

Visualização em Grafo de Redes Metabólicas via Web

Gabriella de Oliveira Esteves

Setembro de 2016

Sumário

- 1 Fundamentação Teórica
 - Biologia molecular
 - Bioinformática
 - Grafos
 - Banco de dados NoSQL
- 2 Problema
- 3 Objetivos
- 4 2Path
 - Design
 - Implementação
- 5 Método
- 6 Referências

Síntese de RNA e síntese de proteína

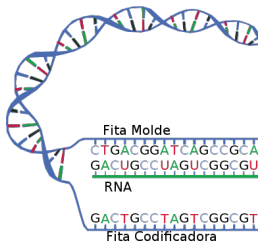


Figure 1: A transcrição ocorre no núcleo celular

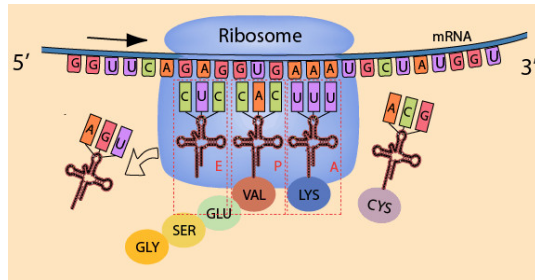


Figure 2: A tradução ocorre no retículo endoplasmático rugoso

Enzima

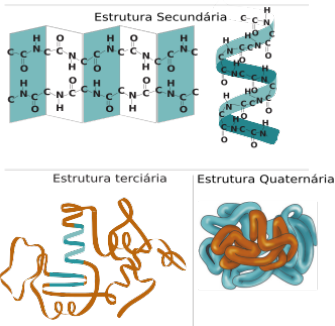


Figure 3: Estruturas das proteínas

- **Enzimas** são proteínas que *catalizam* reações químicas
- **Catabolismo** é a ação de degradar moléculas para produção de energia
- **Isoenzimas** são *enzimas* com as mesmas funções, mas com estruturas diferentes
- **Cofatores** são pequenas moléculas que se ligam às *enzimas* e ou realçam ou limitam sua atividade

Metabolismo

Metabolismo

Metabolismo é o *conjunto* de todas as reações bioquímicas de síntese e degradação de substâncias em certo organismo.

Uma **reação bioquímica** possui uma ou mais entradas (**substratos**) e uma ou mais saídas (**produtos**). Um produto pode ser substrato.

Via metabólica

Sequência de reações bioquímicas que ocorrem num organismo que podem ser catalisadas por enzimas com ajuda de cofatores

Rede metabólica

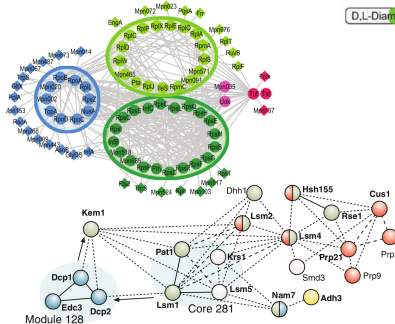
Conjunto de Vias metabólicas

Desafios das ômicas

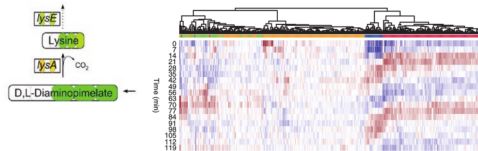
		Desafio na Bioinformática
DNA	Genoma	Sequenciamento de Gene
RNA	Transcriptoma	Expressão do mRNA Sequenciamento de RNA
Proteína	Proteoma	Expressão da proteína Estrutura da proteína
	Metaboloma	Concentração de metabólitos Pathways metabólicos
	Interactoma	Interações entre proteínas

Visualização de dados ômicos

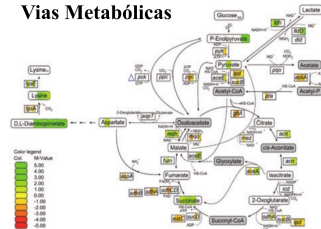
Interação entre proteínas



Expressão dos genes



Vias Metabólicas



Grafos

Grafo é um par ordenado $G = (V_G, E_G)$, onde $V_G = \{v_1, v_2, \dots\}$ é um conjunto de vértices e $E_G = \{e_1, e_2, \dots\}$ é um conjunto de arestas.

