



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE COMPUTAÇÃO

RONALD ASSUNÇÃO SOUZA

**sARCASM - UMA FERRAMENTA SEMI-AUTOMÁTICA
PARA RECONSTRUÇÃO E ANÁLISE DE MATRIZES
ESTEQUIOMÉTRICAS**

Belém – PA
2018

Ronald Assunção Souza

**sARCASM - UMA FERRAMENTA SEMI-AUTOMÁTICA
PARA RECONSTRUÇÃO E ANÁLISE DE MATRIZES
ESTEQUIOMÉTRICAS**

Trabalho de Conclusão de Curso apresentado para obtenção do grau de Bacharel em Ciência da Computação. Instituto de Ciências Exatas e Naturais. Universidade Federal do Pará.

Orientador: Prof. Dra. Regiane Kawasaki

Co-orientador: Msc. Renato R. M. Oliveira

Belém - PA

2018

Ronald Assunção Souza

**sARCASM - UMA FERRAMENTA SEMI-AUTOMÁTICA
PARA RECONSTRUÇÃO E ANÁLISE DE MATRIZES
ESTEQUIOMÉTRICAS**

Trabalho de Conclusão de Curso
apresentado para obtenção do grau de
Bacharel em Ciência da Computação.
Instituto de Ciências Exatas e Naturais.
Universidade Federal do Pará.

Data da defesa: Belém-PA, 22 do Mês de Fevereiro

Conceito: _____

Prof. Dra. Regiane Kawasaki – Orientador
Faculdade de Computação – ICEN/UFPA

Msc. Renato R. M. Oliveira – Coorientador
Faculdade de Computação – ICEN/UFPA

Prof. Dra. – Danielle Costa Carrara Couto
Faculdade de Tecnologia em Geoprocessamento – Campos Ananindeua/UFPA

Prof. Dra. – Luciana Pereira Xavier
Faculdade de Biotecnologia – ICB/UFPA

Dedico este trabalho em memória de meus avós José Souza e Raimunda Souza e em homenagem a meus avós Martiniano Patrício e Maria Bárbara.

AGRADECIMENTOS

À minha orientadora, Dra. Regiane Kawasaki, por ter me dado oportunidade de fazer parte de sua equipe no Laboratório de Bioinformática e computação de Alto desempenho (LabioCad), por ter acreditado no meu potencial, e por sempre confiar no meu trabalho e sempre me incentivar;

Ao meu co-orientador Msc. Renato Oliveira por ter paciência comigo, sempre disposto a me ajudar, por ter me ensinado muitas coisas e por ter me apresentado o tema, com certeza um grande amigo;

Aos meus avós Martiniano e Bárbara que tornaram tudo isso possível, me deram toda a assistência possível. Sempre se preocuparam com a educação de todos os seus filhos, netos e bisnetos. Tudo isso só se tornou possível por meio do projeto chamado “Casinha”, tornando possível não só a minha formação como a de vários outros familiares.

Ao meu pai José Adelmo por tudo o que fez por mim, me incentivando, se preocupando com meu desempenho, torcendo por mim e o principal confiando em mim, obrigado paizão.

À minha mãe Maria do Espírito Santo (Marah) por sua confiança, carinho, amizade, seu bom humor e por sempre estar ao meu lado mesmo estando longe. Agradeço também ao meu padrasto Isaias Farias por sempre me ajudar no que pode, por sua amizade e lealdade.

Aos meus irmãos Mahélen Sousa e Gustavo Sousa e também minha sobrinha Manuella (Manu) pelo carinho que sempre tiveram por mim, pelos momentos felizes que me proporcionaram e também pela nossa amizade, amo vocês.

Quero agradecer à minha linda amada, Samara Teles, uma pessoa muito especial que tenho sempre comigo, é uma mulher incrível que me ajuda em todos os momentos, não importa o quão possa ser difícil. Ela sempre está ali me aconselhando e me guiando, te amo.

Aos meus tios em especial Rosilene Assunção que foi quem me incentivou a sair de Cametá e tentar uma vida nova em Belém, ao Martiniano Junior e sua esposa Eliana Nunes por sempre me ajudarem nos momentos difíceis, ao Alcindo Assunção por seu bom humor, sua companhia e suas caronas, gostaria de agradecer também às minhas segundas mães tia Liduina, tia Nazaré, tia Marilene e todos os outros tios tanto maternos quanto paternos por estarem sempre me ajudando direta ou indiretamente. Todos vocês são parte fundamental desta história.

Ao meu primo Marthyriano Netto meu meio irmão que me ensinou muitas coisas desde convivência até cozinhar e com certeza contribuiu muito na minha caminhada, a minha prima Melyane Gaia que também aprendi muito quando morávamos juntos, ao meu primo Marcelo Gaia que foi o melhor parceiro que tive em todos esses anos e pessoa cuja sua história me inspirou muito, ao meu primo Hugo Rodrigues, aos meus primos Marcos Gaia e Milene Gaia, ao meu primo Caíque Scerne e todos os primos que contribuíram diretamente e indiretamente para a realização deste trabalho.

Agradeço aos meus grandes amigos que conquistei durante todos esses anos de UFPa, Eric Pereira, Paulo Jr., Tiago Araújo, Thiago Bechara (Becharada), Alexandre Freitas, Paulo Souza de Sousa Souza, Brunelli Miranda, Wendel, Talvane Lima, Arthur e Géssica, Enilson Santos, Elvis e Jefferson Figueira obrigado por todos os momentos de companheirismo, conselhos, risadas, e ajudas que foram muito importantes para conclusão deste trabalho.

Aos meus amigos Cássio Barbosa, Rangel Barbosa e Clismerson obrigado pelos conselhos, risadas e ajudas contribuindo muito na minha formação.

À Universidade Federal do Pará pela minha formação profissional no decorrer dos anos de graduação.

A todas as pessoas que contribuíram direta ou indiretamente para a realização deste trabalho, o meu muito obrigado.

“Quando a maré enche, eleva todos os barcos” (Wilson Fisk)

RESUMO

A Biologia de sistemas é a análise do comportamento e interação entre os componentes de um sistema biológico, sendo uma área que vem crescendo bastante e que permite diversos tipos de análises, como a modelagem de processos biológicos e simulações. A análise do balanço de fluxo (FBA) é um método matemático utilizado para analisar o fluxo de metabólitos de uma rede metabólica, permitindo prever a taxa de crescimento de um organismo ou a taxa de produção de um metabólito de interesse. Para o cálculo do FBA, faz-se necessário e essencial o uso de matriz estequiométrica, estrutura que indica os diversos componentes que participam de uma reação (reagentes e produtos) e suas proporções moleculares relativas. Esta matriz acaba sendo difícil de ser construída manualmente, pois demanda tempo pelo fato de que uma via metabólica possui muitas reações e componentes. Em nosso estudo desenvolvemos o sARCASM (semi-Automatic Reconstruction and Analysis of Stoichometric Matrix), um conjunto de *scripts* escrito em Python cujo objetivo é automatizar, padronizar e diminuir o tempo do cálculo comparado a outros *pipelines* que utilizam o FBA. O sARCASM ajuda na reconstrução *in silico* de vias metabólicas de interesse e utiliza as informações de anotação de um genoma alvo para que por meio do banco de dados do KEGG, sejam obtidos os grupos de genes ortólogos, assim como o modelo estequiométrico, geração do modelo no formato SBML e aplicação do método FBA. Foi utilizado como testes a via metabólica da biossíntese de ácidos graxos e o organismo estudado foi a *Exiguobacterium antarcticum* B7 e os resultados obtidos foram bastante aproximados quando comparados com um trabalho de referência.

Palavras-chave: Matriz estequiométrica, reconstrução de vias metabólicas, análise baseada em restrições.

ABSTRACT

Systems Biology is the analysis of the behavior and interaction between the components of a biological system, being an area that has been growing and allowing several types of analysis, such as the modeling of biological processes and simulations. Flux balance analysis (FBA) is a mathematical method used to analyze the metabolic flux of a metabolic network, allowing prediction of the growth rate of an organism or the rate of production of a metabolite of interest. In order to calculate the FBA, it is necessary and essential to use a stoichiometric matrix, a structure that indicates the various components that participate in a reaction (reagents and products) and their relative molecular proportions. This matrix ends up being difficult to be constructed manually, because it demands time for the fact that a metabolic pathway has many reactions and components. In our study, we developed sARCASM (semi-Automatic Reconstruction and Analysis of Stoichiometric Matrix), a set of scripts written in Python whose objective is to automate, standardize and reduce calculation time compared to other pipelines that use the FBA. SARCASM assists in the in silico reconstruction of metabolic pathways of interest and uses the annotation information of a target genome so that, through the KEGG database, the orthologous gene groups are obtained, as well as the stoichiometric model, generation of the model in the SBML format and application of the FBA method. The metabolic pathway of fatty acid biosynthesis was used as tests and the organism studied was *Exiguobacterium antarcticum* B7 and the results obtained were quite approximate when compared with a reference work.

Keywords: Stoichiometric matrix, reconstruction of metabolic pathways, analysis based on constraints.

