506629.220, Tc00.1047053507001.80, Tc00.1047053508465.110, Tc00.1047053508827.40, Tc00.1047053508951.40, Tc00.1047053509153.120, Tc00.1047053509829.20, Tc00.1047053510121.40, Tc00.1047053510303.290 |
| 1.3.99.2 | 1 | Tc00.1047053503559.109, Tc00.1047053506629.220, Tc00.1047053507001.80, Tc00.1047053508465.110, Tc00.1047053508827.40, Tc00.1047053508951.40, Tc00.1047053509153.120, Tc00.1047053509829.20, Tc00.1047053510121.40, Tc00.1047053510303.290, Tc00.1047053511693.90 |
| 1.3.99.2 | 2 | Tc00.1047053506567.70 |
| 1.3.99.3 | 1 | Tc00.1047053506629.220, Tc00.1047053507001.80, Tc00.1047053508465.110, Tc00.1047053508827.40, Tc00.1047053508951.40, Tc00.1047053509153.120, Tc00.1047053509829.20, Tc00.1047053510121.40, Tc00.1047053510303.290 |
| 1.3.99.7 | 1 | Tc00.1047053506629.220, Tc00.1047053507001.80, Tc00.1047053508465.110, Tc00.1047053508951.40, Tc00.1047053509153.120, Tc00.1047053509829.20, Tc00.1047053510121.40, Tc00.1047053510303.290 |
| 1.4.1.2 | 1 | Tc00.1047053505843.10, Tc00.1047053507875.20, Tc00.1047053508111.30, Tc00.1047053509445.39 |
| 1.4.1.3 | 1 | Tc00.1047053507875.20, Tc00.1047053508111.30 |
| 1.4.1.4 | 1 | Tc00.1047053507875.20, Tc00.1047053508111.30 |
| 1.4.3.- | 2 | Tc00.1047053511909.40 |
| 1.4.3.16 | 1 | Tc00.1047053503849.60, Tc00.1047053508535.10, Tc00.1047053510215.10, Tc00.1047053511909.40 |
| 1.4.4.2 | 1 | Tc00.1047053422507.10, Tc00.1047053509163.80, Tc00.1047053510817.5, Tc00.1047053510911.50 |
| 1.4.4.2 | 3 | Tc00.1047053504643.30, Tc00.1047053509161.10 |
| 1.4.4.2 | 4 | Tc00.1047053506885.430, Tc00.1047053508457.30 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 1.5.1.- | 3 | Tc00.1047053507641.60, Tc00.1047053509351.10, Tc00.1047053510943.50 |
| 1.5.1.- | 7 | Tc00.1047053507617.9, Tc00.1047053508461.80, Tc00.1047053509941.100 |
| 1.5.1.12 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509351.10, Tc00.1047053510943.50, Tc00.1047053511751.120 |
| 1.5.1.2 | 1 | Tc00.1047053506857.20, Tc00.1047053509207.90 |
| 1.5.1.35 | 1 | Tc00.1047053506155.60, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053510943.50, Tc00.1047053511751.120 |
| 1.5.1.7 | 4 | Tc00.1047053506163.50, Tc00.1047053508177.90, Tc00.1047053508821.40, Tc00.1047053511691.39 |
| 1.5.3.1 | 1 | Tc00.1047053504017.40, Tc00.1047053509799.140, Tc00.1047053511037.20, Tc00.1047053511231.50 |
| 1.5.99.1 | 1 | Tc00.1047053504017.40, Tc00.1047053511231.50 |
| 1.5.99.8 | 1 | Tc00.1047053506155.60, Tc00.1047053506411.30, Tc00.1047053506943.4, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053509351.10, Tc00.1047053510943.50, Tc00.1047053511751.120 |
| 1.8.1.2 | 1 | Tc00.1047053503753.29, Tc00.1047053506585.70, Tc00.1047053506931.80, Tc00.1047053510657.170, Tc00.1047053510747.60, Tc00.1047053510877.120 |
| 1.8.1.2 | 2 | Tc00.1047053511367.230 |
| 1.8.1.4 | 1 | Tc00.1047053484299.10, Tc00.1047053503555.30, Tc00.1047053504507.5, Tc00.1047053506701.10, Tc00.1047053507089.270, Tc00.1047053507757.60, Tc00.1047053507757.70, Tc00.1047053508933.10, Tc00.1047053508933.20, Tc00.1047053509379.10, Tc00.1047053511025.110 |
| 1.8.1.7 | 1 | Tc00.1047053484299.10, Tc00.1047053503555.30, Tc00.1047053504507.5, Tc00.1047053506701.10, Tc00.1047053507089.270, Tc00.1047053507757.60, Tc00.1047053507757.70, Tc00.1047053508933.10, Tc00.1047053508933.20, Tc00.1047053509379.10, Tc00.1047053511025.110 |
| 2.1.1.- | 1 | Tc00.1047053506661.80, Tc00.1047053511245.79 |
| 2.1.1.- | 10 | Tc00.1047053507009.110, Tc00.1047053510667.50 |
| 2.1.1.- | 15 | Tc00.1047053507011.260, Tc00.1047053510663.80 |
| 2.1.1.- | 2 | Tc00.1047053503515.7, Tc00.1047053503823.160, Tc00.1047053504147.40, Tc00.1047053504427.90, Tc00.1047053504867.110, Tc00.1047053505807.220, Tc00.1047053506147.60, Tc00.1047053506363.130, Tc00.1047053506529.50, Tc00.1047053506853.30, Tc00.1047053506883.80, Tc00.1047053506947.80, Tc00.1047053507017.70, Tc00.1047053507057.30, Tc00.1047053507087.50, Tc00.1047053507207.19, Tc00.1047053507583.20, Tc00.1047053508153.1110, Tc00.1047053508153.370, Tc00.1047053508405.110, Tc00.1047053508461.30, Tc00.1047053508593.110, Tc00.1047053508741.290, Tc00.1047053509331.50, Tc00.1047053509805.40, Tc00.1047053509951.10, Tc00.1047053510187.240, Tc00.1047053510285.60, Tc00.1047053510311.140, Tc00.1047053510321.20, Tc00.1047053510429.20, Tc00.1047053510943.180, Tc00.1047053511269.20, Tc00.1047053511307.6, Tc00.1047053511619.5 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 2.1.1.- | 21 | Tc00.1047053508463.29, Tc00.1047053508625.120, Tc00.1047053509757.20, Tc00.1047053510659.130 |
| 2.1.1.- | 30 | Tc00.1047053503923.30, Tc00.1047053508583.10 |
| 2.1.1.- | 5 | Tc00.1047053504003.70, Tc00.1047053510515.130 |
| 2.1.1.- | 67 | Tc00.1047053439125.10, Tc00.1047053509455.120 |
| 2.1.1.10 | 1 | Tc00.1047053506479.120, Tc00.1047053508993.20 |
| 2.1.2.1 | 1 | Tc00.1047053509937.150, Tc00.1047053510407.90 |
| 2.1.2.1 | 4 | Tc00.1047053505051.30, Tc00.1047053506375.10, Tc00.1047053506375.60, Tc00.1047053507589.14, Tc00.1047053507693.5, Tc00.1047053510647.30, Tc00.1047053510671.20, Tc00.1047053510905.19 |
| 2.1.2.10 | 1 | Tc00.1047053504017.40, Tc00.1047053506801.110, Tc00.1047053511231.50 |
| 2.1.2.10 | 2 | Tc00.1047053506885.430, Tc00.1047053508457.30 |
| 2.1.2.10 | 3 | Tc00.1047053422507.10, Tc00.1047053509163.80, Tc00.1047053510911.50 |
| 2.1.2.9 | 1 | Tc00.1047053509267.100, Tc00.1047053510031.60 |
| 2.1.3.2 | 1 | Tc00.1047053506747.20, Tc00.1047053507059.70, Tc00.1047053507059.80, Tc00.1047053507091.50, Tc00.1047053508373.10, Tc00.1047053508373.20, Tc00.1047053508375.30, Tc00.1047053511643.30, Tc00.1047053511647.4, Tc00.1047053511923.110 |
| 2.1.3.3 | 1 | Tc00.1047053507091.50, Tc00.1047053508375.30, Tc00.1047053511643.30, Tc00.1047053511923.110 |
| 2.3.1.- | 100 | Tc00.1047053511315.20, Tc00.1047053511755.50 |
| 2.3.1.- | 11 | Tc00.1047053506227.230, Tc00.1047053506713.40, Tc00.1047053509719.9, Tc00.1047053511809.120, Tc00.1047053511811.30 |
| 2.3.1.- | 16 | Tc00.1047053508851.40, Tc00.1047053511587.120 |
| 2.3.1.- | 17 | Tc00.1047053407477.10, Tc00.1047053509463.30, Tc00.1047053510507.20, Tc00.1047053511003.60, Tc00.1047053511389.150 |
| 2.3.1.- | 25 | Tc00.1047053504157.20, Tc00.1047053511109.120 |
| 2.3.1.- | 28 | Tc00.1047053503935.20, Tc00.1047053506025.60, Tc00.1047053507601.70, Tc00.1047053509717.20, Tc00.1047053510105.170, Tc00.1047053510351.90 |
| 2.3.1.- | 46 | Tc00.1047053506605.160, Tc00.1047053507611.290, Tc00.1047053507723.110, Tc00.1047053509203.60, Tc00.1047053511239.150 |
| 2.3.1.- | 9 | Tc00.1047053506177.100, Tc00.1047053506181.104 |
| 2.3.1.1 | 2 | Tc00.1047053506713.40, Tc00.1047053509719.9 |
| 2.3.1.12 | 1 | Tc00.1047053503935.20, Tc00.1047053506025.60, Tc00.1047053507601.70, Tc00.1047053509717.20, Tc00.1047053510105.170, Tc00.1047053510351.90, Tc00.1047053511367.70 |
| 2.3.1.16 | 1 | Tc00.1047053407477.10, Tc00.1047053509463.30, Tc00.1047053510507.20, Tc00.1047053511003.60, Tc00.1047053511389.150, Tc00.1047053511393.10 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 2.3.1.29 | 1 | Tc00.1047053503453.100, Tc00.1047053506405.50, Tc00.1047053506959.70, Tc00.1047053511071.140, Tc00.1047053511899.10, Tc00.1047053511899.40 |
| 2.3.1.30 | 1 | Tc00.1047053504013.40, Tc00.1047053510879.80 |
| 2.3.1.31 | 2 | Tc00.1047053510661.250, Tc00.1047053510741.10 |
| 2.3.1.37 | 1 | Tc00.1047053503453.100, Tc00.1047053506405.50, Tc00.1047053506959.70, Tc00.1047053511071.140, Tc00.1047053511899.40 |
| 2.3.1.4 | 1 | Tc00.1047053508831.120, Tc00.1047053511671.70 |
| 2.3.1.61 | 1 | Tc00.1047053503555.30, Tc00.1047053503793.10, Tc00.1047053503935.20, Tc00.1047053506025.60, Tc00.1047053506295.160, Tc00.1047053506337.70, Tc00.1047053506701.10, Tc00.1047053507089.270, Tc00.1047053507601.70, Tc00.1047053507757.60, Tc00.1047053507757.70, Tc00.1047053508933.10, Tc00.1047053508933.20, Tc00.1047053509379.10, Tc00.1047053509717.20, Tc00.1047053510091.80, Tc00.1047053510105.170, Tc00.1047053510351.90, Tc00.1047053510421.320, Tc00.1047053510717.30, Tc00.1047053510797.10, Tc00.1047053511025.110, Tc00.1047053511469.100 |
| 2.3.1.7 | 1 | Tc00.1047053507211.10, Tc00.1047053508355.40, Tc00.1047053509611.140, Tc00.1047053509999.90, Tc00.1047053510857.10, Tc00.1047053511303.40, Tc00.1047053511763.19, Tc00.1047053511807.284 |
| 2.3.1.8 | 1 | Tc00.1047053505183.20, Tc00.1047053508647.270 |
| 2.3.1.9 | 1 | Tc00.1047053407477.10, Tc00.1047053509463.30, Tc00.1047053510507.20, Tc00.1047053511003.60, Tc00.1047053511389.150 |
| 2.3.3.10 | 1 | Tc00.1047053511903.40 |
| 2.4.2.- | 10 | Tc00.1047053506865.2, Tc00.1047053508989.9, Tc00.1047053509569.100 |