Grafo **orientado**: $e_i = (v_j, v_k)$

Grafo **não-orientado**: $e_i = \{v_j, v_k\}$

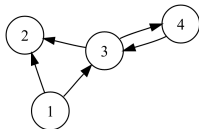


Figure 6: Grafo orientado

Passeio: Sequência de vértices distintos (v_1, v_2, \dots, v_k) tal que (v_i, v_{i+1}) é aresta para $1 \leq i \leq k$

Trilha: Passeio tal que $e_i \neq e_j \forall i, j$

Caminho: Trilha tal que $v_i \neq v_j \forall i, j$

Circuito: Trilha fechada $(v_1 = v_k)$

Ciclo: Caminho fechado $(v_1 = v_k)$

Banco de dados NoSQL orientado a grafo

Banco de dados NoSQL (Não-Relacional ou Não-só-SQL)

- Fácil escalabilidade e distribuição (replicação ou fragmentação) de dados entre vários servidores
- Uso eficiente de indexação distribuída e de memória RAM
- (*CAP Theorem*, 2000) Somente duas destas propriedades: consistência forte, disponibilidade alta ou tolerância a particionamento dos dados na rede

Banco de dados orientado a grafo

- Armazenamento distribuído eficiente
- Relacionamentos (arestas) dispensam uso de JOIN

Análise de redes metabólicas

Problema

Prover uma visualização interativa de redes metabólicas armazenadas em bancos de dados orientados a grafos.

Objetivos

Objetivo

Construir um sistema que acesse redes metabólicas armazenadas em bancos de dados em grafo e gere uma visualização interativa.

Objetivos específicos

- Implementar buscas de vias metabólicas a partir de parâmetros informados pelo pesquisador no sistema
- Recuperar a informação desejada e exibí-la de forma interativa e ergonômica
- Implementar algoritmos de busca em grafos para recuperar eficientemente a informação solicitada e/ou sugerir informação relevante

Login



Sing In

SING IN

[Forgot Password?](#)

New User

SING UP



Busca de chave em organismo



Select one Organism:

Organisms *

Organism A

Hyl

Hydro

Search

New Organism

New Protein

New Reaction

New Pathway

Reports

Update

Help

Figure 8: Página de seleção de Organismo e *keyword* do banco de dados do sistema

Busca específica



Select one of the results:

Results *

Enzyme: hydroxymethylglutaryl-CoA synthase

Reaction: (S)-3-Hydroxy-3-methylglutaryl-CoA + CoA <=> Acetyl-CoA + H2O + Acetoacetyl-CoA

See Details

Voltar

New Organism

New Protein

New Reaction

New Pathway

Reports

Update

Help

Figure 9: Página de seleção de resultado da busca anterior

Java

```
▼ 📁 > EnzymeGraph [EnzymeGraph master]
  ▼ 📁 src/main/java
    ▼ 📁 com.system.model
      ▶ 📄 CellularLocus.java
      ▶ 📄 Cofactor.java
      ▶ 📄 Compound.java
      ▶ 📄 Enzyme.java
      ▶ 📄 Feature.java
      ▶ 📄 Organism.java
      ▶ 📄 Reaction.java
```

```
<modelVersion>4.0.0</modelVersion>
<groupId>com.graph</groupId>
<artifactId>EnzymeGraph</artifactId>
<version>1.0.0-SNAPSHOT</version>
<packaging>war</packaging>

<properties>
  <project.build.sourceEncoding>UTF-8</project.build.sourceEncoding>
  <orientdb.version>2.2.7</orientdb.version>
</properties>

<licenses>
  <license>
    <name>The Apache Software License, Version 2.0</name>
    <url>http://www.apache.org/licenses/LICENSE-2.0.txt</url>
    <distribution>repo</distribution>
  </license>
</licenses>
```



