

Sistemas de visualização de redes metabólicas

Gabriella de O. Esteves
Universidade de Brasília
Departamento de Ciência da Computação
Brasília, Brasil
Email: gabepk.ape@gmail.com

Abstract—The abstract goes here.

I. INTRODUÇÃO

Metabolismo é o conjunto de todas as reações bioquímicas que ocorrem dentro de um organismo vivo. Ele pode ser representado computacionalmente em forma de vias metabólicas, sequências de reações bioquímicas cujo produto e substrato são denominados metabólitos e que podem ser catalisadas por enzimas com ou sem ajuda de co-fatores. A estrutura de dados que melhor representa o metabolismo é o grafo, onde os nós são os elementos químicos ou reações e as arestas são as relações entre eles.

Atualmente os pesquisadores biólogos trabalham com uma quantidade tão extensa de dados sobre redes metabólicas (conjunto de vias metabólicas) que apenas ferramentas de Big Data podem ser usadas para análise, visualização, busca, etc, para um tratamento eficiente. Estes dados, denominados dados ômicos, são armazenados em bancos de dados específicos hoje em dia, mais vários deles colaboram entre si. O foco deste trabalho é apresentar os bancos de dados que representam redes metabólicas em modelo de grafo já existentes bem como suas ferramentas de visualização de vias metabólicas, e apresentar a ferramenta de visualização 2Path formularizada e desenvolvida neste projeto.

II. REDES METABÓLICAS

As reações bioquímicas são alterações químicas que fornecem um ou mais produtos a partir de um ou mais substratos. Uma via metabólica é uma sequência de reações bioquímicas, cujo produto e substrato são denominados de metabólitos, que podem ser catalisados por enzimas, as quais muitas vezes necessitam de compostos químicos chamados de co-fatores para realizarem suas atividades na célula. O conjunto de vias metabólicas de um organismo é chamado de rede metabólica [8]. Todos estes elementos que compõem as redes metabólicas são dados biológicos estudados na área metabômica. Nesta seção são apresentados conceitos básicos de biologia molecular e metabolismo, e três bancos de dados voltados para redes metabólicas: KEGG¹, BioCyc² e Reactome³.

¹<http://www.kegg.jp>.

²<http://biocyc.org>.

³<http://www.reactome.org>.

A. Conceitos de Biologia Molecular

O DNA é um conjunto de biomoléculas em um organismo que armazena informações, chamados de genes, referentes ao funcionamento de todas as suas células. Ele constitui o genoma em todos os seres vivos, com exceção dos vírus. A expressão dos genes é o processo no qual os genes são filtrados e utilizados na síntese de um produto, geralmente proteína. O método é segmentado em três etapas: transcrição (síntese de RNA mensageiro, mRNA, a partir de DNA), *splicing* (filtragem do gene sintetizador de proteínas desejadas do mRNA) e tradução (síntese de proteína a partir do mRNA filtrado). Completo este processo, as proteínas resultantes poderão formar uma configuração tridimensional de até quatro níveis. As enzimas, por exemplo, são proteínas que podem ter estrutura terciária ou quaternária.

B. Conceitos de Metabolismo

Enzimas são proteínas responsáveis por auxiliar a realização de biossíntese (construção) e biodegradação de moléculas no metabolismo com o propósito de catalisar (acelerar) reações bioquímicas. As enzimas possuem um local pré-determinado em formato côncavo chamado de sítio ativo, que comporta um ou mais substratos. Se a enzima comporta apenas um substrato, a estrutura que se forma com o preenchimento do sítio ativo é um complexo enzima-substrato, porém se ela comporta mais de um substrato, a estrutura é chamada de complexo ternário intermediário [9]. Quando a atividade catalítica não pode ser realizada apenas pela enzima, co-fatores auxiliam o processo.

O conjunto de todas as reações bioquímicas que ocorrem dentro de um organismo vivo é chamado de metabolismo, e ele podem ser dividido em dois subconjuntos: catabolismo, quando ocorre a quebra de moléculas complexas produzindo energia, e anabolismo, quando ocorre a síntese de moléculas complexas, o que requer energia. Ele também pode ser classificado em relação à função que exerce no organismo [10]. Se a função realizada é fundamental no organismo, como crescimento, desenvolvimento e reprodução, ele é denominado metabolismo primário. Caso contrário, é denominado metabolismo secundário.