| 2.4.2.17 | 2 | Tc00.1047053506321.310, Tc00.1047053506839.80, Tc00.1047053507773.60, Tc00.1047053507809.30, Tc00.1047053508717.30, Tc00.1047053510431.250, Tc00.1047053511661.90 |
| 2.4.2.28 | 1 | Tc00.1047053506789.280, Tc00.1047053508387.20 |
| 2.5.1.16 | 1 | Tc00.1047053503855.20, Tc00.1047053504033.130, Tc00.1047053510337.40, Tc00.1047053510339.50 |
| 2.5.1.18 | 1 | Tc00.1047053510163.20 |
| 2.5.1.22 | 2 | Tc00.1047053503855.20, Tc00.1047053504033.130, Tc00.1047053510337.40, Tc00.1047053510339.50 |
| 2.5.1.47 | 1 | Tc00.1047053506905.50, Tc00.1047053507165.50, Tc00.1047053507793.20, Tc00.1047053508175.360, Tc00.1047053508177.110, Tc00.1047053508177.120, Tc00.1047053508177.129, Tc00.1047053508241.140, Tc00.1047053509149.9, Tc00.1047053510381.10, Tc00.1047053511691.10, Tc00.1047053511691.20 |
| 2.5.1.47 | 2 | Tc00.1047053510661.250, Tc00.1047053510739.19, Tc00.1047053510741.10 |
| 2.5.1.48 | 1 | Tc00.1047053510661.250, Tc00.1047053510739.19, Tc00.1047053510741.10 |
| 2.5.1.49 | 1 | Tc00.1047053510661.250, Tc00.1047053510739.19, Tc00.1047053510741.10 |
| 2.5.1.6 | 1 | Tc00.1047053506945.160, Tc00.1047053510445.50 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 2.6.1.- | 1 | Tc00.1047053506529.420, Tc00.1047053506529.430, Tc00.1047053510187.20, Tc00.1047053510187.30, Tc00.1047053510187.40, Tc00.1047053510187.50, Tc00.1047053510187.70, Tc00.1047053510565.11, Tc00.1047053510795.10, Tc00.1047053510889.120, Tc00.1047053510889.140, Tc00.1047053511461.20 |
| 2.6.1.- | 2 | Tc00.1047053506457.60, Tc00.1047053509693.100 |
| 2.6.1.- | 7 | Tc00.1047053503679.10, Tc00.1047053503841.70, Tc00.1047053510945.70 |
| 2.6.1.1 | 1 | Tc00.1047053503679.10, Tc00.1047053503747.10, Tc00.1047053503841.70, Tc00.1047053506529.420, Tc00.1047053506529.430, Tc00.1047053510187.20, Tc00.1047053510187.30, Tc00.1047053510187.40, Tc00.1047053510187.50, Tc00.1047053510187.70, Tc00.1047053510565.11, Tc00.1047053510795.10, Tc00.1047053510889.120, Tc00.1047053510889.140, Tc00.1047053510945.70, Tc00.1047053511461.20 |
| 2.6.1.16 | 1 | Tc00.1047053510303.200 |
| 2.6.1.17 | 1 | Tc00.1047053510187.20, Tc00.1047053510187.30, Tc00.1047053510187.40, Tc00.1047053510187.50, Tc00.1047053510187.70, Tc00.1047053510795.10, Tc00.1047053511461.20 |
| 2.6.1.2 | 1 | Tc00.1047053506529.420, Tc00.1047053506529.430, Tc00.1047053510889.120, Tc00.1047053510889.140 |
| 2.6.1.37 | 1 | Tc00.1047053506457.60, Tc00.1047053509693.100 |
| 2.6.1.44 | 1 | Tc00.1047053506457.60, Tc00.1047053509693.100 |
| 2.6.1.45 | 1 | Tc00.1047053506457.60, Tc00.1047053509693.100 |
| 2.6.1.5 | 1 | Tc00.1047053503747.10, Tc00.1047053508535.50, Tc00.1047053508537.5, Tc00.1047053509041.10, Tc00.1047053510187.20, Tc00.1047053510187.30, Tc00.1047053510187.40, Tc00.1047053510187.50, Tc00.1047053510187.70, Tc00.1047053510565.11, Tc00.1047053510795.10, Tc00.1047053511461.20 |
| 2.6.1.5 | 2 | Tc00.1047053503679.10, Tc00.1047053503841.70, Tc00.1047053510945.70 |
| 2.6.1.51 | 1 | Tc00.1047053506457.60, Tc00.1047053509693.100 |
| 2.6.1.57 | 1 | Tc00.1047053503679.10, Tc00.1047053503841.70, Tc00.1047053510945.70 |
| 2.6.1.57 | 2 | Tc00.1047053510187.20, Tc00.1047053510187.30, Tc00.1047053510187.40, Tc00.1047053510187.50, Tc00.1047053510187.70, Tc00.1047053510795.10, Tc00.1047053511461.20 |
| 2.6.1.7 | 1 | Tc00.1047053510187.70, Tc00.1047053511461.20 |
| 2.7.1.25 | 1 | Tc00.1047053506297.300, Tc00.1047053506357.40, Tc00.1047053508041.60, Tc00.1047053508461.130, Tc00.1047053508949.4, Tc00.1047053510119.20, Tc00.1047053510119.9, Tc00.1047053511277.70, Tc00.1047053511367.360, Tc00.1047053511367.370, Tc00.1047053511369.10, Tc00.1047053511369.20, Tc00.1047053511369.30, Tc00.1047053511369.5 |
| 2.7.1.31 | 1 | Tc00.1047053508741.170, Tc00.1047053509139.10 |
| 2.7.1.39 | 1 | Tc00.1047053503597.10, Tc00.1047053509167.150 |
| 2.7.1.71 | 3 | Tc00.1047053508177.20, Tc00.1047053511693.60 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 2.7.2.11 | 1 | Tc00.1047053509067.70, Tc00.1047053511023.10 |
| 2.7.2.4 | 6 | Tc00.1047053422507.10, Tc00.1047053509163.80, Tc00.1047053510911.50 |
| 2.7.3.2 | 1 | Tc00.1047053482369.29, Tc00.1047053507241.30 |
| 2.7.3.2 | 2 | Tc00.1047053508463.29, Tc00.1047053508625.120, Tc00.1047053509757.20, Tc00.1047053510659.130 |
| 2.7.3.3 | 1 | Tc00.1047053482369.29, Tc00.1047053507241.30 |
| 2.7.7.14 | 1 | Tc00.1047053511727.120 |
| 2.7.7.15 | 1 | Tc00.1047053511727.120 |
| 2.7.7.4 | 1 | Tc00.1047053503697.90, Tc00.1047053504013.50, Tc00.1047053504077.40, Tc00.1047053504105.210, Tc00.1047053504109.150, Tc00.1047053506235.10, Tc00.1047053506297.300, Tc00.1047053506357.40, Tc00.1047053506583.60, Tc00.1047053506677.39, Tc00.1047053507081.30, Tc00.1047053508041.60, Tc00.1047053508169.20, Tc00.1047053508169.9, Tc00.1047053508461.130, Tc00.1047053508601.130, Tc00.1047053508949.4, Tc00.1047053509199.10, Tc00.1047053510119.20, Tc00.1047053510119.9, Tc00.1047053510533.120, Tc00.1047053510879.70, Tc00.1047053510963.90, Tc00.1047053510965.5, Tc00.1047053511111.10, Tc00.1047053511277.70, Tc00.1047053511367.360, Tc00.1047053511367.370, Tc00.1047053511369.10, Tc00.1047053511369.20, Tc00.1047053511369.30, Tc00.1047053511369.5, Tc00.1047053511821.110 |
| 2.7.8.- | 2 | Tc00.1047053506559.40, Tc00.1047053506631.20 |
| 2.7.8.- | 4 | Tc00.1047053506795.40, Tc00.1047053509937.30 |
| 2.7.8.- | 8 | Tc00.1047053506581.10, Tc00.1047053508169.60 |
| 2.7.8.1 | 1 | Tc00.1047053508153.920, Tc00.1047053508181.40, Tc00.1047053509791.150, Tc00.1047053511039.3 |
| 2.7.8.2 | 1 | Tc00.1047053508153.920, Tc00.1047053508181.40, Tc00.1047053509791.150 |
| 2.7.8.8 | 1 | Tc00.1047053505071.100, Tc00.1047053511395.20 |
| 2.7.9.3 | 1 | Tc00.1047053504079.6, Tc00.1047053508717.36 |
| 2.8.1.2 | 1 | Tc00.1047053508173.40 |
| 2.8.3.5 | 1 | Tc00.1047053504153.360, Tc00.1047053506301.50 |
| 3.1.2.4 | 1 | Tc00.1047053506727.100, Tc00.1047053506727.90, Tc00.1047053508153.130, Tc00.1047053508185.10, Tc00.1047053511529.160, Tc00.1047053511529.170 |
| 3.1.4.3 | 4 | Tc00.1047053504149.160 |
| 3.3.1.1 | 1 | Tc00.1047053511229.50, Tc00.1047053511589.200 |
| 3.4.-.- | 15 | Tc00.1047053507681.200 |
| 3.4.-.- | 17 | Tc00.1047053506363.70, Tc00.1047053506363.90, Tc00.1047053506975.30, Tc00.1047053506975.40, Tc00.1047053507657.10, Tc00.1047053507657.20, Tc00.1047053507689.30, Tc00.1047053507689.40, Tc00.1047053507689.50, Tc00.1047053508971.40, Tc00.1047053509213.120, Tc00.1047053510257.80, Tc00.1047053510837.20, Tc00.1047053510837.9, Tc00.1047053511391.120 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 3.4.-.- | 20 | Tc00.1047053460127.20, Tc00.1047053510431.40 |
| 3.4.-.- | 27 | Tc00.1047053503697.9, Tc00.1047053507083.60, Tc00.1047053509627.30, Tc00.1047053509901.140, Tc00.1047053511649.60 |
| 3.4.-.- | 29 | Tc00.1047053507047.110, Tc00.1047053509179.150 |
| 3.4.-.- | 38 | Tc00.1047053504147.80, Tc00.1047053506411.10, Tc00.1047053511237.10 |
| 3.4.-.- | 39 | Tc00.1047053503671.10, Tc00.1047053506529.90, Tc00.1047053508153.60, Tc00.1047053508593.150, Tc00.1047053509683.130, Tc00.1047053510325.69, Tc00.1047053511051.70 |
| 3.4.-.- | 4 | Tc00.1047053503991.20, Tc00.1047053507081.110, Tc00.1047053510655.120 |
| 3.4.-.- | 40 | Tc00.1047053487507.10, Tc00.1047053504839.28, Tc00.1047053504839.44, Tc00.1047053505983.9, Tc00.1047053507259.10, Tc00.1047053509445.10, Tc00.1047053509499.14 |
| 3.4.-.- | 46 | Tc00.1047053506739.50, Tc00.1047053510819.60 |
| 3.4.-.- | 48 | Tc00.1047053508665.14, Tc00.1047053508737.100, Tc00.1047053508807.10, Tc00.1047053509125.20 |
| 3.4.-.- | 6 | Tc00.1047053508397.30, Tc00.1047053509885.30 |
| 3.4.-.- | 8 | Tc00.1047053510155.80, Tc00.1047053511181.50, Tc00.1047053511585.80 |
| 3.4.11.2 | 1 | Tc00.1047053506529.90, Tc00.1047053508153.60, Tc00.1047053508593.150, Tc00.1047053509683.130, Tc00.1047053510325.69, Tc00.1047053511051.70 |
| 3.4.11.4 | 1 | Tc00.1047053506513.110, Tc00.1047053508917.49 |
| 3.4.11.5 | 2 | Tc00.1047053504153.140, Tc00.1047053508799.240, Tc00.1047053509859.40, Tc00.1047053510515.60 |
| 3.4.11.5 | 3 | Tc00.1047053510655.120 |
| 3.5.1.- | 2 | Tc00.1047053447255.20, Tc00.1047053506559.80, Tc00.1047053507519.60, Tc00.1047053508207.150 |
| 3.5.1.- | 3 | Tc00.1047053508839.80, Tc00.1047053511381.50 |
| 3.5.1.- | 5 | Tc00.1047053506363.70, Tc00.1047053506363.90, Tc00.1047053506975.30, Tc00.1047053506975.40 |
| 3.5.1.- | 7 | Tc00.1047053506593.60, Tc00.1047053506829.80 |
| 3.5.1.1 | 2 | Tc00.1047053504147.110, Tc00.1047053505555.61, Tc00.1047053510823.80 |
| 3.5.1.14 | 1 | Tc00.1047053506363.70, Tc00.1047053506363.80, Tc00.1047053506363.90, Tc00.1047053506975.30, Tc00.1047053506975.34, Tc00.1047053506975.40 |
| 3.5.1.14 | 2 | Tc00.1047053509213.120, Tc00.1047053510257.80 |