LISTA DE ILUSTRAÇÕES

Figura 1: Diagrama proposto por Francis Crick. Retirado de (Watson et al., 2015).	23
Figura 2: A mudança na ênfase da pesquisa biológica. Retirado de (Palsson, 2000).	25
Figura 3: Exemplo de via metabólica, a-e representam os metabólitos, R1 e R6 possuem apenas uma seta por serem irreversíveis, as demais com duas setas são reversíveis e seta com duas pontas são reações externas. Adaptado de (Schilling et al.,2000).....	28
Figura 4: Esquema de reação representando parte do metabolismo do monossacarídeo. As reações reversíveis são indicadas por pontas de flecha duplas. A definição de sinal para as reações reversíveis é tal que o fluxo é tomado positivo se ele for de baixo para cima. Retirado de (Schuster et al., 2000).	29
Figura 5: Exemplo de uma via Metabólica, via de biossíntese de ácidos graxos. Retirado de KEGG (2017).....	31
Figura 6: Visão geral do KGML. Retirado de (Kanehisa et al.,2004).	32
Figura 7: O procedimento geral do KAAS. Adaptado de (Moriya et al.,2007)	34
Figura 8: Adaptação da representação do modelo de uma matriz estequiométrica proposto por Palsson, 2013.	36
Figura 9: Representação de uma via metabólica fictícia e sua matriz estequiométrica. Adaptado de (Szallasi; Stelling; Periwál, 2006).....	37

Figura 10: A base conceitual de modelagem baseada em restrições. Da esquerda para a direita: O primeiro gráfico mostra que a distribuição sem restrições, pode estar em qualquer ponto; O segundo mostra o espaço após as restrições; e o terceiro mostra a função objetivo Z maximizada. Adaptado de Orth et al., 2010. 39

Figura 11: Formulação de um problema FBA. Adaptado de Orth et al., 2010. 41

Figura 12: Fluxograma de utilização do sARCASM. 43

Figura 13: Via metabólica da biossíntese de ácidos graxos. Retirado de KEGG (2017)..... 48

Figura 14: Resultado retirado de Kawasaki (2016) 53

Figura 15: Resultado obtido pelo sARCASM..... 53

LISTA DE TABELAS

Tabela 1: Informações sobre o arquivo baixado que contém os genes a serem utilizadas na criação da Matriz estequiométrica. Fonte: O autor.	47
Tabela 2: Genes relacionados com suas respectivas reações	49
Tabela 3: Fluxos de entrada e saída do FBA.	51
Tabela 4: Objetivos e suas devidas porcentagens obtidas no cálculo do FBA.	52
Tabela 5: Tabela de comparação de resultados de FBA.	54

LISTA DE SIGLAS

FBA	<i>Flux balance analysis</i>
GPR	Gene proteína reação
HTML	<i>HyperText Markup Language</i>
KAAS	<i>KEGG Automatic Annotation Server</i>
KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
KGML	<i>KEGG Markup Language</i>
KO	<i>KEGG Orthology</i>
RNA	<i>Ribonucleic acid</i>
sARCASM	<i>Semi-Automatic reconstruction and analysis of stoichiometric matrix</i>
XML	<i>eXtensible Markup Language</i>
SBML	<i>Systems Biology Markup Language</i>
ATP	<i>Adenosine triphosphate</i>
ADP	<i>Adenosine diphosphate</i>

SUMÁRIO

1. INTRODUÇÃO.....	17
1.1 Justificativa	17
1.2 Contexto.....	18
1.3 Objetivos.....	20
1.3.1 Objetivo Geral	20
1.3.2 Objetivo Específico	20
1.4 Organização do Trabalho	21
2. REFERÊNCIAL TEÓRICO.....	22
2.1 Biologia de sistemas.....	22
2.2 Reconstrução de Modelos Metabólicos.....	26
2.2.1 Vias metabólicas.	27
2.3 Bancos de Dados e Ferramentas Genômicas.....	30
2.3.1 <i>Kyoto Encyclopedia of Genes and Genomes (KEGG)</i>	30
2.3.2 <i>KEGG Markup Language (KGML)</i>	31
2.3.3 <i>KEGG Automatic Annotation Server (KAAS)</i>	33

2.3.4 <i>System Biology Markup Language (SBML)</i>	34
2.4 Análise e modelagem de modelos metabólicos	36
2.4.1 Matrizes estequiométricas	36
2.4.2 FBA (<i>Flux Balance Analysis</i>)	38
3. UMA FERRAMENTA SEMI-AUTOMÁTICA PARA RECONSTRUÇÃO E ANÁLISE DE MATRIZES ESTEQUIOMÉTRICAS	42
3.1 Módulos do sARCASM.....	43
3.1.1 <i>Script: sARCASM_KtM.py</i>	43
3.1.2 <i>Script: sARCASM_CtS.py</i>	44
3.1.3 <i>Script: sARCASM_FBA.py</i>	45
4. RESULTADOS.....	46
4.1 Criação de matriz estequiométrica da <i>Exiguobacterium antarcticum</i> B7.....	46
4.2 Geração do Arquivo no formato SBML	50
4.3 Cálculo e análise do FBA	50
5. CONSIDERAÇÕES FINAIS	55
5.1 Trabalhos futuros	56
REFERÊNCIAS.....	58

1. INTRODUÇÃO

1.1 Justificativa

O processo de análise e reconstrução de vias metabólicas apresenta alguns problemas e limitações quando a quantidade de informações da via metabólica é muito grande.

Um dos problemas e limitações que mais devem ser considerados é o fato de que o processo de análise e reconstrução é feito em parte por processos manuais como a criação de matrizes estequiométricas, esse processo apesar de não ser complexo leva muito tempo para ser realizado. A matriz estequiométrica possui muitas reações que estão relacionados a diversos metabólitos, informações de reversibilidade e ainda devem conter uma lista relacionando cada reação aos seus genes correspondentes.

Outro problema relacionado ao processo de análise e reconstrução de vias metabólicas é a dificuldade de se obter o arquivo SBML, modelo computacional que busca padronizar o formato de arquivo utilizados em ferramentas relacionadas a biologia de sistema. Existe uma grande dificuldade em se obter esse tipo de arquivo a partir de matrizes estequiométricas ou de arquivos no formato KGML. Existem conversores que prometem fazer a conversão de SBML para KGML (Wrzodek, 2013), porém podem conter erros devido a formatos SBML incompatíveis com determinadas ferramentas, acaba não fazendo a conversão de matrizes estequiométricas o que dificulta modificações específicas em um organismo estudado.

1.2 Contexto

No início dos estudos biológicos, acerca de organismos biológicos, a biologia era baseada na ideia reducionista, ou seja, estudar os organismos e seus componentes de forma isolada, compreender cada uma de suas funções afim de compreender o organismo como um todo. Porém após os primeiros sequenciamentos de genoma houve uma grande demanda por novas ideias para trabalhar com a crescente quantidade de dados que se obteve. Foi nesse momento então que surgiu a necessidade de mudar o paradigma reducionista e ter uma abordagem holística na biologia (Palsson, 2000).

A biologia de sistemas surge e utiliza a abordagem holística para decifrar as complexidades dos sistemas biológicos. Ela parte da compreensão de que todas as redes e componentes que formam os organismos vivos são mais do que a soma de suas partes. A biologia de sistemas é multidisciplinar pois integra muitas disciplinas científicas como a biologia, informática, engenharia, bioinformática, física e outros. E tem como objetivo prever como esses sistemas mudam ao longo do tempo e em condições variadas e para desenvolver soluções para os problemas ambientais e de saúde.

São necessários alguns passos como a modelagem e a simulação de genes para chegar no objetivo principal da biologia de sistemas que é a predição e geração de novas hipóteses biológicas. Modelagem é o processo de transformar informações biológicas em um padrão matemático para que possam ser utilizados métodos matemáticos para obter uma solução desejada. Um exemplo de modelagem dentro da biologia é a matriz estequiométrica, uma matriz que contém o balanceamento das reações, ou seja, suas estequiometrias e será mostrado com mais detalhes no decorrer deste trabalho. Já a simulação é feita por meio dos dados modelados, utilizando-se de processamento computacional para simular os modelos metabólicos. Existem diversos métodos de

simulação e um deles é o FBA (*Flux Balance Analysis*), um método matemático para simular o metabolismo de redes metabólicas que utiliza a programação linear para calcular uma solução de fluxos dentre os possíveis caminhos em uma via metabólica, permitindo assim prever a taxa de crescimento ou a taxa de produção de um determinado metabólito em um organismo.

Na elaboração do trabalho foi necessário trabalhar com dois formatos de arquivo que são padrões de ferramentas diferentes que se unificadas em um só formato irá facilitar muito o processo de modelagem e análise. Um desses formatos é o KGML (*KEGG Markup Language*) criado pelo KEGG, uma forma de visualização em um formato baseado em XML que tem como objetivo facilitar a análise de vias metabólicas. Outro formato de arquivo utilizado neste trabalho é o SBML (*System Biology Markup Language*), outra linguagem baseada em XML criada para facilitar os estudos da biologia de sistema e padronizar os formatos das diferentes ferramentas utilizado por biólogos de sistemas.

Esse trabalho tem como objetivo filtrar os componentes de uma via metabólica de interesse retirando do genoma apenas os genes ortólogos que fazem parte da via, ortólogos também são chamados de KO (*KEGG Ortholog*), são genes que possuem a mesma função dentro de um organismo e possuem um ancestral comum. Outro objetivo é facilitar a modelagem de vias metabólicas, ou seja, simplificar a tradução de arquivos KGML em matriz, posteriormente em arquivo SBML e permitir utilizar o SBML tanto no cálculo do FBA ou em qualquer outra ferramenta de análise. Foi então construído um *script* que contem três módulos objetivando poupar tempo na construção de matrizes, geração de arquivo no formato SBML e por consequência utilizar o cálculo do FBA.

1.3 Objetivos

1.3.1 Objetivo Geral

Desenvolver uma ferramenta de software para obtenção de matrizes estequiométricas, reconstrução e análise de vias metabólicas.

1.3.2 Objetivo Específico

- Obter da via de biossíntese de ácidos graxos;
- Criação dos *scripts* para criar a Matriz, SBML e FBA;
- Realizar testes para validar o processo de criação da matriz e cálculo do FBA;
- Curadoria da matriz estequiométrica;
- Analisar a via de biossíntese de ácidos graxos e genoma da *Exiguobacterium antarcticum* B7;
- Calcular o FBA.

1.4 Organização do Trabalho

Este trabalho está organizado em 5 seções:

1 INTRODUÇÃO (página 17): Apresenta o trabalho, mostrando sua relevância para a área.

Assim como os objetivos do estudo.

2 REFERÊNCIAL TEÓRICO (página 22): Apresenta os conceitos e ferramentas importantes para o entendimento e realização deste trabalho.

3 MATERIAIS E MÉTODOS (página 42): Apresenta com detalhes os três módulos que compõem o sARCASM, os cenários de testes, sua construção, utilização, funcionalidade e as métricas para avaliação da interação que foram utilizadas no sARCASM.

4 RESULTADOS (página 46): Apresenta como se deu a aplicação da solução para dados reais extraídos de um micro-organismo, assim como a descrição dos resultados e a discussão do mesmo.

5 CONSIDERAÇÕES FINAIS (página 55): Apresenta algumas considerações sobre o trabalho desenvolvido e trabalhos futuros.

2. REFERÊNCIAL TEÓRICO

2.1 Biologia de sistemas

A primeira sequência de genoma completo foi disponibilizada por Fleischmann (1995), dando início a era do genoma. Depois disso, um grande número de tecnologias de alto desempenho permitiu conhecer várias peças moleculares de componentes celulares, tornando possível reconstruir o conjunto de reações bioquímicas em escala genômica que ocorrem dentro das células. Este processo de reconstrução de rede é a essência da biologia de sistemas (Palsson, 2006).

Segundo Koide e seus colaboradores (2009) a biologia de sistemas usa uma abordagem integrativa unindo diversas disciplinas científicas como a biologia, informática, engenharia, bioinformática, física entre outros e entender como um organismo funciona por inteiro, estudando suas inter-relações funcionais levando em consideração o ambiente entre suas partes constituintes (por exemplo, genes, RNAs, proteínas e metabólitos), entender as interconexões entre os distintos membros desse conjunto de interações, até a compreensão do organismo como um todo.