C. Banco de Dados de redes metabólicas

1) **KEGG**: O KEGG (*Kyoto Encyclopedia of Genes and Genomes*) é uma base de informações sobre sistemas

biológicos em nível molecular, sobretudo sobre conjuntos de dados em larga escala gerados por sequenciamento de genoma [2]. As informações sobre os sistemas podem ser dadas em forma de módulos, unidades funcionais com identificação otimizada para análise dos dados, em forma de *brite*, coleção de arquivos estruturados hierarquicamente sobre as funções das entidades biológicas, ou em forma de vias, mapa de interações moleculares e reações químicas. Dado que o metabolismo é um conjunto de reações e transformações químicas, a maneira natural de representá-lo é por meio de uma rede de interações, ou seja, em forma de vias. O KEGG oferece uma ferramenta de busca de vias metabólicas sobre várias rede metabólica, dos vários organismos que constituem o banco de dados.

2) *BioCyc*: O BioCyc é um sistema de coleção de aproximadamente 7 mil bancos de dados chamados PGDBs (*Pathway/Genome Databases*) pois possuem duas maneiras diferentes de representar as informações: modelo de vias metabólicas, que enfatiza as sequências de reações, substratos e produtos de múltiplos organismos, ou modelo de sequência genômica, que destaca a localização e descrição dos genes de cada organismo específico [3]. Os bancos PGDBs são organizado em três camadas de acordo com a frequência de atualizações/refinações e da maneira com que os dados foram obtidos. O BioCyc possui um banco de dados específico para redes metabólicas determinadas experimentalmente, chamado MetaCyc. Este é o único banco de dados multi-organismos do grupo BioCyc e ele é referência na ferramenta gratuita *Pathway Tools* desenvolvida pelo instituto de pesquisa *SRI International*.

3) *Reactome*: O Reactome é um banco de dados de reações de mudança de estado, ou seja, além de reações bioquímica, ele também abrange reações de ativação, de degradação e de ligação, por exemplo [4]. Ele faz uma ligação sistemática entre as proteínas de um certo organismo e as funções moleculares do mesmo, fornecendo uma base de funções que pode ser utilizada para pesquisas sobre expressão de genes ou mutações somáticas. O Reactome disponibiliza o *Pathway Browser*, uma rede geral para cada organismo, que representa os vários seus sistemas, como reprodução e metabolismo, por exemplo. Algumas sub-redes estão conectadas (por exemplo, replicação de DNA e ciclo de célula), outra não (por exemplo, contração muscular e reprodução). Nesta rede, cada nó representa uma via cujo número de entidade se reflete no raio do nó, e cada aresta representa a relação entre estas vias. O site ainda possui uma ferramenta de análise de dados baseada nas correspondências entre as reações na redes dos organismos comparados.

III. FERRAMENTAS DE VISUALIZAÇÃO DE REDES METABÓLICAS

A. KEGG Pathway Maps

O KEGG oferece uma visao geral e uma visualização específica para análise de redes metabólicas. Na primeira, o grafo interativo apresenta uma perspectiva global das entidades do banco de dados do KEGG, chamadas de objetos KEGG.

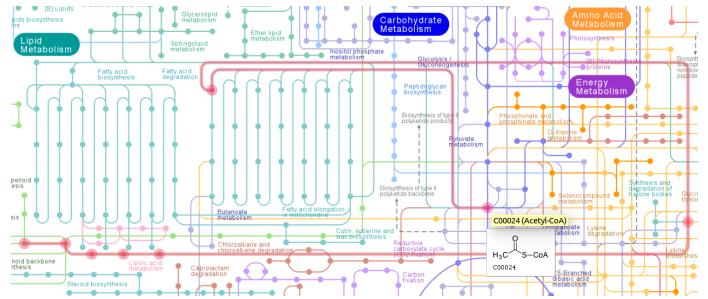


Figure 1. Visão geral da rede metabólica de referência do KEGG, com destaque na via metabólica de biosíntese do backbone de terpenóides [11].

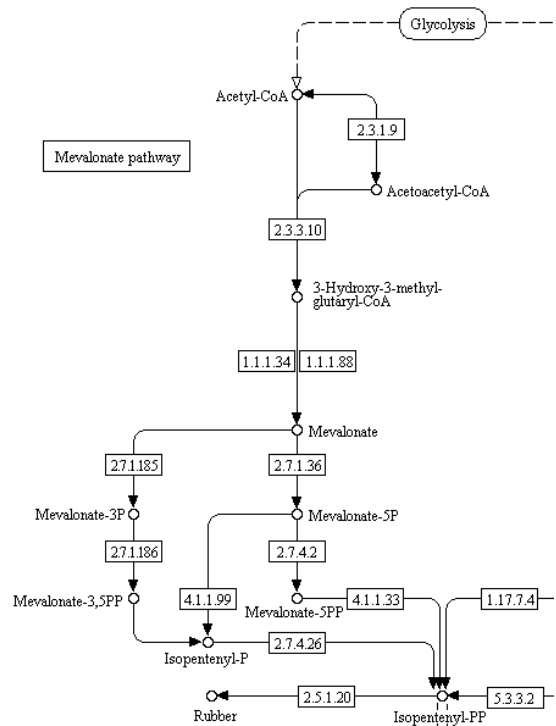


Figure 2. Via metabólica de biosíntese do backbone de terpenóides [12].