| 3.5.1.16 | 1 | Tc00.1047053507657.10, Tc00.1047053507657.20, Tc00.1047053507689.30, Tc00.1047053507689.40, Tc00.1047053507689.50, Tc00.1047053508971.40, Tc00.1047053509213.120, Tc00.1047053510257.80, Tc00.1047053510837.20, Tc00.1047053510837.9, Tc00.1047053511391.120 |
| 3.5.1.16 | 2 | Tc00.1047053504105.30, Tc00.1047053509011.20 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 3.5.1.18 | 1 | Tc00.1047053507657.10, Tc00.1047053507657.20, Tc00.1047053507689.30, Tc00.1047053507689.40, Tc00.1047053508971.40, Tc00.1047053509213.120, Tc00.1047053510257.80, Tc00.1047053510837.20, Tc00.1047053510837.9, Tc00.1047053511391.120 |
| 3.5.1.3 | 1 | Tc00.1047053510241.120, Tc00.1047053510579.20 |
| 3.5.1.32 | 1 | Tc00.1047053506363.70, Tc00.1047053506363.90, Tc00.1047053506975.30, Tc00.1047053506975.40 |
| 3.5.1.4 | 2 | Tc00.1047053508839.80, Tc00.1047053511381.50 |
| 3.5.1.47 | 1 | Tc00.1047053506363.70, Tc00.1047053506363.90, Tc00.1047053506975.30, Tc00.1047053506975.40 |
| 3.5.1.54 | 1 | Tc00.1047053504835.20, Tc00.1047053509913.10 |
| 3.5.1.6 | 1 | Tc00.1047053509979.150, Tc00.1047053510039.40 |
| 3.5.1.78 | 1 | Tc00.1047053504427.10, Tc00.1047053508479.110, Tc00.1047053509099.50, Tc00.1047053509319.90, Tc00.1047053509331.134 |
| 3.5.2.- | 3 | Tc00.1047053509803.40 |
| 3.5.2.14 | 1 | Tc00.1047053508889.4, Tc00.1047053509803.40, Tc00.1047053511273.59 |
| 3.5.2.2 | 1 | Tc00.1047053506747.20, Tc00.1047053507059.70, Tc00.1047053508373.20 |
| 3.5.2.7 | 1 | Tc00.1047053441401.10, Tc00.1047053506247.220, Tc00.1047053508741.140, Tc00.1047053509137.30 |
| 3.5.2.9 | 1 | Tc00.1047053508889.4, Tc00.1047053509803.40, Tc00.1047053511273.59 |
| 3.5.3.1 | 1 | Tc00.1047053509497.30, Tc00.1047053510947.40 |
| 3.5.3.11 | 1 | Tc00.1047053507031.90, Tc00.1047053507963.20 |
| 3.5.3.8 | 2 | Tc00.1047053507031.90, Tc00.1047053507963.20 |
| 3.5.5.1 | 1 | Tc00.1047053509979.150, Tc00.1047053510039.40 |
| 3.7.1.2 | 1 | Tc00.1047053510657.50, Tc00.1047053510749.10 |
| 3.7.1.3 | 1 | Tc00.1047053503881.10, Tc00.1047053508119.110 |
| 3.7.1.5 | 1 | Tc00.1047053510657.50, Tc00.1047053510749.10 |
| 4.1.1.- | 11 | Tc00.1047053511017.20 |
| 4.1.1.- | 3 | Tc00.1047053505965.50, Tc00.1047053510903.50 |
| 4.1.1.- | 4 | Tc00.1047053510657.50, Tc00.1047053510749.10 |
| 4.1.1.- | 9 | Tc00.1047053509051.30, Tc00.1047053509965.380 |
| 4.1.1.15 | 1 | Tc00.1047053506941.150, Tc00.1047053511511.150 |
| 4.1.1.15 | 2 | Tc00.1047053503847.20, Tc00.1047053503905.10, Tc00.1047053504103.30, Tc00.1047053504867.120, Tc00.1047053506685.50, Tc00.1047053506779.190, Tc00.1047053507057.4, Tc00.1047053507093.210, Tc00.1047053507445.50, Tc00.1047053507739.110, Tc00.1047053507811.120, Tc00.1047053509663.10, Tc00.1047053510943.190, Tc00.1047053511151.100, Tc00.1047053511527.70, Tc00.1047053511649.80 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 4.1.1.25 | 1 | Tc00.1047053506941.150, Tc00.1047053511511.150 |
| 4.1.1.50 | 1 | Tc00.1047053504257.30, Tc00.1047053504257.60, Tc00.1047053509167.110, Tc00.1047053509167.120 |
| 4.1.1.65 | 1 | Tc00.1047053407335.9, Tc00.1047053508459.18, Tc00.1047053509551.149 |
| 4.1.1.68 | 1 | Tc00.1047053510657.50, Tc00.1047053510749.10 |
| 4.1.1.9 | 1 | Tc00.1047053506401.50, Tc00.1047053509603.19, Tc00.1047053509605.10 |
| 4.1.2.- | 2 | Tc00.1047053503991.39, Tc00.1047053507081.130 |
| 4.1.2.11 | 1 | Tc00.1047053508671.20, Tc00.1047053509695.210, Tc00.1047053509695.220, Tc00.1047053509695.230 |
| 4.1.3.- | 6 | Tc00.1047053503991.39, Tc00.1047053507081.130 |
| 4.1.3.27 | 1 | Tc00.1047053511807.110 |
| 4.1.3.4 | 1 | Tc00.1047053506635.80, Tc00.1047053508027.129, Tc00.1047053508029.10 |
| 4.2.1.- | 15 | Tc00.1047053507257.60, Tc00.1047053507669.10, Tc00.1047053509879.40, Tc00.1047053510329.270 |
| 4.2.1.- | 32 | Tc00.1047053437575.9, Tc00.1047053508897.10, Tc00.1047053510027.10 |
| 4.2.1.- | 5 | Tc00.1047053506727.100, Tc00.1047053506727.90, Tc00.1047053506959.40, Tc00.1047053507547.40, Tc00.1047053508153.130, Tc00.1047053508185.10, Tc00.1047053508441.70, Tc00.1047053508981.39, Tc00.1047053510997.40, Tc00.1047053511277.210, Tc00.1047053511529.160, Tc00.1047053511529.170 |
| 4.2.1.- | 6 | Tc00.1047053506567.70, Tc00.1047053507049.60 |
| 4.2.1.11 | 1 | Tc00.1047053504105.140, Tc00.1047053511029.10, Tc00.1047053511529.90 |
| 4.2.1.17 | 1 | Tc00.1047053504233.4, Tc00.1047053506727.100, Tc00.1047053506727.90, Tc00.1047053506959.40, Tc00.1047053507107.40, Tc00.1047053507547.40, Tc00.1047053508153.130, Tc00.1047053508185.10, Tc00.1047053508441.70, Tc00.1047053508981.39, Tc00.1047053509261.30, Tc00.1047053509701.10, Tc00.1047053509717.90, Tc00.1047053510105.240, Tc00.1047053510997.40, Tc00.1047053511277.210, Tc00.1047053511529.160, Tc00.1047053511529.170 |
| 4.2.1.18 | 1 | Tc00.1047053506959.40, Tc00.1047053508153.130, Tc00.1047053508185.10, Tc00.1047053510997.40, Tc00.1047053511277.210 |
| 4.2.1.22 | 1 | Tc00.1047053447925.10, Tc00.1047053506905.50, Tc00.1047053507165.50, Tc00.1047053507793.20, Tc00.1047053508175.360, Tc00.1047053508177.110, Tc00.1047053508177.120, Tc00.1047053508177.129, Tc00.1047053508241.140, Tc00.1047053508241.149, Tc00.1047053509149.9, Tc00.1047053510381.10, Tc00.1047053511691.10, Tc00.1047053511691.20 |
| 4.2.1.33 | 1 | Tc00.1047053511277.290 |
| 4.2.1.36 | 1 | Tc00.1047053511277.290 |
| 4.2.1.49 | 1 | Tc00.1047053504045.110, Tc00.1047053504047.5 |
| 4.2.1.9 | 1 | Tc00.1047053511003.190 |
| 4.2.3.1 | 1 | Tc00.1047053504213.80, Tc00.1047053508857.160 |
| 4.2.3.1 | 2 | Tc00.1047053505183.80, Tc00.1047053506337.210 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 4.3.1.19 | 1 | Tc00.1047053506825.70 |
| 4.3.1.3 | 1 | Tc00.1047053506247.220, Tc00.1047053508741.140, Tc00.1047053509137.30 |
| 4.3.1.5 | 1 | Tc00.1047053506247.220 |
| 4.3.2.2 | 1 | Tc00.1047053503855.30, Tc00.1047053510337.30 |
| 4.4.1.1 | 1 | Tc00.1047053510661.250, Tc00.1047053510739.19, Tc00.1047053510741.10 |
| 4.4.1.1 | 3 | Tc00.1047053506629.70, Tc00.1047053508465.10, Tc00.1047053509831.10, Tc00.1047053510659.44 |
| 4.4.1.11 | 1 | Tc00.1047053510661.250, Tc00.1047053510739.19, Tc00.1047053510741.10 |
| 4.4.1.16 | 1 | Tc00.1047053506629.70, Tc00.1047053508465.10, Tc00.1047053509831.10, Tc00.1047053510659.44 |
| 4.4.1.16 | 2 | Tc00.1047053509163.80, Tc00.1047053510911.50 |
| 4.4.1.16 | 3 | Tc00.1047053510661.250, Tc00.1047053510739.19, Tc00.1047053510741.10 |
| 4.4.1.8 | 1 | Tc00.1047053510661.250, Tc00.1047053510739.19, Tc00.1047053510741.10 |
| 5.1.1.1 | 1 | Tc00.1047053508303.10, Tc00.1047053511179.150 |
| 5.1.1.10 | 1 | Tc00.1047053506681.70, Tc00.1047053506825.70 |
| 5.1.1.11 | 1 | Tc00.1047053503559.100, Tc00.1047053504157.20, Tc00.1047053511109.120, Tc00.1047053511693.100 |
| 5.1.1.4 | 1 | Tc00.1047053506795.80, Tc00.1047053509935.29 |
| 5.3.1.23 | 1 | Tc00.1047053503971.40, Tc00.1047053508269.30 |
| 5.3.3.- | 1 | Tc00.1047053506959.40, Tc00.1047053508153.130, Tc00.1047053508185.10, Tc00.1047053510997.40, Tc00.1047053511277.210 |
| 5.3.3.- | 2 | Tc00.1047053510657.50, Tc00.1047053510749.10 |
| 5.3.3.10 | 1 | Tc00.1047053510657.50, Tc00.1047053510749.10 |
| 5.4.2.9 | 1 | Tc00.1047053508851.189, Tc00.1047053511589.140 |
| 6.1.1.1 | 1 | Tc00.1047053506573.30, Tc00.1047053508321.50, Tc00.1047053509207.50, Tc00.1047053509937.180, Tc00.1047053511217.200 |
| 6.1.1.10 | 1 | Tc00.1047053506573.30, Tc00.1047053508321.50, Tc00.1047053509207.50, Tc00.1047053509247.50, Tc00.1047053511217.200 |
| 6.1.1.11 | 1 | Tc00.1047053506777.80, Tc00.1047053511163.10 |
| 6.1.1.11 | 2 | Tc00.1047053506155.60, Tc00.1047053507023.120, Tc00.1047053507641.60, Tc00.1047053511751.120 |
| 6.1.1.12 | 1 | Tc00.1047053503815.20, Tc00.1047053505807.120, Tc00.1047053506559.350, Tc00.1047053507233.94, Tc00.1047053508971.30, Tc00.1047053509037.50, Tc00.1047053509039.4, Tc00.1047053509695.10, Tc00.1047053509967.30, Tc00.1047053510777.20, Tc00.1047053511911.21 |
| 6.1.1.14 | 2 | Tc00.1047053504017.79, Tc00.1047053511229.110, Tc00.1047053511589.260 |
| 6.1.1.15 | 1 | Tc00.1047053503939.80, Tc00.1047053506831.70, Tc00.1047053507913.39, Tc00.1047053508277.160, Tc00.1047053509805.130, Tc00.1047053511731.60 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 6.1.1.15 | 2 | Tc00.1047053509167.100, Tc00.1047053509797.40, Tc00.1047053510443.9 |
| 6.1.1.16 | 1 | Tc00.1047053506531.60, Tc00.1047053506905.50, Tc00.1047053507165.50, Tc00.1047053507297.20, Tc00.1047053507537.30, Tc00.1047053507793.20, Tc00.1047053508175.360, Tc00.1047053508177.110, Tc00.1047053508177.120, Tc00.1047053508177.129, Tc00.1047053508241.140, Tc00.1047053509149.9, Tc00.1047053509399.9, Tc00.1047053510381.10, Tc00.1047053511691.10, Tc00.1047053511691.20 |