Antes de prosseguir é importante falar sobre os paradigmas da biologia molecular, ou seja, o dogma central da biologia, nome proposto por Francis Crick em 1956 e representado pela Figura 1, a figura mostra o fluxo de informação genica e as setas indicam a direção que o fluxo deve obedecer (Watson et al., 2015).

Na Figura 1, da esquerda para a direita, mostra o DNA (*deoxyribonucleic acid*), traduzindo para o português, ácido desoxirribonucleico, encontrado em sua grande maioria no núcleo da célula, tem formato de dupla hélice como uma escada espiral, unidas por ponte de hidrogênio entre suas bases nitrogenadas que são: adenina sempre pareada com a timina, e a citosina com a guanina. A seta circundando o DNA significando que o DNA é o molde para a própria replicação,

transmitindo exatamente as mesmas informações nele contido para as suas cópias (Gonçalvez 2016; Watson et al., 2015).

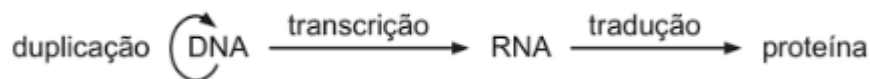


Figura 1: Diagrama proposto por Francis Crick. Retirado de (Watson et al., 2015).

A segunda seta na Figura 1 entre o DNA e o RNA indica a síntese de RNA chamada de transcrição. O RNA (*RiboNucleic Acid*), traduzindo para o português, significa ácido ribonucleico, tem uma constituição parecida com o DNA exceto por duas diferenças: o açúcar constituinte do DNA é a desoxirribose e do RNA é a ribose e a segunda diferença é que o RNA possui a base nitrogenada uracila no lugar da timina do DNA. Quando uma célula precisa de um determinado componente, o segmento do DNA responsável por esse componente é então codificado em RNA, esse segmento do DNA é chamado de gene que pode abranger alguns pares de nucleotídeos ou até mesmo centenas, cada gene codifica a produção de uma molécula específica de RNA e esse processo é conhecido como transcrição (Alberts, 2006).

A seta chamada tradução, na Figura 1, é a etapa de produção da proteína, nessa etapa é retirado um trecho do RNA, chamado de códon, e então é produzido uma sequência de aminoácidos, moléculas orgânicas que servem como unidade fundamental para a etapa seguinte que é a formação das proteínas. Proteínas são macromoléculas que contém várias cópias de um componente menor, elas estão em todas as células e possuem diferentes tamanho, são formados por combinações de 20 aminoácidos e essas combinações conseguem formar uma enorme variedade de compostos com funções muito diferentes entre si como por exemplo: enzimas, hormônios, anticorpos, músculos e participam de quase todas as funções celulares (Cox, 2008; Alberts, 2006; Watson et al., 2015).

Vale destacar as enzimas que são responsáveis pelas reações químicas dentro e fora das células e por isso são muito importantes para este trabalho.

Com a necessidade de compreender melhor as mínimas partes de um organismo surgiram as tecnologias "ômicas": como a genômica que estuda a estrutura, função e expressão de todos os genes em um organismo, a transcriptômica, estudo do RNA mensageiros (mRNA) o responsável pela transcrição, dentro de uma célula ou organismo, a proteômica que estuda as proteínas e a metabolômica que estuda de metabolitos em um sistema (Horgan et al., 2011). Segundo Harrington e seus colaboradores (2008), após o surgimento das tecnologias "ômicas" obteve-se uma grande quantidade de dados biológicos oriundos de diversos *habitats* e com objetivo de aproveitar todo potencial desses dados, necessita de tecnologia para converter sequência genômica em conhecimento biológico.

Segundo Palsson (2000) a necessidade de compreender o organismo como um todo e o aproveitamento dos dados "ômicos" contribuiu para a mudança da abordagem biológica de reducionista para integrativa ou holística.

A abordagem reducionista estuda os menores componentes de um organismo de forma isolada (Looijen, 1998) como mostrado ao lado esquerdo da Figura 2 representados pela genômica, transcriptômica e proteômica.

A abordagem integrativa ou holística, abordagem que se baseia em estudar o sistema com todas as partes constituintes no mesmo ecossistema pois possuem um comportamento diferente quando estão juntos de que quando estão separadas (Looijen, 1998) mostrado ao lado direito da Figura 2 representados pela bioinformática, modelagens matemáticas e a simulação *in silico*.

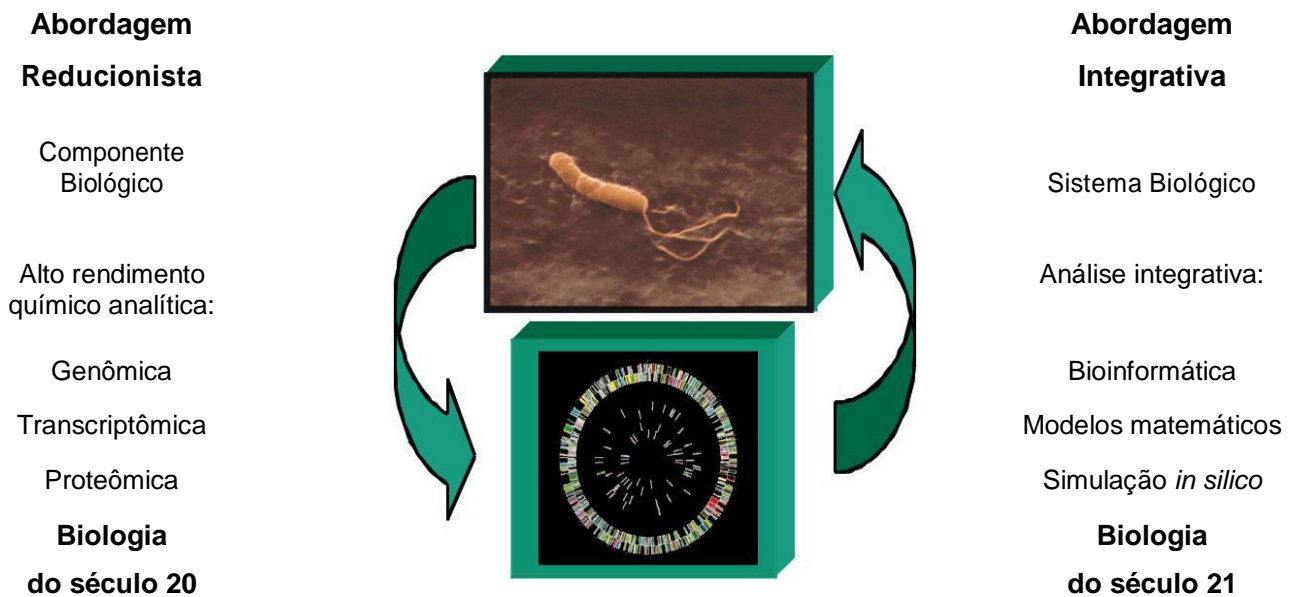


Figura 2: A mudança na ênfase da pesquisa biológica. Retirado de (Palsson, 2000).

Após a mudança de ênfase da pesquisa na biologia obteve-se listas mais detalhadas de componentes biológicos, foi possível determinar suas interações e gerar um amplo conjunto de dados genômicos, isso tudo levou ao surgimento de um paradigma fundamental para a biologia de sistemas (Ideker et al., 2011): começa com os componentes que fazem parte do organismo e suas interações e ligações, depois uma representação matemática, ou seja, modelagem *in silico* e a simulação para predizer e gerar hipóteses que podem então ser testadas experimentalmente (Bruggeman et al., 2007; Palsson, 2015).

As próximas seções mostrarão mais detalhes sobre cada etapa do paradigma da biologia de sistemas como a reconstrução dos modelos metabólicos, bancos de dados e ferramentas e métodos de modelagem e simulação.

Inicia-se pela reconstrução de redes metabólicas responsáveis por listar componentes que compõe um organismo e a interação entre esses componentes. Segundo Thiele e Palsson (2010), o objetivo comum para estudos da biologia de sistemas de um organismo alvo é a reconstrução da rede metabólica em escala genômica de alta qualidade.

2.2 Reconstrução de Modelos Metabólicos

A reconstrução de uma rede representa a base de conhecimento bioquímico, genético e genomicamente estruturada que contém informações de qualidade detalhadas sobre um organismo.

A ideia principal é reconstruir as mesmas reações por meio das informações encontradas no genoma do organismo, essas reconstruções contêm informações como as estequiometrias, ou seja, a quantidade exata de reagentes que deve ser usada e a quantidade de produtos que será obtida nas reações metabólicas (Fogaça, 2017), fórmulas químicas, cargas de metabólitos e as associações entre genes, proteínas e reações (GPRs). Essas reconstruções formam uma base para a formulação de relações genótipo-fenótipo mecânicas e, portanto, computáveis, do genoma (Palsson, 2009).

Para iniciar a reconstrução precisa-se da anotação do genoma, experiências bioquímicas e dados da fisiologia celular fornecendo informações para descrever todas as reações, quais são os substratos, cofatores, produtos gerados, estequiometria, reversibilidade e localização celular dentro de um sistema. Uma rede de reações é criada a partir do conjunto de diversos dados definindo todas as diferentes reações que ocorrem em um organismo. Estes dados são então utilizados para criar um modelo matemático que relaciona todas as reações dentro de uma rede com todos os metabólitos participantes num dado organismo, uma parte muito importante para análise *in silico* (Papin et al.,2003).

Todo o processo de reconstrução de redes metabólicas, baseado no modelo proposto por (Thiele e Palsson, 2010), pode ser resumido em quatro passos:

1) No primeiro estágio deve-se definir todos os componentes do sistema. Usar esses componentes, juntamente com o conhecimento bioquímico e genético prévio, para formular um modelo inicial. O primeiro estágio consiste na reconstrução baseado na anotação do genoma do organismo alvo. Segundo Médigue e seus colaboradores (2007) anotação, em termos gerais, é a extração de conhecimento biológico a partir de sequências brutas de nucleotídeos, e utilização de bancos de dados bioquímicos;

2) São estudadas as interações entre os componentes gerados no primeiro passo, os "diagramas de fiação" dos circuitos genéticos são reconstruídos e os mapas da escala do genoma são formados de forma gradual. Este processo é uma reconstrução da rede de reação bioquímica;

3) Na terceira fase, as redes reconstruídas são convertidas em um formato matemático que descreve formalmente o conhecimento biológico subjacente à rede reconstruída. Modelos de computador são então gerados para analisar, interpretar e prever as funções biológicas que podem surgir a partir de redes reconstruídas;

4) Na última fase o modelo será simulado e a predição resulta em gerar hipóteses específicas que serão testadas experimentalmente. Estes modelos *in silico* de redes reconstruídas são então melhorados de forma iterativa (Palsson, 2015).