OrientDB

conexão com banco

```
public List<Organism> getAllOrganisms() {  
    List<Organism> allOrganisms = new ArrayList<>();  
  
    OrientGraph graph = new OrientGraph("plocal:DBs/Teste");  
    try {  
        for (Vertex v : graph.getVerticesOfClass("Organism")) {  
            Organism organism = new Organism((String) v.getProperty("taxName"));  
            organism.setId((Long) v.getProperty("id"));  
            organism.setTaxName((String) v.getProperty("taxName"));  
            allOrganisms.add(organism);  
        }  
    } finally {  
        graph.shutdown();  
    }  
  
    return allOrganisms;  
}
```



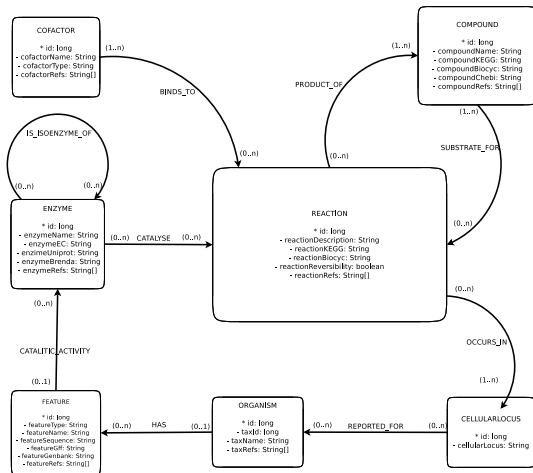


Figure 10: Modelo de dados proposto para o 2Path (Waldeyr e Maristela)

AngularJS

Aprendendo:

- **Learn AngularJS 1.X**, Code Academy
<https://www.codecademy.com/learn/learn-angularjs>
- **D3 for Mere Mortals**, por Luke Franchi
<http://www.recursion.org/d3-for-mere-mortals/>
- **D3 Tutorials**, por Scott Murray
<http://alignedleft.com/tutorials/d3>



Cronograma

Table 1: Cronograma

Atividades	2016			
	Set	Out	Nov	Dez
Familiaridade com as ferramentas e linguagens	X			
Implementação da aplicação	X	X		
Implementação de algoritmos de busca em grafo		X	X	
Interpretação dos resultados			X	X
Defesa				X

Referências

- 1 Vincent Lacroix, Ludovic Cottret, Patricia Thébault, and Marie-France Sagot. An introduction to metabolic networks and their structural analysis. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 5(4):594–617, 2008
- 2 João Carlos Setubal and João Meidanis. Introduction to computational molecular biology. PWS Publishing Company, 1997
- 3 Jing Han, E. Haihong, Guan Le, and Jian Du. Survey on NoSQL database. In *Pervasive Computing and Applications (ICPCA)*, 2011 6th International Conference on, pages 363–366. IEEE, October 2011.

Figuras

- 1 History of biology, visitado em 2016-09-23,
https://en.wikipedia.org/wiki/History_of_biology
- 2 Transcription (Gene Expression), visitado em 2016-01-02,
https://geneed.nlm.nih.gov/topic_subtopic.php?tid=15&sid=22
- 3 Protein structure, 2009, visitado em 2016-09-22,
<http://www.particlesciences.com/news/technical-briefs/2009/protein-structure.html>
- 4 Prof. Doutor Pedro Silva. Uma panorâmica geral das vias metabólicas, 2002, visitado em 2016-09-22,
<http://homepage.ufp.pt/pedros/bq/integracao.htm>

Perguntas e Respostas

Perguntas?