| 6.1.1.16 | 3 | Tc00.1047053503939.80, Tc00.1047053509805.130 |
| 6.1.1.17 | 1 | Tc00.1047053503939.80, Tc00.1047053506831.70, Tc00.1047053507913.39, Tc00.1047053508277.160, Tc00.1047053509805.130, Tc00.1047053511731.60 |
| 6.1.1.18 | 1 | Tc00.1047053506831.70, Tc00.1047053507913.39, Tc00.1047053507915.5, Tc00.1047053508277.160, Tc00.1047053511731.60 |
| 6.1.1.19 | 1 | Tc00.1047053508355.320 |
| 6.1.1.2 | 1 | Tc00.1047053505051.30, Tc00.1047053506375.10, Tc00.1047053506375.60, Tc00.1047053507589.14, Tc00.1047053507693.5, Tc00.1047053509365.61, Tc00.1047053510647.30, Tc00.1047053510671.20, Tc00.1047053510905.19 |
| 6.1.1.20 | 1 | Tc00.1047053506127.30, Tc00.1047053508271.10, Tc00.1047053511727.140 |
| 6.1.1.21 | 1 | Tc00.1047053507019.40 |
| 6.1.1.22 | 1 | Tc00.1047053506559.350, Tc00.1047053509695.10, Tc00.1047053509967.30, Tc00.1047053510777.20 |
| 6.1.1.3 | 1 | Tc00.1047053508299.80, Tc00.1047053511181.80 |
| 6.1.1.4 | 1 | Tc00.1047053509167.100, Tc00.1047053509679.9, Tc00.1047053509797.40, Tc00.1047053510443.9, Tc00.1047053511049.40 |
| 6.1.1.5 | 1 | Tc00.1047053504137.10, Tc00.1047053509167.100, Tc00.1047053509797.40, Tc00.1047053510443.9 |
| 6.1.1.5 | 2 | Tc00.1047053503939.80, Tc00.1047053509805.130 |
| 6.1.1.6 | 1 | Tc00.1047053503815.20, Tc00.1047053505807.120, Tc00.1047053508971.30, Tc00.1047053510777.20 |
| 6.1.1.7 | 1 | Tc00.1047053504883.50, Tc00.1047053506511.30, Tc00.1047053506863.10, Tc00.1047053508299.80, Tc00.1047053511181.80, Tc00.1047053511825.220 |
| 6.1.1.7 | 2 | Tc00.1047053509153.90, Tc00.1047053510303.320 |
| 6.1.1.9 | 1 | Tc00.1047053504257.70, Tc00.1047053509167.100, Tc00.1047053509797.40, Tc00.1047053510443.9 |
| 6.3.1.1 | 1 | Tc00.1047053503625.10, Tc00.1047053503899.90 |
| 6.3.1.2 | 1 | Tc00.1047053503405.10, Tc00.1047053508175.370 |
| 6.3.1.8 | 1 | Tc00.1047053504427.10, Tc00.1047053508479.110, Tc00.1047053509099.50, Tc00.1047053509319.90, Tc00.1047053509331.134 |
| 6.3.2.- | 1 | Tc00.1047053509611.100, Tc00.1047053510857.50 |
| 6.3.2.- | 11 | Tc00.1047053503559.100, Tc00.1047053504427.110, Tc00.1047053509331.30, Tc00.1047053511693.100 |
| 6.3.2.- | 14 | Tc00.1047053508303.4, Tc00.1047053511179.160 |
| 6.3.2.- | 15 | Tc00.1047053511003.80 |

| EC NUMBER | CLUSTER | TCR HIT |
|---|---|---|
| 6.3.2.- | 2 | Tc00.1047053504147.190, Tc00.1047053506275.90, Tc00.1047053507099.40, Tc00.1047053507515.70, Tc00.1047053507515.90, Tc00.1047053507669.210, Tc00.1047053507875.10, Tc00.1047053508111.40, Tc00.1047053508211.60, Tc00.1047053508213.9, Tc00.1047053508231.60, Tc00.1047053508257.250, Tc00.1047053508439.30, Tc00.1047053508547.201, Tc00.1047053508971.50, Tc00.1047053509053.10, Tc00.1047053509875.250, Tc00.1047053509965.350, Tc00.1047053510747.170, Tc00.1047053510749.4, Tc00.1047053511725.160 |
| 6.3.2.- | 21 | Tc00.1047053507797.9, Tc00.1047053510165.30, Tc00.1047053510645.20, Tc00.1047053511807.60 |
| 6.3.2.- | 22 | Tc00.1047053506477.30, Tc00.1047053506743.210, Tc00.1047053506775.30, Tc00.1047053506885.270, Tc00.1047053506941.80, Tc00.1047053508409.170, Tc00.1047053508797.10, Tc00.1047053508995.30, Tc00.1047053509253.20, Tc00.1047053510141.20, Tc00.1047053510729.150, Tc00.1047053511169.29, Tc00.1047053511277.440, Tc00.1047053511511.80 |
| 6.3.2.- | 27 | Tc00.1047053504427.250, Tc00.1047053508043.30, Tc00.1047053509777.100, Tc00.1047053510155.120, Tc00.1047053511585.130, Tc00.1047053511655.69 |
| 6.3.2.- | 4 | Tc00.1047053503897.40, Tc00.1047053509563.30 |
| 6.3.2.2 | 5 | Tc00.1047053507625.99, Tc00.1047053507787.100 |
| 6.3.2.3 | 2 | Tc00.1047053506659.30, Tc00.1047053508865.10 |
| 6.3.4.16 | 1 | Tc00.1047053507059.80, Tc00.1047053507091.50, Tc00.1047053508373.10, Tc00.1047053508375.30, Tc00.1047053511643.30, Tc00.1047053511647.4, Tc00.1047053511923.110 |
| 6.3.4.16 | 2 | Tc00.1047053504835.20, Tc00.1047053509913.10 |
| 6.3.4.4 | 1 | Tc00.1047053508731.60, Tc00.1047053509683.80 |
| 6.3.4.6 | 1 | Tc00.1047053504835.20, Tc00.1047053509913.10 |
| 6.3.5.1 | 1 | Tc00.1047053427789.30, Tc00.1047053506221.80 |
| 6.3.5.2 | 1 | Tc00.1047053508085.10, Tc00.1047053511807.110 |
| 6.3.5.5 | 1 | Tc00.1047053506747.20, Tc00.1047053507059.70, Tc00.1047053507059.80, Tc00.1047053507091.50, Tc00.1047053508373.10, Tc00.1047053508373.20, Tc00.1047053508375.30, Tc00.1047053511643.30, Tc00.1047053511647.4, Tc00.1047053511923.110 |
| 6.4.1.1 | 1 | Tc00.1047053504835.20, Tc00.1047053506773.40, Tc00.1047053508373.10, Tc00.1047053508799.170, Tc00.1047053509913.10, Tc00.1047053511647.4 |
| 6.4.1.3 | 1 | Tc00.1047053504835.20, Tc00.1047053506773.40, Tc00.1047053508799.170, Tc00.1047053509913.10 |
| 6.4.1.4 | 1 | Tc00.1047053504835.20, Tc00.1047053509913.10 |
| 6.4.1.4 | 2 | Tc00.1047053506773.40, Tc00.1047053508799.170 |

**Anexo II:** Material Suplementar referente ao artigo “Structural Modelling and Comparative Analysis of Homologous, Analogous and Specific Proteins from *Trypanosoma cruzi* versus *Homo sapiens*: Putative Drug Targets for Chagas' Disease Treatment”.

A) Tabela com a lista completa dos modelos tridimensionais das enzimas homólogas, análogas e específicas de *T. cruzi versus H. sapiens.*

| CATEGORIES | | MHOLline QUALITY | TCR ID | EC NUMBER | EC DESCRIPTION |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 1 | Tc00.1047053510187.30 | 2.6.1.5 | Tyrosinetransaminase |
| Matching Annotation | Homologous | 1 | Tc00.1047053510187.40 | 2.6.1.5 | Tyrosinetransaminase |
| Matching Annotation | Homologous | 1 | Tc00.1047053510187.50 | 2.6.1.5 | Tyrosinetransaminase |
| Matching Annotation | Homologous | 1 | Tc00.1047053505999.90 | 2.7.2.3 | Phosphoglyceratekinase |
| Matching Annotation | Homologous | 1 | Tc00.1047053509065.60 | 1.2.1.12 | Glyceraldehyde-3-phosphatedehydrogenase(phosphorylating) |
| Matching Annotation | Homologous | 1 | Tc00.1047053509065.70 | 1.2.1.12 | Glyceraldehyde-3-phosphatedehydrogenase(phosphorylating) |
| Matching Annotation | Homologous | 1 | Tc00.1047053506943.50 | 1.2.1.12 | Glyceraldehyde-3-phosphatedehydrogenase(phosphorylating) |
| Matching Annotation | Homologous | 1 | Tc00.1047053506943.60 | 1.2.1.12 | Glyceraldehyde-3-phosphatedehydrogenase(phosphorylating) |
| Matching Annotation | Homologous | 1 | Tc00.1047053511461.20 | 2.6.1.5 | Tyrosinetransaminase |
| Matching Annotation | Homologous | 1 | Tc00.1047053511419.40 | 2.7.2.3 | Phosphoglyceratekinase |
| Matching Annotation | Homologous | 1 | Tc00.1047053510663.70 | 1.1.1.44 | Phosphogluconatedehydrogenase(decarboxylating) |
| Matching Annotation | Homologous | 1 | Tc00.1047053510795.10 | 2.6.1.5 | Tyrosinetransaminase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 2 | Tc00.1047053508153.130 | 4.2.1.17 | Enoyl-CoAhydratase |
| Matching Annotation | Homologous | 2 | Tc00.1047053506925.300 | 5.2.1.8 | Peptidylprolylisomerase |
| Matching Annotation | Homologous | 2 | Tc00.1047053506925.319 | 1.1.1.42 | Isocitratedehydrogenase(NADP(+)) |
| Matching Annotation | Homologous | 2 | Tc00.1047053510187.60 | 1.2.1.12 | Glyceraldehyde-3-phosphatedehydrogenase(phosphorylating) |
| Matching Annotation | Homologous | 2 | Tc00.1047053508175.360 | 4.2.1.22 | Cystathioninebeta-synthase |
| Matching Annotation | Homologous | 2 | Tc00.1047053507993.80 | 2.7.11.26 | [Tauprotein]kinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053509965.380 | 4.1.1.35 | UDP-glucuronatedecarboxylase |
| Matching Annotation | Homologous | 2 | Tc00.1047053510105.230 | 1.2.1.12 | Glyceraldehyde-3-phosphatedehydrogenase(phosphorylating) |
| Matching Annotation | Homologous | 2 | Tc00.1047053510303.290 | 1.3.99.3 | Acyl-CoAdehydrogenase |
| Matching Annotation | Homologous | 2 | Tc00.1047053504427.110 | 6.2.1.1 | Acetate--CoAligase |
| Matching Annotation | Homologous | 2 | Tc00.1047053510421.320 | 1.2.4.1 | Pyruvatedehydrogenase(acetyl-transferring) |
| Matching Annotation | Homologous | 2 | Tc00.1047053507089.270 | 1.8.1.4 | Dihydrolipoyldehydrogenase |
| Matching Annotation | Homologous | 2 | Tc00.1047053511903.30 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 2 | Tc00.1047053509233.180 | 3.6.3.14 | H(+)-transportingtwo-sectorATPase |
| Matching Annotation | Homologous | 2 | Tc00.1047053509695.190 | 2.7.11.22 | Cyclin-dependentkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053509331.30 | 6.2.1.1 | Acetate--CoAligase |
| Matching Annotation | Homologous | 2 | Tc00.1047053508541.220 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053508541.230 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 2 | Tc00.1047053508541.240 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053508815.110 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 2 | Tc00.1047053509717.80 | 1.2.1.12 | Glyceraldehyde-3-phosphatedehydrogenase(phosphorylating) |