Para representar a reconstrução da rede de reações bioquímicas é utilizado a via metabólica, que se trata de uma série de reações químicas, interligadas, que ocorrem dentro de uma célula.

2.2.1 Vias metabólicas.

Metabolismo é a soma de todas as transformações químicas que ocorrem em uma célula ou organismo que ocorre através de uma série de reações catalisadas por enzimas que constituem as

vias metabólicas (Cox; Nelson.,2008). Os reagentes, produtos e intermediários de uma reação enzimática são conhecidos como metabólitos representados pelas letras (A-E) na Figura 3. Em uma via metabólica, o produto de uma enzima atua como substrato para a próxima (Schuster et al., 2000).

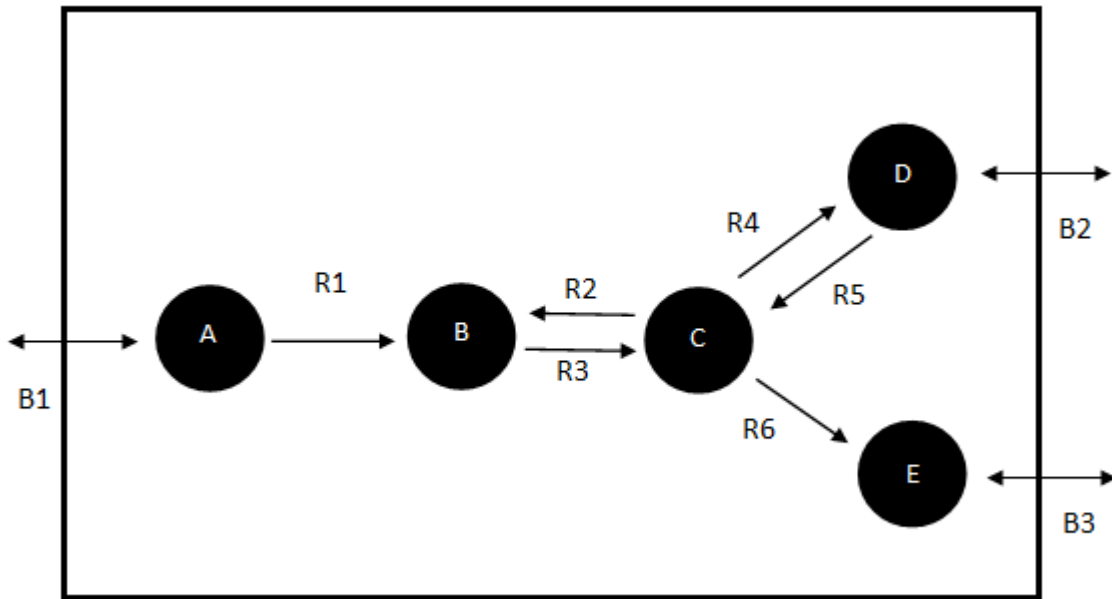


Figura 3: Exemplo de via metabólica, a-e representam os metabólitos, R1 e R6 possuem apenas uma seta por serem irreversíveis, as demais com duas setas são reversíveis e seta com duas pontas são reações externas. Adaptado de (Schilling et al.,2000)

Vias metabólicas podem estar em lugares diferentes dentro de uma da célula, por exemplo, o ciclo do ácido cítrico, cadeia de transporte de elétrons e fosforilação oxidativa ocorre na membrana mitocondrial. Em contraste, a glicólise, a via de pentose fosfato e a biossíntese de ácidos graxos ocorrem no citosol de uma célula (Pratt e Voet., 2013).

A Figura 4 mostra a distinção entre reações reversíveis e irreversíveis. A irreversibilidade aqui significa que o fluxo líquido sempre tem o mesmo sinal em condições fisiológicas. Os metabólitos

são classificados como internos ou externos de acordo com a condição de não cumprirem a condição do estado quasi-estável. Em outras palavras, a taxa total de produção de cada metabólito interno é igual à taxa total de seu consumo. Em contrapartida, os metabólitos externos não cumprem esta condição porque participam de reações adicionais que não estão envolvidas no sistema em estudo (Schuster et al., 2000).

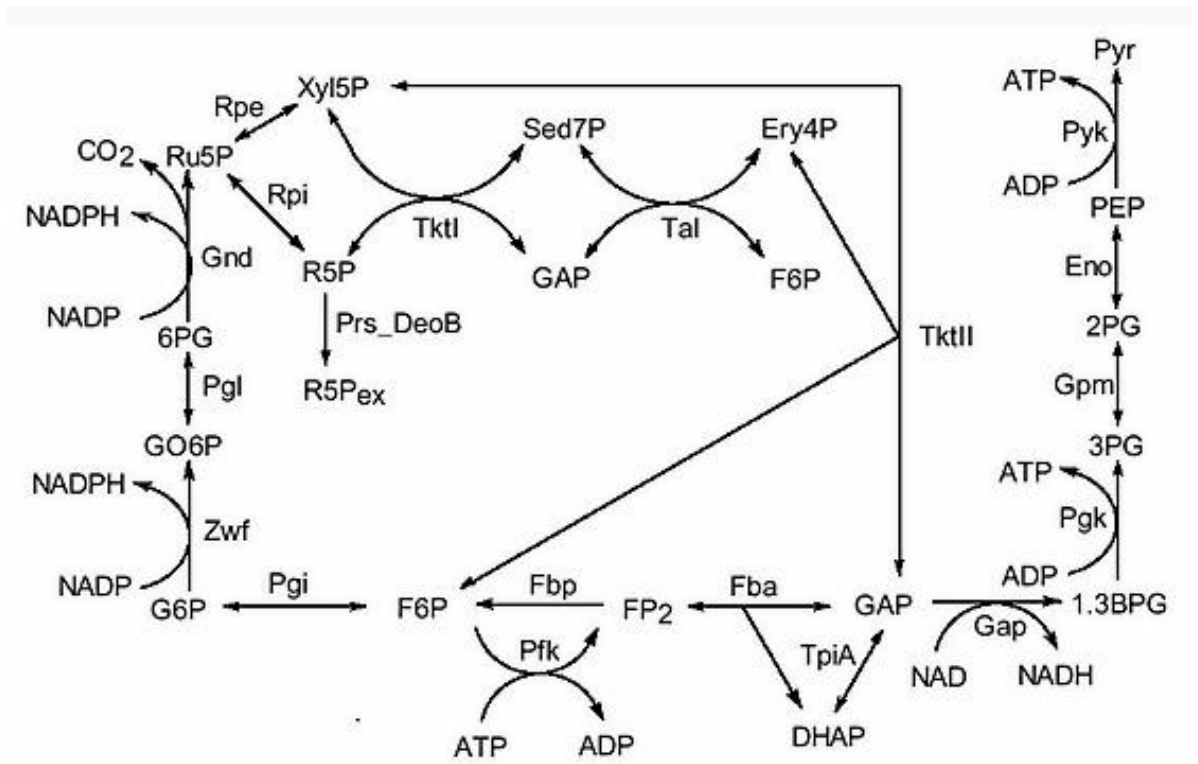


Figura 4: Esquema de reação representando parte do metabolismo do monossacarídeo. As reações reversíveis são indicadas por pontas de flecha duplas. A definição de sinal para as reações reversíveis é tal que o fluxo é tomado positivo se ele for de baixo para cima. Retirado de (Schuster et al., 2000).

Diversos bancos de dados e ferramentas genômicas foram criados para agrupar as informações geradas por novas tecnologias de sequenciamento, concentrar informações sobre genomas, proteína e nucleotídeos, facilitar análises de genes e genomas, entre outras funcionalidades.

2.3 Bancos de Dados e Ferramentas Genômicas

2.3.1 *Kyoto Encyclopedia of Genes and Genomes (KEGG)*

O *Kyoto Encyclopedia of Genes and Genomes* é um recurso de banco de dados integrado composto por dezesseis bancos de dados curados manualmente para a compreensão de informações genômicas, químicas e sistêmicas funcionais, como a célula, o organismo e o ecossistema, a partir de informações de nível molecular, especialmente conjuntos de dados moleculares em grande escala gerados pelo sequenciamento do genoma e outras tecnologias experimentais (Kanehisa et al., 2016).

Segundo Kanehisa (2016) entre todos os bancos de dados do KEGG o *PATHWAY* é o principal, ele contém uma coleção de diagramas de caminho de referência do KEGG (mapas), cada um correspondente a uma rede conhecida de significância funcional. O banco de dados *PATHWAY* também contém caminhos específicos do organismo, que são gerados automaticamente pela sobreposição (coloração) de genes em determinados organismos como mostrado na Figura 5, onde temos uma via metabólica e os quadrados verdes representam a via específica do organismo em estudo.

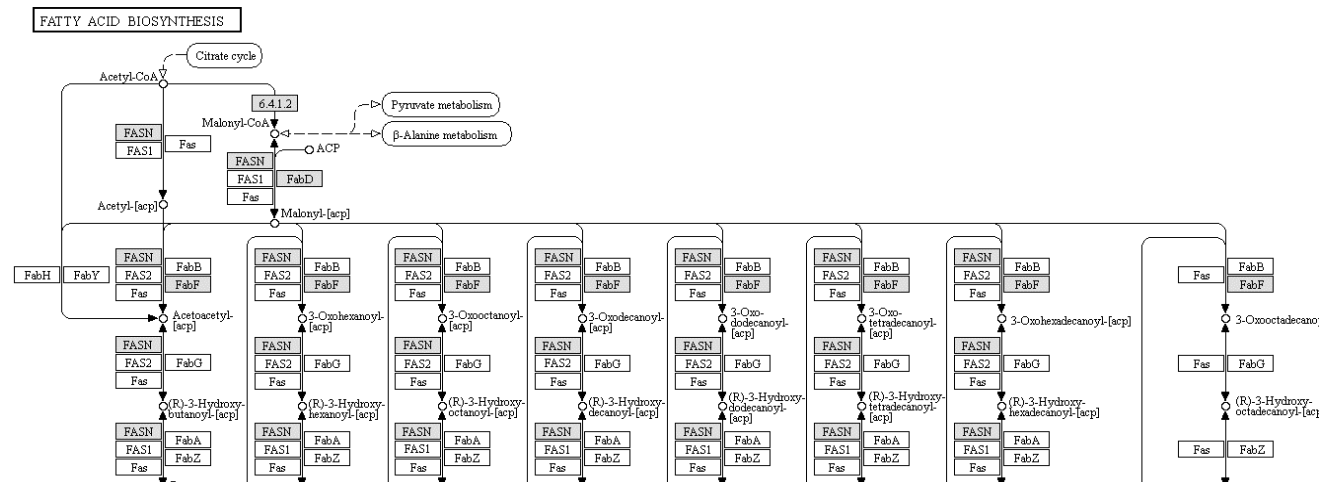


Figura 5: Exemplo de uma via Metabólica, via de biossíntese de ácidos graxos. Retirado de KEGG (2017).