Neste panorama, a rede metabólica selecionada é destacada e é acompanhada de interações externas, que não possuem ligação direta com a rede analisada. Os nós representam compostos químicos e as arestas podem representar enzimas, reações e/ou ortólogos. Na página do Atlas, a visualização das arestas pode ser filtrada, bem como as vias, de acordo com suas funções. A Figura 1 apresenta uma parte da via de biosíntese do backbone de terpenóides, que tem início no composto Acetyl-CoA, destacada sobre as demais. A outra maneira de visualizar esta via é acessando o objeto KEGG do tipo *map* o qual ela representa (*KEGG Pathway Map*). O mapa de vias é um diagrama de interações/reações moleculares entre enzimas desenhado manualmente. A via da Figura 2 apresenta a mesma via destacada da Figura 1.

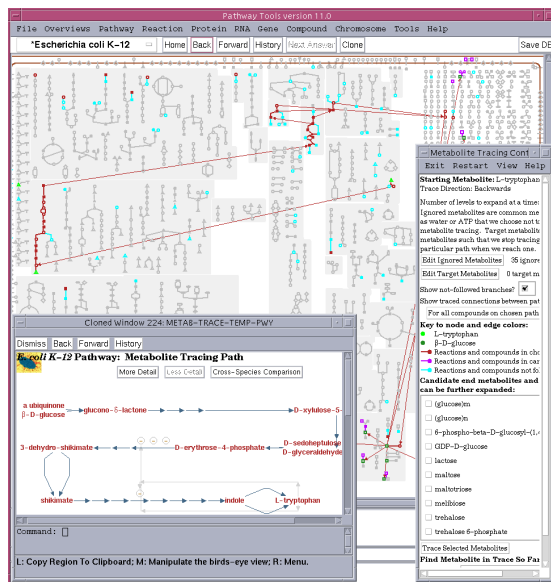


Figure 3. BioCyc

B. MetaCyc Pathway Tools

O Instituto SRI Internacional oferece a ferramenta *Pathway Tools* para busca, visualização e edição de dados do MetaCyc [5]. Ela possui três componentes de análise:

- 1 **Pathologic**: Um PGDB é gerado a partir do genoma de um organismo e de vias metabólicas produzidas por um software de predição de vias;
- 2 **Pathway/Genome Editor**: Permite que curadores (especialistas responsáveis por verificar na literatura a correção dos dados) possam refinar manualmente um PGDB, bem como importar ou exportar entidades de um PGDB para outro;
- 3 **Pathway/Genome Navigator**: Oferece mecanismos de busca, visualização e análise de redes metabólicas.

Além dessas funcionalidades principais, o Pathway Tools dispõe também de meios para geração de posters de genomas e mapas metabólicos a partir de PGDB, comparação de mapas metabólicos inteiros e exportação de PGDB em vários formatos [6]. Via web, o MetaCyc oferece mecanismos básicos de análise metabolômica, tais como coloração de nós sobre redes (ou colagens de redes) metabólicas para geração de diagramas customizados e análise de níveis de metabólitos, por exemplo [7].

C. Reactome

- <http://wiki.reactome.org/index.php/Usersguide>
- The Reactome pathway knowledgebase

D. Cytoscape

- http://www.cytoscape.org/what_is_cytoscape.html
- Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks

IV. CONCLUSÃO

Conclusão

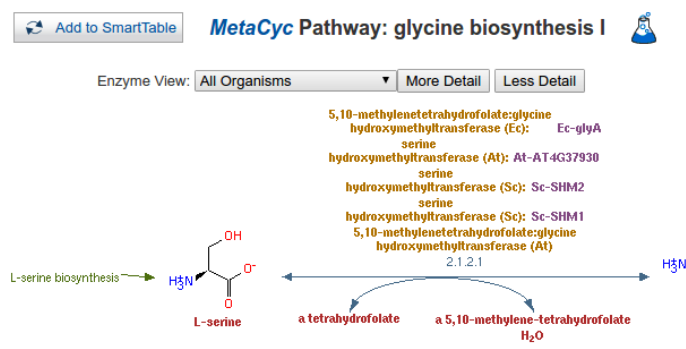


Figure 4. BioCyc

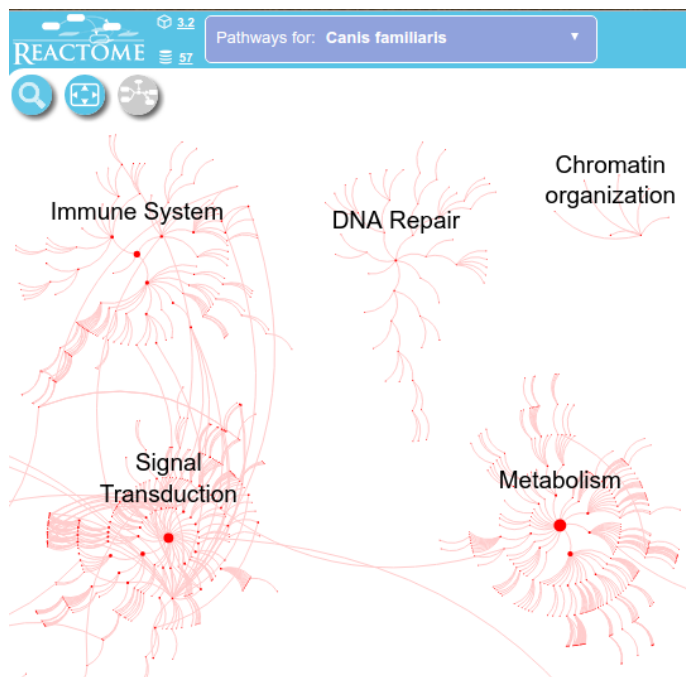


Figure 5. Reactome

REFERENCES

- [1] W.M.C. Silva, *Método para reconstrução in silico de redes metabólicas de fungos: um estudo de caso para o Paracoccidioides lutzii*. Universidade de Brasília, 2014.
- [2] Kyoto Encyclopedia of Genes and Genome, *KEGG Overview* <http://www.kegg.jp/kegg/kegg1a.html> visitado em 01 de Outubro de 2016.
- [3] SRI International and BioCyc Database Collection, *Introduction to BioCyc* <http://biocyc.org/intro.shtml> visitado em 01 de Outubro de 2016.
- [4] Reactome, a curated pathway database, *Usersguide* <http://wiki.reactome.org/index.php/Usersguide> visitado em 01 de Outubro de 2016.
- [5] Caspi, Ron et al. *MetaCyc: A Multiorganism Database of Metabolic Pathways and Enzymes*. Nucleic Acids Research 34.Database issue (2006): D511–D516. PMC. Web. 5 Oct. 2016.
- [6] SRI International and BioCyc Database Collection, *Comparison of Desktop and Web Modes of Pathway Tools* <http://biocyc.org/desktop-vs-web-mode.shtml> visitado em 03 de Outubro de 2016.
- [7] SRI International and BioCyc Database Collection, *Omics Data Analysis* <http://metacyc.org/PToolsWebsiteHowto.shtml#omicsDataAnalysis> visitado em 04 de Outubro de 2016.

- [8] D.L. Nelson and M.C. Michael, *Princípios de Bioquímica de Lehninger - Ed. Comemorativa 25 Anos* Artmed, São Paulo - SP, 5nd edition, 2010.
- [9] G. Michal and D. Schomburg, *The Cell and Its Contents*, pages 14–36. John Wiley & Sons, Inc., 2012.
- [10] R. Wüschiers, M. Jahn, D. Jahn, I. Schomburg, S. Peifer, E. Heinze, H. Bartscher, J. Garbe, A. Steen, M. Schobert, D. Oesterheld, J. Wachtveitl, and A. Chang, *Metabolism*, pages 37–209. John Wiley & Sons, Inc., 2012.
- [11] Kyoto Encyclopedia of Genes and Genome, *Metabolic pathways - Reference pathway* http://www.kegg.jp/kegg-bin/show_pathway?map01100 visitado em 03 de Outubro de 2016.
- [12] Kyoto Encyclopedia of Genes and Genome, *Terpenoid backbone biosynthesis - Reference pathway* http://www.kegg.jp/kegg-bin/show_pathway?map=map00900&show_description=show visitado em 03 de Outubro de 2016.