| Matching Annotation | Homologous | 2 | Tc00.1047053511301.110 | 1.1.1.205 | IMPdehydrogenase |
| Matching Annotation | Homologous | 2 | Tc00.1047053509153.120 | 1.3.99.3 | Acyl-CoAdehydrogenase |
| Matching Annotation | Homologous | 2 | Tc00.1047053510609.70 | 2.7.11.22 | Cyclin-dependentkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053511019.90 | 1.15.1.1 | Superoxidedismutase |
| Matching Annotation | Homologous | 2 | Tc00.1047053511071.140 | 2.3.1.29 | GlycineC-acetyltransferase |
| Matching Annotation | Homologous | 2 | Tc00.1047053511071.70 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 2 | Tc00.1047053506457.10 | 3.4.11.18 | Methionylaminopeptidase |
| Matching Annotation | Homologous | 2 | Tc00.1047053506201.30 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 2 | Tc00.1047053506201.70 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 2 | Tc00.1047053506201.80 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 2 | Tc00.1047053511575.60 | 1.1.1.42 | Isocitratedehydrogenase(NADP(+)) |
| Matching Annotation | Homologous | 2 | Tc00.1047053510761.60 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053508241.140 | 4.2.1.22 | Cystathioninebeta-synthase |
| Matching Annotation | Homologous | 2 | Tc00.1047053511179.160 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 2 | Tc00.1047053508177.110 | 4.2.1.22 | Cystathioninebeta-synthase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 2 | Tc00.1047053508177.120 | 4.2.1.22 | Cystathioninebeta-synthase |
| Matching Annotation | Homologous | 2 | Tc00.1047053508303.4 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 2 | Tc00.1047053511025.110 | 1.8.1.4 | Dihydrolipoyldehydrogenase |
| Matching Annotation | Homologous | 2 | Tc00.1047053510091.80 | 1.2.4.1 | Pyruvatedehydrogenase(acetyl-transferring) |
| Matching Annotation | Homologous | 2 | Tc00.1047053507883.109 | 1.1.1.37 | Malatedehydrogenase |
| Matching Annotation | Homologous | 2 | Tc00.1047053507757.50 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 2 | Tc00.1047053510381.10 | 4.2.1.22 | Cystathioninebeta-synthase |
| Matching Annotation | Homologous | 2 | Tc00.1047053506905.50 | 4.2.1.22 | Cystathioninebeta-synthase |
| Matching Annotation | Homologous | 2 | Tc00.1047053503617.10 | 2.7.11.22 | Cyclin-dependentkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053509499.14 | 1.11.1.15 | Peroxiredoxin |
| Matching Annotation | Homologous | 2 | Tc00.1047053506583.40 | 2.7.11.22 | Cyclin-dependentkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053509051.30 | 4.1.1.35 | UDP-glucuronatedecarboxylase |
| Matching Annotation | Homologous | 2 | Tc00.1047053509633.50 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 2 | Tc00.1047053509633.60 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 2 | Tc00.1047053507211.40 | 1.1.1.205 | IMPdehydrogenase |
| Matching Annotation | Homologous | 2 | Tc00.1047053510647.30 | 6.1.1.2 | Tryptophan--tRNAligase |
| Matching Annotation | Homologous | 2 | Tc00.1047053508137.30 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 2 | Tc00.1047053511899.40 | 2.3.1.29 | GlycineC-acetyltransferase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 2 | Tc00.1047053507541.30 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053511691.20 | 4.2.1.22 | Cystathioninebeta-synthase |
| Matching Annotation | Homologous | 2 | Tc00.1047053504181.40 | 2.7.11.22 | Cyclin-dependentkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053507305.20 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053507305.30 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053506677.20 | 2.7.11.22 | Cyclin-dependentkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053510027.10 | 4.2.1.47 | GDP-mannose4,6-dehydratase |
| Matching Annotation | Homologous | 2 | Tc00.1047053504929.10 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 2 | Tc00.1047053508185.10 | 4.2.1.17 | Enoyl-CoAhydratase |
| Matching Annotation | Homologous | 3 | Tc00.1047053507031.120 | 3.4.11.18 | Methionylaminopeptidase |
| Matching Annotation | Homologous | 3 | Tc00.1047053509693.50 | 3.4.11.18 | Methionylaminopeptidase |
| Matching Annotation | Homologous | 3 | Tc00.1047053505999.100 | 2.7.2.3 | Phosphoglyceratekinase |
| Matching Annotation | Homologous | 3 | Tc00.1047053508541.225 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 3 | Tc00.1047053506795.80 | 5.1.1.4 | Prolineracemase |
| Matching Annotation | Homologous | 3 | Tc00.1047053508897.110 | 5.2.1.8 | Peptidylprolylisomerase |
| Matching Annotation | Homologous | 3 | Tc00.1047053510755.138 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 3 | Tc00.1047053508413.40 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 3 | Tc00.1047053504111.20 | 3.4.11.18 | Methionylaminopeptidase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 3 | Tc00.1047053511419.50 | 2.7.2.3 | Phosphoglyceratekinase |
| Matching Annotation | Homologous | 3 | Tc00.1047053509149.9 | 4.2.1.22 | Cystathioninebeta-synthase |
| Matching Annotation | Homologous | 3 | Tc00.1047053510259.50 | 5.2.1.8 | Peptidylprolylisomerase |
| Matching Annotation | Homologous | 3 | Tc00.1047053453445.20 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511277.630 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506885.400 | 5.2.1.8 | Peptidylprolylisomerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510187.20 | 2.6.1.5 | Tyrosinetransaminase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510187.234 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510187.500 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510187.70 | 2.6.1.5 | Tyrosinetransaminase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508355.380 | 3.1.2.15 | Ubiquitinthiolesterase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508741.320 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506435.300 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053507993.190 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506855.180 | 2.7.4.3 | Adenylatekinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506247.230 | 3.4.21.26 | Prolyloligopeptidase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508277.160 | 6.1.1.18 | Glutamine--tRNAligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509429.290 | 3.1.2.6 | Hydroxyacylglutathionehydrolase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 4 | Tc00.1047053510105.240 | 1.1.1.35 | 3-hydroxyacyl-CoAdehydrogenase |
| Matching Annotation | Homologous | 4 | Tc00.1047053507023.120 | 1.2.1.5 | Aldehydedehydrogenase(NAD(P)(+)) |
| Matching Annotation | Homologous | 4 | Tc00.1047053507023.200 | 2.4.2.29 | QueuinetRNA-ribosyltransferase |
| Matching Annotation | Homologous | 4 | Tc00.1047053507641.60 | 1.2.1.3 | Aldehydedehydrogenase(NAD(+)) |
| Matching Annotation | Homologous | 4 | Tc00.1047053506321.310 | 2.7.6.1 | Ribose-phosphatediphosphokinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511817.40 | 1.6.2.2 | Cytochrome-b5reductase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510431.250 | 2.7.6.1 | Ribose-phosphatediphosphokinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510421.180 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506195.80 | 2.7.4.3 | Adenylatekinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506945.110 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506945.20 | 2.7.2.3 | Phosphoglyceratekinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510729.10 | 3.5.4.6 | AMPdeaminase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511421.60 | 2.5.1.46 | Deoxyhypusinesynthase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509733.180 | 2.7.4.3 | Adenylatekinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053507603.230 | 3.1.2.6 | Hydroxyacylglutathionehydrolase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511751.120 | 1.2.1.5 | Aldehydedehydrogenase(NAD(P)(+)) |
| Matching Annotation | Homologous | 4 | Tc00.1047053511751.30 | 2.4.2.29 | QueuinetRNA-ribosyltransferase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508647.270 | 1.1.1.40 | Malatedehydrogenase(oxaloacetate-decarboxylating)(NADP(+)) |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 4 | Tc00.1047053508647.280 | 1.1.1.40 | Malatedehydrogenase(oxaloacetate-decarboxylating)(NADP(+)) |
| Matching Annotation | Homologous | 4 | Tc00.1047053510661.60 | 2.7.1.30 | Glycerolkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509695.10 | 6.1.1.12 | Aspartate--tRNAligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053507053.70 | 2.5.1.60 | ProteingeranylgeranyltransferasetypeII |