No passado os diagramas de caminho estavam disponíveis apenas em arquivos de imagem.

Para facilitar a análise e modelagem foi criado o *KEGG Markup Language* (KGML)

2.3.2 KEGG Markup Language (KGML)

O KGML é um formato de representação de vias metabólicas do KEGG, permite o desenho automático das vias KEGG e fornece facilidades para análise computacional e modelagem de redes de genes/proteínas e redes químicas.

No passado, os diagramas de caminho eram disponibilizados apenas em arquivos de imagem GIF (ou PNG). Então foi criado o KGML como uma especificação de objetos de gráfico no KEGG. Todas as vias metabólicas e alguns caminhos regulatórios são disponibilizados no formato KGML, permitindo a reconstrução computacional e a manipulação de caminhos (Kanehisa et al., 2004).

A Figura 6 mostra o diagrama da estrutura do KGML. O elemento do *pathway* (via) é um elemento raiz, e um elemento *pathway* é especificado como um mapa de via metabólica no KGML. Os elementos de entrada, *relation* (relação) e *reaction* (reação) especificam as informações da via metabólica, e os demais elementos são usados para especificar informações mais detalhadas sobre nós da via metabólica.

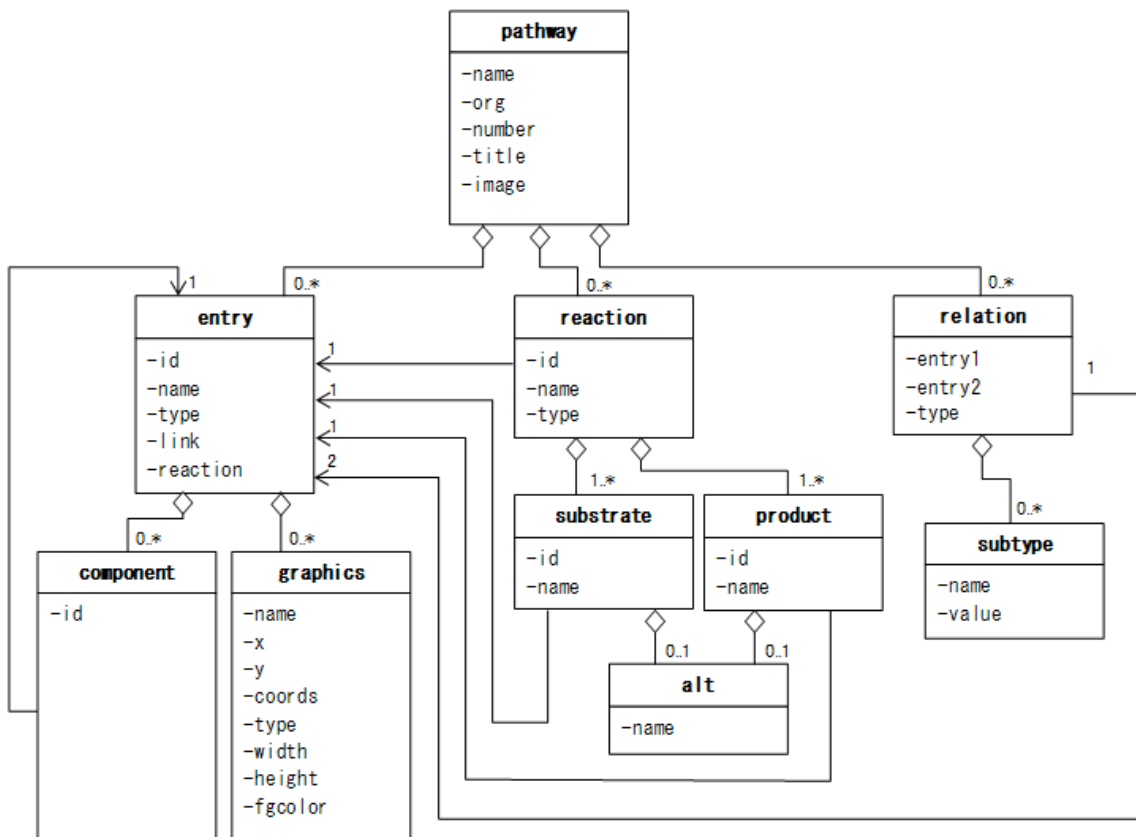


Figura 6: Visão geral do KGML. Retirado de (Kanehisa et al.,2004).

Além do KGML o KEGG oferece outras funções importantes para quem deseja fazer uma análise completa sobre um determinado organismo, um deles é o *KEGG Automatic Annotation Server* (KAAS), um método que se baseia em semelhanças de sequências para garantir a maior qualidade de um genoma anotado.

2.3.3 *KEGG Automatic Annotation Server (KAAS)*

Quando se deu início ao projeto KEGG, a integração de informações genômicas e informações de rede era realizado pelo código das enzimas chamados de “EC”. Os números da EC eram utilizados como identificadores comuns para combinar genes no genoma e produtos de genes (enzimas) na via metabólica. Em seguida, para superar vários problemas inerentes à nomenclatura enzimática introduziu-se um novo esquema de identificação por meio da Ortologia dos genes, genes ortólogos partilham um ancestral comum e têm funções iguais. O *KEGG Orthology (KO)* é uma extensão adicional deste esquema com base na análise computacional, a fim de classificar todas as funções genéticas e explorar caminhos desconhecidos (Kanehisa et al., 2004).

KEGG Automatic Annotation Server (KAAS) é um servidor automático para anotação de genomas, um método rápido para atribuir o identificador *K* (número de identificação de grupos de genes) automaticamente a genes no genoma, permitindo a reconstrução de caminhos de KEGG. O método baseia-se em semelhanças de sequências, informações bidirecionais de melhor impacto, e comparado com o banco de dados do KEGG alcançou um alto grau de precisão (Moriya et al., 2007).

A Figura 7 mostra um fluxograma de funcionamento do KAAS. Primeiro é feito um BLAST que é uma ferramenta de pesquisa em bancos de dados para encontrar similaridades de sequência (Altschul et al., 1997). O BLAST é feito com a sequência de entrada do usuário e o conjunto de sequência de referência (extraídos do banco de dados KEGG GENES) então os homólogos são encontrados no conjunto de referência. Em seguida, os homólogos são então classificados com base na similaridade obtida no BLAST e na taxa de sucesso bidirecional (explicada mais a frente), os que estiverem acima do limite são selecionados como candidatos. Taxa de sucesso bidirecional é um cálculo onde o genoma é comparado contra cada genoma no conjunto de referência do banco

de dados KEGG GENES por pesquisas BLAST em direções direta e reversa, levando cada gene no genoma A como uma consulta comparada com todos os genes no genoma B e vice-versa. Os genes que tiveram uma taxa de similaridade abaixo de 60% são retirados.

Os candidatos a ortólogos são então divididos em grupos KO de acordo com a anotação do banco de dados KEGG GENES. Finalmente, a pontuação de atribuição é calculada com base na probabilidade e heurística para cada grupo KO. Então, o número K do grupo KO com maior pontuação é atribuído à sequência de consulta (Moriya et al.,2007).

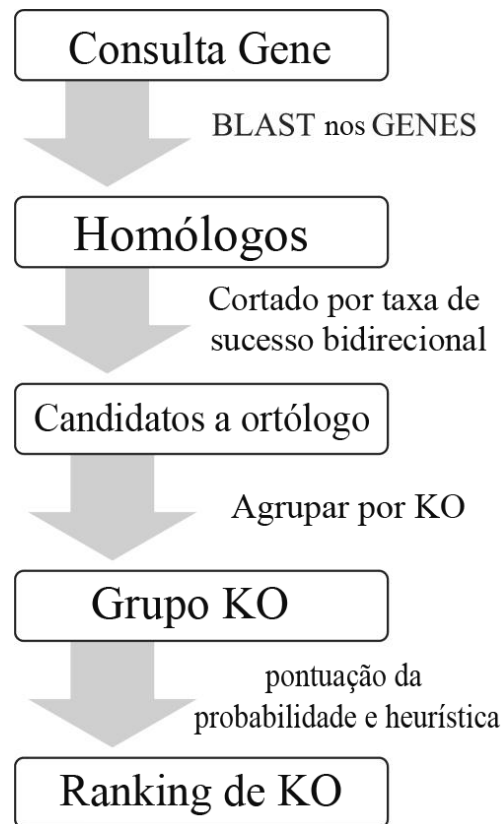


Figura 7: O procedimento geral do KAAS. Adaptado de (Moriya et al.,2007)

2.3.4 *System Biology Markup Language (SBML)*

A biologia de sistemas é caracterizada por sua interdisciplinaridade e misturando teorias, modelagem computacional e simulações. Com o surgimento de uma variedade de ferramentas computacionais para biólogos de sistemas, para processar a grande quantidade de informações

biológicas existente, ocasionou em uma grande diversidade de software acompanhada de uma variedade de incompatibilidades, caracterizado pela falta de um formato comum para descrever modelos (Hucka et al., 2003).

A dificuldade de trocar modelos entre diferentes ferramentas de simulação e análise resultou na criação de um fórum de Plataformas de Software para Biologia de Sistemas que decidiu na primeira reunião desenvolver uma linguagem simples baseada em XML (*eXtensible Markup Language*) para representar e trocar modelos entre ferramentas de simulação ou análise: o SBML (Systems Biology Markup Language) (Hucka et al., 2003).

Resumindo, o SBML é um formato legível por máquina para representar modelos. A estrutura do SBML é adequada para representar diversos modelos como vias de sinalização celular, caminhos metabólicos, reações bioquímicas, regulação de genes e muitos outros (SBML, 2012).

O desenvolvimento do SBML é dividido em níveis para organizar mudanças arquitetônicas. As principais edições do SBML são denominadas níveis ou *Levels* e apresentam mudanças substanciais na composição e estrutura da linguagem. Os modelos definidos nos níveis mais baixos do SBML, ou seja, a primeiras versões criadas podem sempre ser representados em níveis mais altos, embora seja necessária alguma tradução. O inverso também é possível, porém não é garantido.