| Matching Annotation | Homologous | 4 | Tc00.1047053506503.69 | 1.1.1.37 | Malatedehydrogenase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508181.140 | 3.6.1.1 | Inorganicdiphosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510089.170 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509167.100 | 6.1.1.9 | Valine--tRNAligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509109.120 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509109.130 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053505807.120 | 6.1.1.6 | Lysine--tRNAligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509179.140 | 3.1.3.25 | Inositol-phosphatephosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053507047.120 | 3.1.3.25 | Inositol-phosphatephosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506743.110 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506743.130 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510861.140 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508577.160 | 3.4.25.1 | Proteasomeendopeptidasecomplex |
| Matching Annotation | Homologous | 4 | Tc00.1047053509647.70 | 2.7.4.8 | Guanylatekinase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 4 | Tc00.1047053507521.50 | 3.4.25.1 | Proteasomeendopeptidasecomplex |
| Matching Annotation | Homologous | 4 | Tc00.1047053509717.90 | 1.1.1.35 | 3-hydroxyacyl-CoAdehydrogenase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506779.50 | 3.4.25.1 | Proteasomeendopeptidasecomplex |
| Matching Annotation | Homologous | 4 | Tc00.1047053510603.60 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510603.80 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510575.180 | 2.7.4.3 | Adenylatekinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511859.170 | 1.6.2.2 | Cytochrome-b5reductase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509967.30 | 6.1.1.12 | Aspartate--tRNAligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053505183.20 | 1.1.1.40 | Malatedehydrogenase(oxaloacetate-decarboxylating)(NADP(+)) |
| Matching Annotation | Homologous | 4 | Tc00.1047053505183.30 | 1.1.1.40 | Malatedehydrogenase(oxaloacetate-decarboxylating)(NADP(+)) |
| Matching Annotation | Homologous | 4 | Tc00.1047053507831.70 | 1.2.4.1 | Pyruvatedehydrogenase(acetyl-transferring) |
| Matching Annotation | Homologous | 4 | Tc00.1047053508569.20 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511153.140 | 3.4.25.1 | Proteasomeendopeptidasecomplex |
| Matching Annotation | Homologous | 4 | Tc00.1047053509213.70 | 5.1.3.1 | Ribulose-phosphate3-epimerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506835.70 | 2.7.2.3 | Phosphoglyceratekinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510667.60 | 5.2.1.8 | Peptidylprolylisomerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510515.60 | 3.4.11.1 | Leucylaminopeptidase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506649.70 | 3.1.3.11 | Fructose-bisphosphatase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 4 | Tc00.1047053509011.20 | 2.7.7.6 | DNA-directedRNApolymerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509607.10 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509607.70 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053504055.81 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511293.69 | 1.1.1.37 | Malatedehydrogenase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506839.60 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506839.80 | 2.7.6.1 | Ribose-phosphatediphosphokinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509775.40 | 1.15.1.1 | Superoxidedismutase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511837.70 | 2.7.11.24 | Mitogen-activatedproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053507009.100 | 5.2.1.8 | Peptidylprolylisomerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510257.30 | 5.1.3.1 | Ribulose-phosphate3-epimerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508781.110 | 1.2.4.4 | 3-methyl-2-oxobutanoatedehydrogenase(2-methylpropanoyl-transferring) |
| Matching Annotation | Homologous | 4 | Tc00.1047053507061.30 | 1.15.1.1 | Superoxidedismutase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508717.10 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510243.50 | 2.7.4.3 | Adenylatekinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508971.30 | 6.1.1.6 | Lysine--tRNAligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503479.30 | 2.7.4.3 | Adenylatekinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508357.40 | 3.1.2.15 | Ubiquitinthiolesterase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 4 | Tc00.1047053509047.110 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506585.70 | 1.6.2.4 | NADPH--hemoproteinreductase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511857.80 | 3.5.4.6 | AMPdeaminase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510131.50 | 2.7.7.7 | DNA-directedDNApolymerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510131.70 | 2.7.7.7 | DNA-directedDNApolymerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053509859.40 | 3.4.11.1 | Leucylaminopeptidase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506831.70 | 6.1.1.18 | Glutamine--tRNAligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053507883.100 | 1.1.1.37 | Malatedehydrogenase |
| Matching Annotation | Homologous | 4 | Tc00.1047053507883.80 | 2.7.4.3 | Adenylatekinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511299.70 | 2.7.11.24 | Mitogen-activatedproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510945.70 | 2.6.1.1 | Aspartatetransaminase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506853.50 | 1.2.4.1 | 3-methyl-2-oxobutanoatedehydrogenase(2-methylpropanoyl-transferring) |
| Matching Annotation | Homologous | 4 | Tc00.1047053506853.50 | 1.2.4.4 | 3-methyl-2-oxobutanoatedehydrogenase(2-methylpropanoyl-transferring) |
| Matching Annotation | Homologous | 4 | Tc00.1047053510437.40 | 1.1.1.37 | Malatedehydrogenase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503755.10 | 2.7.7.7 | DNA-directedDNApolymerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503755.30 | 2.7.7.7 | DNA-directedDNApolymerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511047.40 | 1.6.2.2 | Cytochrome-b5reductase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503841.70 | 2.6.1.1 | Aspartatetransaminase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 4 | Tc00.1047053510123.20 | 2.7.11.24 | Mitogen-activatedproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503687.40 | 5.2.1.8 | Peptidylprolylisomerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508351.10 | 3.1.3.11 | Fructose-bisphosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503851.24 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503581.20 | 5.2.1.8 | Peptidylprolylisomerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053506937.10 | 1.1.1.37 | Malatedehydrogenase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503471.10 | 3.1.3.48 | Protein-tyrosine-phosphatase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503513.10 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053504167.30 | 2.7.11.24 | Mitogen-activatedproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510221.39 | 3.5.4.6 | AMPdeaminase |
| Matching Annotation | Homologous | 4 | Tc00.1047053511691.10 | 4.2.1.22 | Cystathioninebeta-synthase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503815.20 | 6.1.1.6 | Lysine--tRNAligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503679.10 | 2.6.1.1 | Aspartatetransaminase |
| Matching Annotation | Homologous | 4 | Tc00.1047053507107.40 | 5.3.3.8 | Dodecenoyl-CoAisomerase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510671.20 | 6.1.1.2 | Tryptophan--tRNAligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053503753.29 | 1.6.2.4 | NADPH--hemoproteinreductase |
| Matching Annotation | Homologous | 4 | Tc00.1047053510777.20 | 6.1.1.12 | Aspartate--tRNAligase |
| Matching Annotation | Homologous | 4 | Tc00.1047053508129.9 | 3.5.4.6 | AMPdeaminase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 4 | Tc00.1047053508795.19 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 4 | Tc00.1047053461927.9 | 3.5.4.6 | AMPdeaminase |
| Matching Annotation | Homologous | 5 | Tc00.1047053508153.820 | 3.6.1.1 | Inorganicdiphosphatase |
| Matching Annotation | Homologous | 5 | Tc00.1047053506559.524 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 5 | Tc00.1047053506559.530 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 5 | Tc00.1047053511127.400 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 5 | Tc00.1047053506435.30 | 2.7.7.7 | DNA-directedDNApolymerase |
| Matching Annotation | Homologous | 5 | Tc00.1047053504105.30 | 2.7.7.6 | DNA-directedRNApolymerase |
| Matching Annotation | Homologous | 5 | Tc00.1047053511727.300 | 1.1.1.271 | GDP-L-fucosesynthase |
| Matching Annotation | Homologous | 5 | Tc00.1047053510431.140 | 2.7.12.1 | Dual-specificitykinase |
| Matching Annotation | Homologous | 5 | Tc00.1047053506195.90 | 2.7.4.3 | Adenylatekinase |
| Matching Annotation | Homologous | 5 | Tc00.1047053509167.120 | 4.1.1.50 | Adenosylmethioninedecarboxylase |
| Matching Annotation | Homologous | 5 | Tc00.1047053507723.189 | 5.3.4.1 | Proteindisulfide-isomerase |
| Matching Annotation | Homologous | 5 | Tc00.1047053511825.210 | 2.7.1.24 | Dephospho-CoAkinase |
| Matching Annotation | Homologous | 5 | Tc00.1047053506821.210 | 1.6.2.2 | Cytochrome-b5reductase |
| Matching Annotation | Homologous | 5 | Tc00.1047053509029.10 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 5 | Tc00.1047053503893.70 | 1.6.2.2 | Cytochrome-b5reductase |
| Matching Annotation | Homologous | 5 | Tc00.1047053511491.100 | 3.1.3.16 | Phosphoproteinphosphatase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 5 | Tc00.1047053509911.90 | 3.4.24.56 | Insulysin |
| Matching Annotation | Homologous | 5 | Tc00.1047053509901.170 | 4.2.99.18 | DNA-(apurinicorapyrimidinicsite)lyase |
| Matching Annotation | Homologous | 5 | Tc00.1047053507601.10 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 5 | Tc00.1047053508273.10 | 3.5.4.6 | AMPdeaminase |
| Matching Annotation | Homologous | 5 | Tc00.1047053506947.90 | 2.7.4.3 | Adenylatekinase |
| Matching Annotation | Homologous | 5 | Tc00.1047053506221.30 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 5 | Tc00.1047053511237.10 | 3.4.24.16 | Neurolysin |
| Matching Annotation | Homologous | 5 | Tc00.1047053510749.40 | 3.1.2.15 | Ubiquitinthiolesterase |
| Matching Annotation | Homologous | 5 | Tc00.1047053504013.110 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 5 | Tc00.1047053507883.90 | 2.7.4.3 | Adenylatekinase |
| Matching Annotation | Homologous | 5 | Tc00.1047053506569.10 | 3.1.2.15 | Ubiquitinthiolesterase |
| Matching Annotation | Homologous | 5 | Tc00.1047053509215.10 | 2.7.7.7 | DNA-directedDNApolymerase |
| Matching Annotation | Homologous | 5 | Tc00.1047053511481.50 | 4.2.99.18 | DNA-(apurinicorapyrimidinicsite)lyase |
| Matching Annotation | Homologous | 5 | Tc00.1047053507389.70 | 3.4.24.56 | Insulysin |
| Matching Annotation | Homologous | 5 | Tc00.1047053504443.10 | 3.1.2.15 | Ubiquitinthiolesterase |
| Matching Annotation | Homologous | 5 | Tc00.1047053511543.60 | 1.11.1.9 | Glutathioneperoxidase |
| Matching Annotation | Homologous | 5 | Tc00.1047053504005.10 | 4.2.99.18 | DNA-(apurinicorapyrimidinicsite)lyase |
| Matching Annotation | Homologous | 5 | Tc00.1047053506411.10 | 3.4.24.16 | Neurolysin |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 5 | Tc00.1047053510259.6 | 2.7.7.7 | DNA-directedDNApolymerase |
| Matching Annotation | Homologous | 5 | Tc00.1047053506701.10 | 1.8.1.4 | Dihydrolipoyldehydrogenase |
| Matching Annotation | Homologous | 5 | Tc00.1047053509379.10 | 1.8.1.4 | Dihydrolipoyldehydrogenase |
| Matching Annotation | Homologous | 5 | Tc00.1047053503873.10 | 1.6.2.2 | Cytochrome-b5reductase |
| Matching Annotation | Homologous | 5 | Tc00.1047053443397.9 | 1.6.2.2 | Cytochrome-b5reductase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506559.350 | 6.1.1.22 | Asparagine--tRNAligase |
| Matching Annotation | Homologous | 6 | Tc00.1047053508461.400 | 2.7.4.6 | Nucleoside-diphosphatekinase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506247.10 | 5.3.4.1 | Proteindisulfide-isomerase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506789.180 | 2.3.2.5 | Glutaminyl-peptidecyclotransferase |
| Matching Annotation | Homologous | 6 | Tc00.1047053511367.260 | 3.1.2.15 | Ubiquitinthiolesterase |
| Matching Annotation | Homologous | 6 | Tc00.1047053511727.40 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506945.220 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506859.10 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506265.30 | 2.7.7.7 | DNA-directedDNApolymerase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506739.190 | 2.7.12.1 | Dual-specificitykinase |
| Matching Annotation | Homologous | 6 | Tc00.1047053505807.10 | 5.2.1.8 | Peptidylprolylisomerase |
| Matching Annotation | Homologous | 6 | Tc00.1047053507083.30 | 4.2.99.18 | DNA-(apurinicorapyrimidinicsite)lyase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506871.100 | 3.5.1.88 | Peptidedeformylase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 6 | Tc00.1047053508569.130 | 2.7.1.105 | 6-phosphofructo-2-kinase |
| Matching Annotation | Homologous | 6 | Tc00.1047053508177.129 | 4.2.1.22 | Cystathioninebeta-synthase |
| Matching Annotation | Homologous | 6 | Tc00.1047053508699.10 | 5.99.1.3 | DNAtopoisomerase(ATP-hydrolyzing) |
| Matching Annotation | Homologous | 6 | Tc00.1047053510173.90 | 2.4.2.30 | NAD(+)ADP-ribosyltransferase |
| Matching Annotation | Homologous | 6 | Tc00.1047053503733.20 | 2.7.1.105 | 6-phosphofructo-2-kinase |
| Matching Annotation | Homologous | 6 | Tc00.1047053503723.90 | 1.2.1.12 | Glyceraldehyde-3-phosphatedehydrogenase(phosphorylating) |
| Matching Annotation | Homologous | 6 | Tc00.1047053509033.30 | 5.2.1.8 | Peptidylprolylisomerase |
| Matching Annotation | Homologous | 6 | Tc00.1047053509203.70 | 5.99.1.3 | DNAtopoisomerase(ATP-hydrolyzing) |
| Matching Annotation | Homologous | 6 | Tc00.1047053508865.4 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506315.90 | 2.7.12.1 | Dual-specificitykinase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506483.69 | 2.7.11.24 | Mitogen-activatedproteinkinase |
| Matching Annotation | Homologous | 6 | Tc00.1047053511499.50 | 3.1.2.15 | Ubiquitinthiolesterase |
| Matching Annotation | Homologous | 6 | Tc00.1047053506025.60 | 2.3.1.61 | Dihydrolipoyllysine-residuesuccinyltransferase |
| Matching Annotation | Homologous | 6 | Tc00.1047053509721.60 | 2.4.2.30 | NAD(+)ADP-ribosyltransferase |
| Matching Annotation | Homologous | 6 | Tc00.1047053509505.10 | 5.3.4.1 | Proteindisulfide-isomerase |
| Matching Annotation | Homologous | 6 | Tc00.1047053507897.30 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 6 | Tc00.1047053511735.60 | 1.15.1.1 | Superoxidedismutase |
| Matching Annotation | Homologous | 6 | Tc00.1047053509261.30 | 5.3.3.8 | Dodecenoyl-CoAisomerase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 6 | Tc00.1047053511521.30 | 1.15.1.1 | Superoxidedismutase |
| Matching Annotation | Homologous | 6 | Tc00.1047053507693.5 | 6.1.1.2 | Tryptophan--tRNAligase |
| Matching Annotation | Homologous | 6 | Tc00.1047053503935.20 | 2.3.1.61 | Dihydrolipoyllysine-residuesuccinyltransferase |
| Matching Annotation | Homologous | 6 | Tc00.1047053511813.20 | 6.3.2.19 | Ubiquitin--proteinligase |
| Matching Annotation | Homologous | 6 | Tc00.1047053508585.29 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 6 | Tc00.1047053508537.10 | 1.2.1.12 | Glyceraldehyde-3-phosphatedehydrogenase(phosphorylating) |
| Matching Annotation | Homologous | 7 | Tc00.1047053511277.20 | 2.7.7.6 | DNA-directedRNApolymerase |
| Matching Annotation | Homologous | 7 | Tc00.1047053506885.390 | 3.5.4.6 | AMPdeaminase |
| Matching Annotation | Homologous | 7 | Tc00.1047053508707.180 | 3.1.2.15 | Ubiquitinthiolesterase |
| Matching Annotation | Homologous | 7 | Tc00.1047053507063.100 | 2.7.7.7 | DNA-directedDNApolymerase |
| Matching Annotation | Homologous | 7 | Tc00.1047053511367.70 | 2.3.1.12 | Dihydrolipoyllysine-residueacetyltransferase |
| Matching Annotation | Homologous | 7 | Tc00.1047053506357.90 | 2.7.7.6 | DNA-directedRNApolymerase |
| Matching Annotation | Homologous | 7 | Tc00.1047053510729.299 | 2.7.11.22 | Cyclin-dependentkinase |
| Matching Annotation | Homologous | 7 | Tc00.1047053508387.120 | 2.3.2.5 | Glutaminyl-peptidecyclotransferase |
| Matching Annotation | Homologous | 7 | Tc00.1047053511825.220 | 6.1.1.7 | Alanine--tRNAligase |
| Matching Annotation | Homologous | 7 | Tc00.1047053510879.10 | 3.1.3.16 | Phosphoproteinphosphatase |
| Matching Annotation | Homologous | 7 | Tc00.1047053506733.20 | 3.5.4.6 | AMPdeaminase |
| Matching Annotation | Homologous | 7 | Tc00.1047053506375.10 | 6.1.1.2 | Tryptophan--tRNAligase |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Homologous | 7 | Tc00.1047053510241.120 | 3.5.1.88 | Peptidedeformylase |