As revisões menores do SBML são denominadas Versões e constituem mudanças dentro de um nível para corrigir, ajustar e aperfeiçoar recursos da linguagem (SBML, 2012).

Atualmente, a linguagem SBML encontra-se na versão 1, nível 3 (Hucka et al., 2016).

2.4 Análise e modelagem de modelos metabólicos

2.4.1 Matrizes estequiométricas

Uma rede biológica é composta por reações que podem ser representadas por equações químicas. Uma matriz estequiométrica é construída a partir dessas equações químicas.

A matriz estequiométrica contém informações sobre todas as transformações metabólicas incluídas em uma reconstrução metabólica para um determinado microrganismo e é formada pelos coeficientes estequiométricos das reações que constituem uma rede de reação. É organizado de tal forma que cada coluna corresponde a uma reação e cada linha corresponde a um composto. As entradas na matriz são coeficientes estequiométricos que são inteiros. (Dubitzky, 2013).

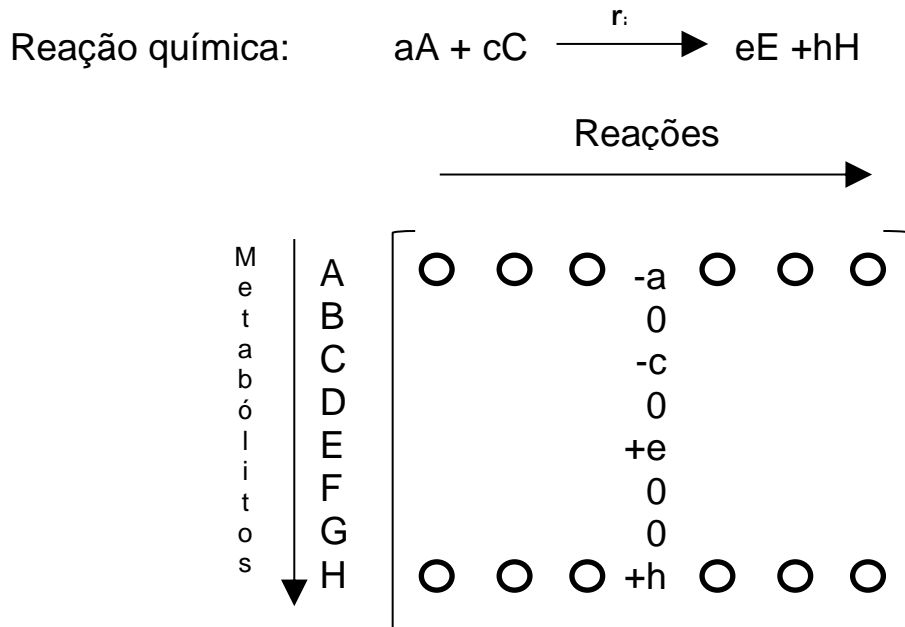


Figura 8: Adaptação da representação do modelo de uma matriz estequiométrica proposto por Palsson, 2013.

Os coeficientes estequiométricos são inteiros (a , c , e , h na Figura 8) que representam o número de moléculas de espécies químicas (A , C , E , H , na Figura 8) que são transformadas nesta reação química particular. Estes coeficientes são constantes (isto é, não são dependentes da condição, que

são funções de temperatura, pressão, pH, etc.). Além disso, eles são biologicamente universais, ou seja, essa reação metabólica procede da mesma maneira em todas as células; por exemplo, o hexoquinase sempre catalisa a reação:



A matriz toda é representada por “S” e para a formação de uma coluna em S cada metabólito tem uma linha na matriz estequiométrica e cada reação possui uma coluna. Os coeficientes estequiométricos são utilizados para formar uma coluna, com o coeficiente estequiométrico que corresponde a um metabólito particular que aparece na linha a que corresponde. Se um metabólito é formado pela reação, o coeficiente tem um sinal positivo, se ele é consumido pela reação, o coeficiente estequiométrico aparece com um sinal negativo. Todas as outras linhas (correspondentes a metabólitos que não participam da reação) ficam com o valor zero. Os coeficientes estequiométricos são geralmente unidade, ou seja, 1 ou -1 (Palsson, 2003).

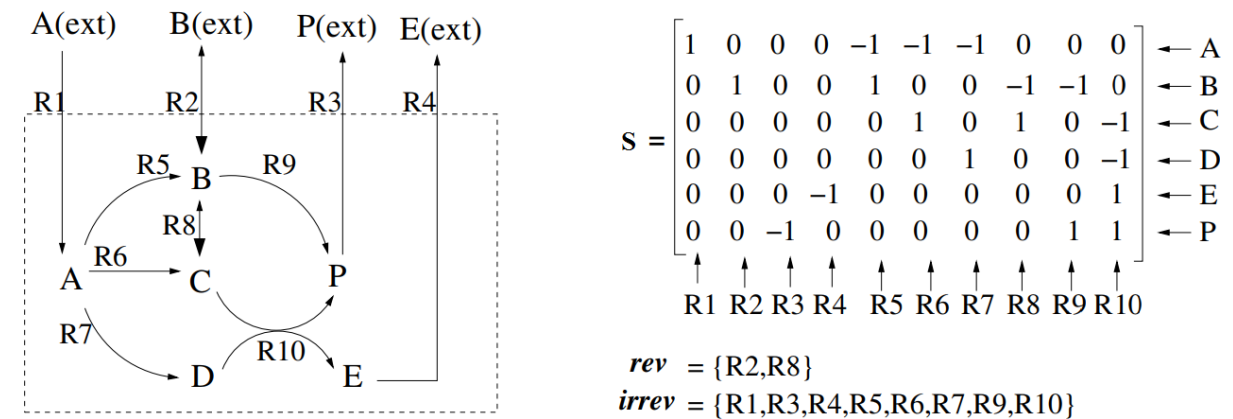


Figura 9: Representação de uma via metabólica fictícia e sua matriz estequiométrica. Adaptado de (Szallasi; Stelling; Periwal, 2006).

A Figura 9 mostra a via e as variáveis correspondentes de uma rede de exemplo simples. Esta rede compreende 6 metabolitos (internos) e 10 reações, as reações dentro de uma via podem ser reversíveis ou irreversíveis, no exemplo duas delas são reversíveis. As principais características de uma rede são suas restrições e suas conexões. Relacionado a esta questão estão os compostos internos e externos. Os compostos internos são aqueles que são mostrados explicitamente no modelo de rede, ou seja, na matriz S . Por outro lado, os compostos externos são chamados de pias ou fontes que podem estar fisicamente fora do sistema (substratos ou produtos como os quatro compostos externos na Figura 9 A(ext), B(ext), P(ext), E(ext)), mas também podem estar localizados dentro a célula (Szallasi; Stelling; Periwal, 2006).

Após a modelagem pode-se então simular o organismo *in silico* que irá predizer e gerará hipóteses que serão testadas experimentalmente. Para essa simulação e especificamente para análise de reconstruções de redes metabólicas é utilizado o FBA.

2.4.2 FBA (*Flux Balance Analysis*)

Flux Balance Analysis (FBA) é uma abordagem amplamente utilizada para o estudo de redes bioquímicas, em particular as reconstruções de redes metabólicas em escala de genoma que foram construídas na última década (Duarte, 2007).

Estas reconstruções de rede contém todas as reações metabólicas conhecidas num organismo e os genes que codificam cada enzima. A FBA calcula o fluxo de metabólitos através desta rede metabólica, tornando assim possível prever a taxa de crescimento de um organismo ou a taxa de produção de um metabólito biotecnologicamente importante (Orth, 2010).

O ponto principal do FBA são as restrições, que são definidos por coeficientes estequiométricos de cada reação de uma via metabólica e são de dois tipos, de balanço e de limites. Uma restrição de balanço é representada por uma equação por exemplo, a conservação da massa, onde em um estado estacionário, não há acumulação ou perda de compostos, restrição 1) na Figura 10. As restrições de limites são controladas por intervalos numéricos de variáveis e parâmetros individuais, tais como concentrações, fluxos ou constantes cinéticas, restrição 2) na Figura 10.

Por mais que as restrições definam uma grande quantidade de soluções, é possível identificar e analisar pontos únicos dentro do espaço da solução. Por exemplo, podemos estar interessados em identificar à taxa máxima de crescimento ou à produção máxima de um determinado metabólito em um organismo, dado o seu conjunto específico de restrições. O FBA é um método para identificar esses pontos ótimos dentro de um espaço restrito (Figura 10) (Orth et al., 2010).

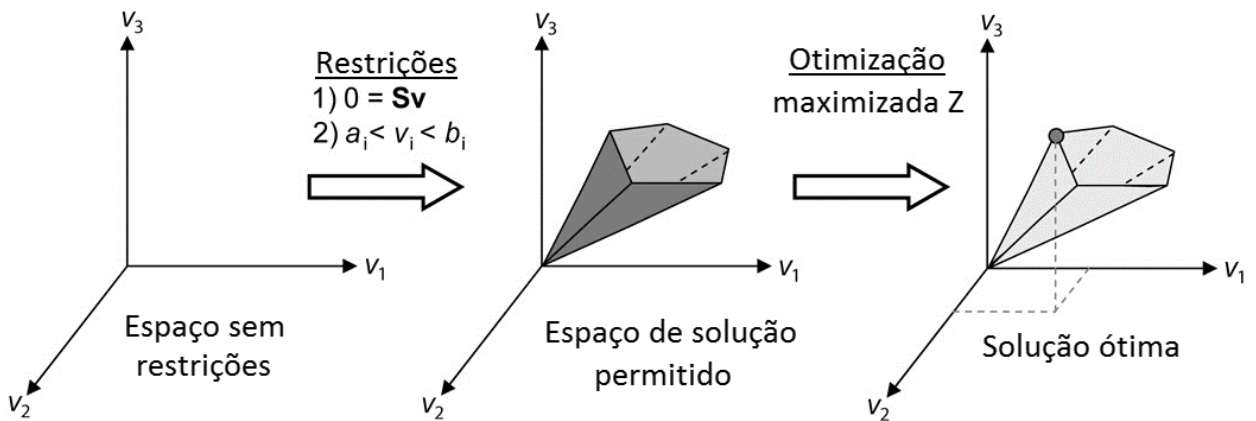


Figura 10: A base conceitual de modelagem baseada em restrições. Da esquerda para a direita: O primeiro gráfico mostra que a distribuição sem restrições, pode estar em qualquer ponto; O segundo mostra o espaço após as restrições; e o terceiro mostra a função objetivo Z maximizada. Adaptado de Orth et al., 2010.

Conforme descrito por Orth et al., 2010, o primeiro passo na FBA é representar matematicamente reações metabólicas (visto no subtópico anterior). A característica central desta representação é uma tabulação, na forma de uma matriz numérica, dos coeficientes estequiométricos de cada reação, determinando restrições ao FBA (Figura 11 a, b). Então as reações metabólicas são representadas como uma matriz estequiométrica (S), de tamanho $m * n$. Cada linha dessa matriz representa um composto único (m compostos) e cada coluna representa uma reação (n reações).

O fluxo através de todas as reações em uma rede é representado pelo vetor v , que possui um comprimento de n e as concentrações de todos os metabólitos são representadas pelo vetor x , com comprimento m . O sistema de equações de balanço de massa no estado estacionário ($d x / dt = 0$) é (Figura 11 c):

$$S * v = 0$$

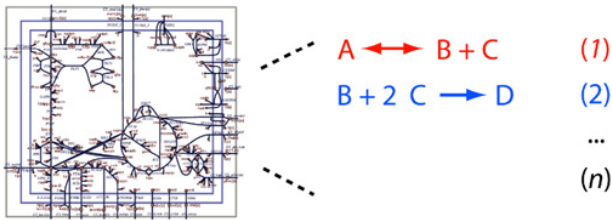
Ou seja, qualquer v que satisfaça esta equação é dito estar no espaço nulo de S . Em qualquer modelo metabólico realista em larga escala, existem mais reações do que os compostos ($n > m$). Em outras palavras, como v vai sempre ter o tamanho de n então existem diversas variáveis desconhecidas, portanto, não existe uma solução única para este sistema de equações.

O FBA procura maximizar ou minimizar uma função objetiva $Z = c^T v$, que pode ser qualquer combinação linear de fluxos, onde c é um vetor de pesos, indicando quanto cada reação contribui para a função objetivo. Na prática, quando apenas uma reação é desejada para maximização ou minimização, c é um vetor de zeros com 1 na posição da reação de interesse (Figura 11d).

A otimização de tal sistema é realizada com a ajuda da programação linear (Figura 11e). O FBA pode ser definido como o uso da programação linear para resolver a equação $Sv = 0$ dado um conjunto de limites superiores e inferiores em v e uma combinação linear de fluxos como

função objetiva. A saída do FBA é uma distribuição de fluxo particular, v , que maximiza ou minimiza a função objetivo.

a Reações metabólicas curadas



b Formular matriz S

		Reações			
		1	2	...	n
Metabólitos	A	-1			
	B	1	-1		
	C	1	-2		
	D		1		
	...				
		S			
		m			

c Aplicar restrições de balanço de massa

$S \ (m \times n)$

-1
1 -1
1 -2
1

$v \ (n \times 1)$

v_1
v_2
...
v_n

$\ast = 0 \rightarrow$

m equação de balanço de massa

$-v_1 + \dots = 0$
 $v_1 - v_2 + \dots = 0$
 $v_1 - 2 v_2 + \dots = 0$
 $v_2 + \dots = 0$
 \dots

d Define função objetivo Z

$Z =$

$c^T \ (1 \times n)$

1	0	...	0
---	---	-----	---

\ast

$v \ (n \times 1)$

v_1
v_2
...
v_n

Define a reação 1 como objetivo

e Otimize Z usando programação linear

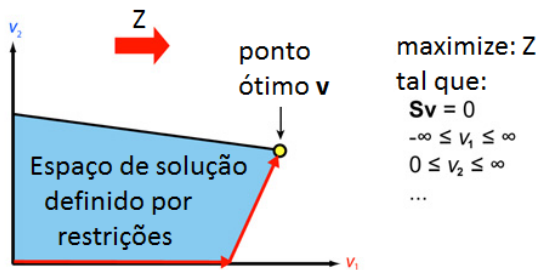


Figura 11: Formulação de um problema FBA. Adaptado de Orth et al., 2010.

3. UMA FERRAMENTA SEMI-AUTOMÁTICA PARA RECONSTRUÇÃO E ANÁLISE DE MATRIZES ESTEQUIOMÉTRICAS

Nesta seção é descrita como a ferramenta nomeada de sARCASM (*Semi-Automatic reconstruction and analysis of stoichiometric matrix*) foi desenvolvida, quais métodos foram utilizados, os ambientes em que foram testados, assim como as tecnologias que foram utilizadas e como elas contribuíram para a criação da ferramenta. O sARCASM tem como objetivo principal modelar vias metabólicas, ou seja, criar matrizes estequiométricas de vias metabólicas e simular um ambiente de restrições com a ajuda do FBA.

Para criação e utilização do algoritmo implementado neste trabalho foi necessário instalar as seguintes ferramentas: Python 3.5.2 (Python, 2017), PyCharm Community Edition 2016.3.2 (JetBrains, 2017), Anaconda 4.3.22 (Anaconda, 2017), e as bibliotecas: BeautifulSoup (Beautiful Soup, 2017), numpy (Numpy, 2017), ElementTree (Python, 2017), openpyxl (Openpyxl, 2017) e COBRApy (Ebrahim, 2013). O computador utilizado para o desenvolvimento da ferramenta apresenta a seguinte configuração: desktop Modelo L4800, com processador 4x Intel(R) Core(TM) i5-4440 CPU @ 3.10GHz e 8 GB de RAM, executando Ubuntu 14.04.4 LTS (Ubuntu, 2017), kernel Linux 3.16.0-67-generic (x86_64).

Visando facilitar a implementação, compartilhamento e reutilização da ferramenta foi utilizado a plataforma de programação Anaconda Distribution, um gerenciador de pacotes, gerenciador de ambiente e distribuição Python grátis e fácil de instalar com uma coleção de mais de 720 pacotes de código aberto com suporte gratuito à comunidade (Anaconda Documentation).

A implementação deste trabalho foi dividida em três módulos como na Figura 12: sARCASM_KtM.py, sARCASM_CtS.py, sARCASM_FBA.py. Este modelo de programação foi utilizado para facilitar o entendimento, permitir a curadoria principalmente da matriz estequiométrica e a reutilização do código.

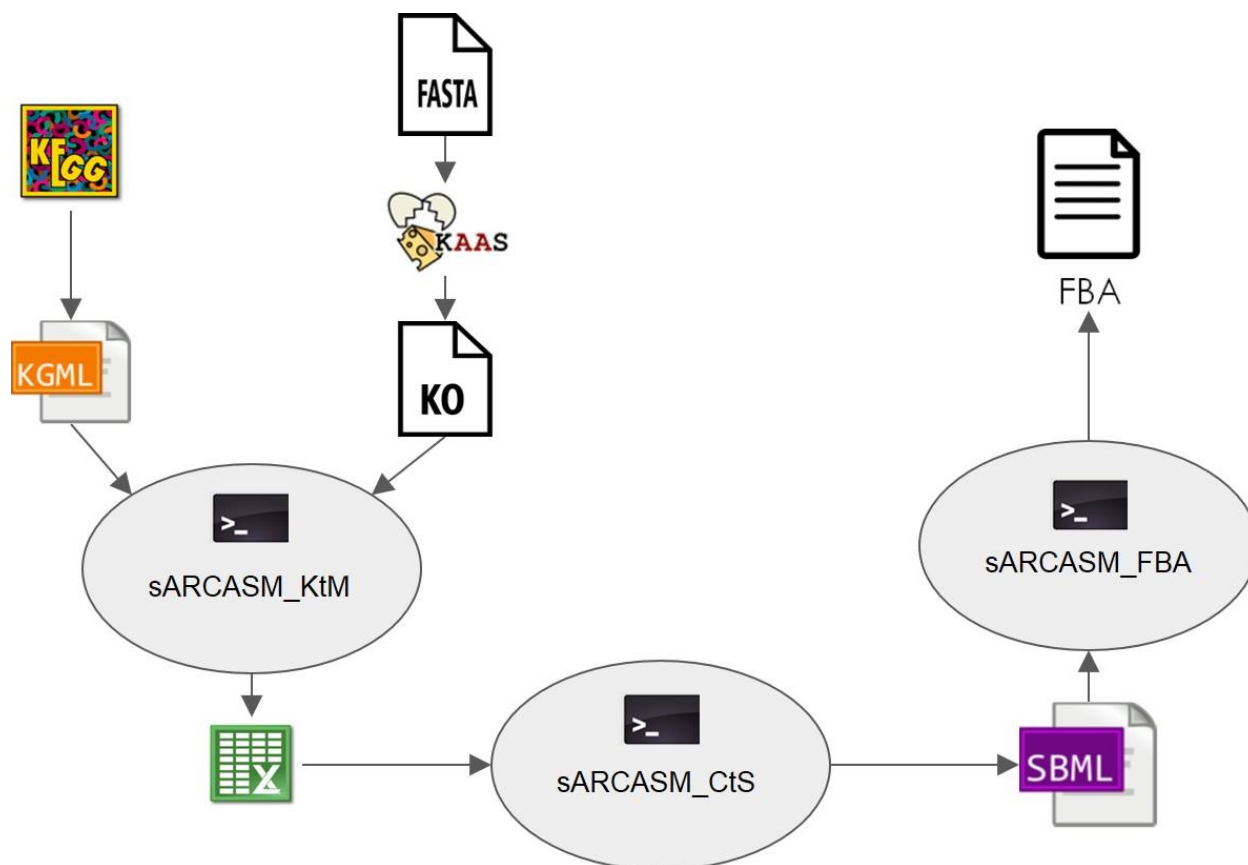


Figura 12: Fluxograma de utilização do sARCASM.

3.1 Módulos do sARCASM

3.1.1 Script: sARCASM_KtM.py

O módulo sARCASM_KtM.py tem como objetivo filtrar informações do arquivo KGML a partir dos genes ortólogos obtidos no KAAS e transformá-los em matriz no formato XLSX. Para a construção do sARCASM_KtM.py foram utilizadas as bibliotecas: Numpy um pacote fundamental para a computação científica na montagem de listas e matrizes multidimensionais (Numpy, 2017), *Beautifulsoup* é uma biblioteca Python utilizada para extrair dados de arquivos

HTML e XML (Beautiful Soup, 2017), *ElementTree* uma biblioteca para análise e construção de XML baseada em árvores. XML é um formato de dados inerentemente hierárquico, e a maneira mais natural de representá-lo é com uma árvore (Python, 2017).

Para utilizar o `sARCASM_KtM.py` precisa-se de dois arquivos: O KGML, que fornece informações para análise computacional e modelagem de redes de genes/proteínas e redes químicas, e os genes ortólogos a serem modelado que podem ser obtidos submetendo seus genes de interesse ao KAAS.