| Matching Annotation | Homologous | 7 | Tc00.1047053509073.30 | 3.1.2.15 | Ubiquitinthiolesterase |
| Matching Annotation | Homologous | 7 | Tc00.1047053508995.40 | 3.6.4.6 | Vesicle-fusingATPase |
| Matching Annotation | Homologous | 7 | Tc00.1047053510579.20 | 3.5.1.88 | Peptidedeformylase |
| Matching Annotation | Homologous | 7 | Tc00.1047053506627.10 | 3.5.4.6 | AMPdeaminase |
| Matching Annotation | Homologous | 7 | Tc00.1047053506863.10 | 6.1.1.7 | Alanine--tRNAligase |
| Matching Annotation | Homologous | 7 | Tc00.1047053506477.20 | 3.6.4.6 | Vesicle-fusingATPase |
| Matching Annotation | Homologous | 7 | Tc00.1047053506419.20 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 7 | Tc00.1047053511235.20 | 1.2.1.12 | Glyceraldehyde-3-phosphatedehydrogenase(phosphorylating) |
| Matching Annotation | Homologous | 7 | Tc00.1047053506851.10 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 7 | Tc00.1047053510597.9 | 2.7.11.1 | Non-specificserine/threonineproteinkinase |
| Matching Annotation | Homologous | 7 | Tc00.1047053510905.19 | 6.1.1.2 | Tryptophan--tRNAligase |
| Matching Annotation | Homologous | 7 | Tc00.1047053510443.9 | 6.1.1.9 | Valine--tRNAligase |
| Matching Annotation | Analogous | 4 | Tc00.1047053509941.100 | 1.3.1.34 | 2,4-dienoyl-CoAreductase(NADPH) |
| Matching Annotation | Analogous | 5 | Tc00.1047053510303.210 | 1.3.1.34 | 2,4-dienoyl-CoAreductase(NADPH) |
| Matching Annotation | Specific of T. cruzi | 1 | Tc00.1047053503555.30 | 1.8.1.12 | Trypanothione-disulfidereductase |
| Matching Annotation | Specific of T. cruzi | 3 | Tc00.1047053508595.50 | 3.4.22.51 | Cruzipain |
| Matching Annotation | Specific of T. cruzi | 3 | Tc00.1047053507297.10 | 3.4.22.51 | Cruzipain |

| | | | | | |
|---|---|---|---|---|---|
| Matching Annotation | Specific of T. cruzi | 3 | Tc00.1047053504507.5 | 1.8.1.12 | Trypanothione-disulfidereductase |
| Matching Annotation | Specific of T. cruzi | 4 | Tc00.1047053507165.50 | 2.5.1.47 | Cysteinesynthase |
| Matching Annotation | Specific of T. cruzi | 4 | Tc00.1047053507793.20 | 2.5.1.47 | Cysteinesynthase |
| Matching Annotation | Specific of T. cruzi | 5 | Tc00.1047053506649.20 | 3.6.3.6 | Proton-exportingATPase |
| Matching Annotation | Specific of T. cruzi | 6 | Tc00.1047053506529.550 | 3.4.22.51 | Cruzipain |
| Matching Annotation | Specific of T. cruzi | 6 | Tc00.1047053507537.20 | 3.4.22.51 | Cruzipain |
| Matching Annotation | Specific of T. cruzi | 6 | Tc00.1047053505763.19 | 3.6.3.6 | Proton-exportingATPase |
| Matching Annotation | Specific of T. cruzi | 7 | Tc00.1047053509429.320 | 3.4.22.51 | Cruzipain |
| Matching Annotation | Specific of T. cruzi | 7 | Tc00.1047053507603.260 | 3.4.22.51 | Cruzipain |
| Matching Annotation | Specific of T. cruzi | 7 | Tc00.1047053507603.270 | 3.4.22.51 | Cruzipain |
| Matching Annotation | Specific of T. cruzi | 7 | Tc00.1047053509401.30 | 3.4.22.51 | Cruzipain |
| Matching Annotation | Conflicting Clustering (a) (Homologous/Analogous) | 7 | Tc00.1047053511895.10 | 4.6.1.1 | Adenylatecyclase |
| Matching Annotation | Conflicting Clustering (a) (Homologous/Analogous) | 7 | Tc00.1047053510581.20 | 4.6.1.1 | Adenylatecyclase |
| Matching Annotation | Conflicting Clustering (a) (Homologous/Analogous) | 7 | Tc00.1047053509449.10 | 4.6.1.1 | Adenylatecyclase |
| Surface | Homologous | 4 | Tc00.1047053510503.100 | 3.4.24.36 | Leishmanolysin |
| Surface | Homologous | 4 | Tc00.1047053509369.50 | 3.4.24.36 | Leishmanolysin |
| Surface | Homologous | 4 | Tc00.1047053410797.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Homologous | 6 | Tc00.1047053508165.310 | 3.4.24.36 | Leishmanolysin |
| Surface | Homologous | 6 | Tc00.1047053506289.140 | 3.4.24.36 | Leishmanolysin |
| Surface | Homologous | 6 | Tc00.1047053506289.170 | 3.4.24.36 | Leishmanolysin |
| Surface | Homologous | 6 | Tc00.1047053506289.210 | 3.4.24.36 | Leishmanolysin |
| Surface | Homologous | 6 | Tc00.1047053511257.100 | 3.4.24.36 | Leishmanolysin |

| | | | | | |
|---|---|---|---|---|---|
| Surface | Homologous | 6 | Tc00.1047053511151.54 | 3.4.24.36 | Leishmanolysin |
| Surface | Homologous | 6 | Tc00.1047053506515.29 | 3.2.1.18 | Exo-alpha-sialidase |
| Surface | Homologous | 6 | Tc00.1047053509483.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053511211.90 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053510565.150 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053507623.110 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053508699.100 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053508699.90 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053509011.80 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053506587.100 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053509205.100 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053506163.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053506163.20 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053508813.40 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053505965.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053506257.50 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053510899.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053505931.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053505931.20 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053511203.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053504397.20 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053506921.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053508475.30 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053505615.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053508825.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 4 | Tc00.1047053510873.20 | 3.4.24.36 | Leishmanolysin |
| Surface | Specific of T. cruzi | 6 | Tc00.1047053507197.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053511277.610 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053506401.380 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053506435.370 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053507993.350 | 3.4.24.36 | Leishmanolysin |

| | | | | | |
|---|---|---|---|---|---|
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053508999.170 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053506321.240 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053508693.100 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053510657.200 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053510747.40 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053511257.60 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053510761.80 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053511281.50 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053505989.70 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053508545.40 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053506867.40 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053510263.30 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053508609.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053508611.30 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053511723.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053507919.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053511035.10 | 3.4.24.36 | Leishmanolysin |
| Surface | Conflicting Clustering (a) (Homologous/Specific of T. cruzi) | 4 | Tc00.1047053507917.10 | 3.4.24.36 | Leishmanolysin |

| | | | | | |
|---|---|---|---|---|---|
| Hypothetical | Homologous | 5 | Tc00.1047053507057.20 | 2.7.4.3 | Adenylatekinase |
| Hypothetical | Homologous | 6 | Tc00.1047053510303.230 | 3.1.1.29 | Aminoacyl-tRNAhydrolase |
| Hypothetical | Homologous | 6 | Tc00.1047053511763.19 | 2.3.1.7 | CarnitineO-acetyltransferase |
| Hypothetical | Homologous | 7 | Tc00.1047053509207.130 | 5.2.1.8 | Peptidylprolylisomerase |
| Hypothetical | Homologous | 7 | Tc00.1047053506857.60 | 5.2.1.8 | Peptidylprolylisomerase |
| Hypothetical | Analogous | 6 | Tc00.1047053509005.50 | 3.1.1.3 | Triacylglycerollipase |
| Conflicting Annotation (b) | Homologous | 5 | Tc00.1047053506959.70 | 2.3.1.29 | GlycineC-acetyltransferase |
| Conflicting Annotation (b) | Homologous | 5 | Tc00.1047053503453.100 | 2.3.1.29 | GlycineC-acetyltransferase |

(a) Conflicting clustering between results obtained by KEGG and Swiss-Prot databases using AnEnPi methodology.
(b) Conflicting annotation between the methodology proposed in this work and GeneDB