O KGML deve ser obtido no site (<http://www.genome.jp/kegg/>) nele contém diversos vias. Depois de escolher a via de interesse deve-se baixar o arquivo KGML, contendo todas as informações bioquímicas do organismo.

Para obter os KO deve-se acessar o site (<http://www.genome.jp/tools/kaas/>) e submeter os genes de interesse, informar o seu e-mail e então aguardar o servidor lhe devolver os KO.

Em posse desses dois arquivos o usuário poderá então utilizar o `sARCASM_KtM.py`, utilizando como entrada os genes ortólogos obtidos no KAAS e a via de interesse obtida no KEGG no formato KGML. O *script* vai comparar os genes obtidos no KAAS com os genes da via obtidos pelo KEGG, retirar apenas as reações que fazem parte dos genes ortólogos de entrada e a partir dessas reações cria uma matriz estequiométrica no formato XLSX. XLSX é um padrão de arquivo do pacote Microsoft Office mais especificamente do excel (Microsoft, 2017), que facilita a visualização da matriz gerada além de facilitar os ajustes na etapa de curadoria.

3.1.2 Script: `sARCASM_CtS.py`

O `sARCASM_CtS.py` tem como objetivo de traduzir a matriz estequiométrica para o formato padrão da biologia de sistemas SBML (Hucka et al., 2003). Na construção deste algoritmo foram

utilizadas as seguintes bibliotecas: `openpyxl`, uma biblioteca Python para ler e gravar arquivos do Excel 2010 `xlsx` (Openpyxl, 2017), e `COBRApy` que é um pacote para modelagem baseada em restrições de redes biológicas (Ebrahim, 2013) e pelo fato de ter funções que criam arquivos no formato SBML foi adaptado para traduzir matriz estequiométrica em SBML.

Para a utilização do módulo o usuário irá precisar da matriz estequiométrica curada. Durante o processo de curadoria, o pesquisador deverá revisar as reações, metabólitos, reversibilidade e todas as referências bibliográficas disponíveis. Esse processo varia de acordo com a via metabólica e do microrganismo que está sendo analisado. Em seguida o processo de curadoria deve identificar os componentes externos que são utilizados porém não são produzidos na via a ser estudada e os componentes externos de saída que são produzidos e não são utilizados. Com todas as informações adicionais de compostos chamados externos e suas saídas inclusas na matriz estequiométrica, pode-se então utilizar como entrada na utilização do `sARCASM_CtS.py`.

3.1.3 Script: `sARCASM_FBA.py`

O `sARCASM_FBA` tem como objetivo aplicar o cálculo do FBA. Na construção deste *script* foram importadas algumas funções da biblioteca `COBRApy` (Ebrahim, 2013).

Foram utilizadas funções disponíveis na biblioteca `COBRApy`. Necessita-se da via de interesse modelada no formato SBML para ser executada, que pode ser obtida com o `sARCASM_CtS.py`.

O *script* deve mostrar na tela os resultados do FBA contendo todas os metabólitos envolvidos na função objetivo, ou seja, o caminho que o metabólito de interesse deve percorrer para obter o máximo de aproveitamento e suas respectivas taxas de produção.

4. RESULTADOS

Para analisar se a semi-automatização para a reconstrução e análise de matrizes estequiométricas obteria bons resultados, foi realizado um estudo de caso onde o objetivo seria modelar e simular um organismo modelo já publicado, a fim de se comparar os resultados obtidos em ambos os métodos.

O organismo escolhido para esse estudo foi a *Exiguobacterium antarcticum* B7, bactéria extremófila que possui propriedades de interesse biotecnológico, por exemplo, a capacidade de adaptação ao frio, que tornam este gênero um alvo para descobrir novas enzimas, como lipases e proteases, além de melhorar a compreensão dos mecanismos de adaptação e sobrevivência a baixas temperaturas (Carneiro et al., 2012). Este organismo foi mesmo utilizado no trabalho de um outro autor (Kawasaki et al., 2016), e foi escolhido para meios de comparação dos resultados, assim validando os dados deste trabalho.

O ambiente de testes foram os mesmos da criação dos *scripts* do sARCASM: desktop Modelo L4800, com processador 4x Intel(R) Core(TM) i5-4440 CPU @ 3.10GHz e 8 GB de RAM, executando Ubuntu 14.04.4 LTS (Ubuntu, 2017), kernel Linux 3.16.0-67-generic (x86_64).

4.1 Criação de matriz estequiométrica da *Exiguobacterium antarcticum* B7

Para a criação da matriz estequiométrica obteve-se o genoma da *Exiguobacterium antarcticum* B7 através do site do NCBI (2018) onde foi obtido os dados referentes a *E. antarcticum* B7 cujo o a sequência de referência do NCBI é “NC_018665.1”. O formato de arquivo do organismo disponibilizado era o formato “.fasta” facilitando a manipulação do arquivo. Mais detalhes sobre o organismo podem ser observados na tabela.

Tabela 1: Informações sobre o arquivo baixado que contém os genes a serem utilizadas na criação da Matriz estequiométrica. Fonte: O autor.

Site em que foi baixado	www.ncbi.nlm.nih.gov
Conteúdo do arquivo	2.928 genes
Organismo	<i>Exiguobacterium antarcticum</i> B7
Tamanho do arquivo	2.7 MB

Fonte: www.ncbi.nlm.nih.gov

O próximo passo foi obter os grupos ortólogos ou KO referentes ao genoma para que possamos posteriormente comparar os genes obtidos no NCBI com os genes do banco de dados do KEGG.

Para obter-se os KO foi utilizado o KAAS. Foi submetido os genes no site e como resultado ele retornou 1.497 KO dentre os 2.928 genes iniciais. Esses resultados então foram salvos em um arquivo de texto “.txt”.

A via metabólica a que foi analisada para este estudo foi a via de biossíntese de ácidos graxos pelo fato da sua capacidade de se transformar em substâncias biologicamente mais ativas, com funções especiais no equilíbrio homeostático, em componente estrutural das membranas celulares e do tecido cerebral e nervoso (Takahashi, 2018).

A biossíntese de ácidos graxos ocorre em diferentes partes da célula e requer a participação de um intermediário de três carbonos, o malonil-CoA. A formação de malonil-CoA a partir de acetil-CoA é um processo irreversível, catalisado por acetil-CoA carboxilase (Cox; Nelson.,2008).

A via usada como modelo foi obtida no site do KEGG e foi utilizada como um filtro para retirar apenas os KO pertencentes a via de biossíntese de ácidos graxos do restante dos KO do genoma da *E. antarcticum* B7, tornando assim possível estudá-la separado dos outros genes.

A Figura 13 mostra a via metabólica extraída do KEGG, os quadrados coloridos com a cor verde representam os genes específicos da *Exiguobacterium antarcticum*. O identificador para a via é o “ean00061” onde “ean” indica o organismo e “00061” indica a via metabólica.

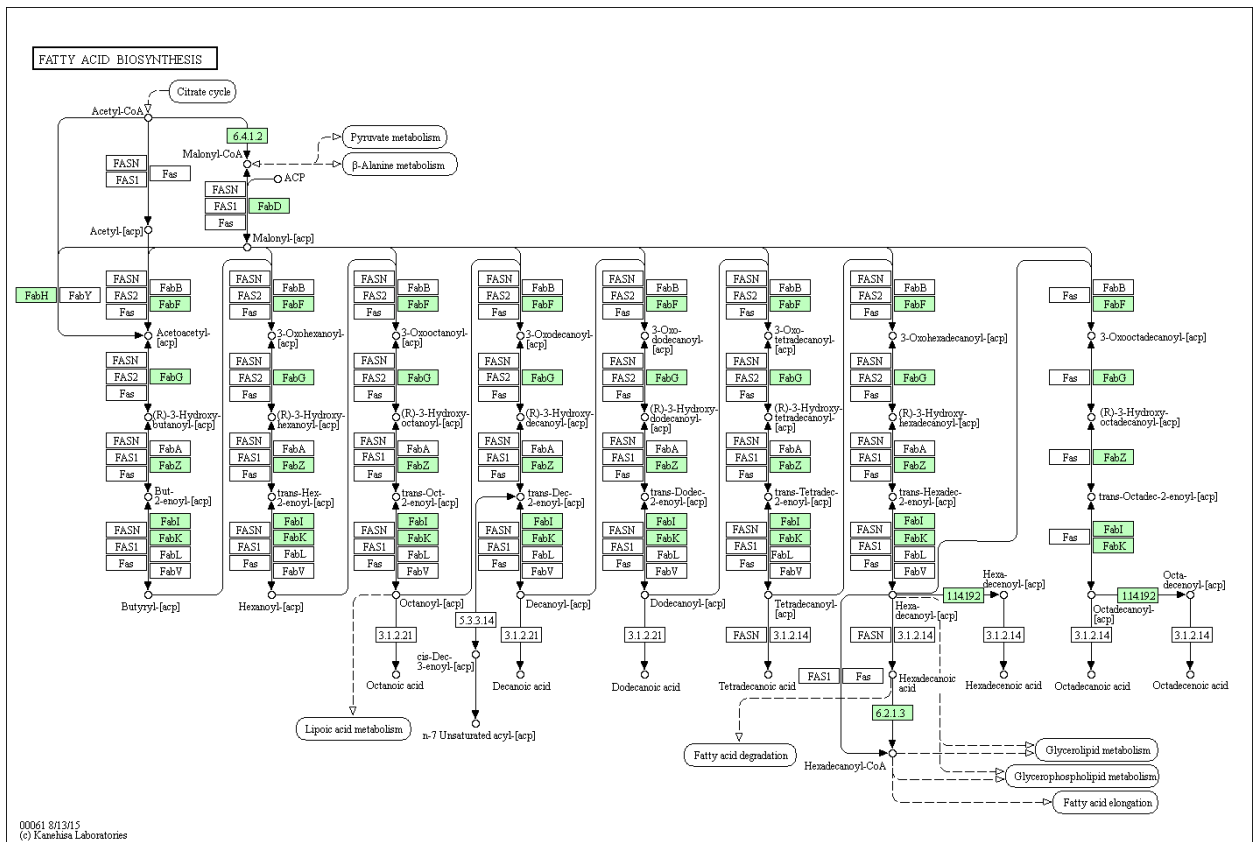


Figura 13: Via metabólica da biossíntese de ácidos graxos. Retirado de KEGG (2017)

Após baixar o KGML da via no KEGG (2017) foi utilizado então o SARCASM. O script comparou os KO tanto da via metabólica quanto do genoma e selecionou apenas os KO em comum e gerou a matriz estequiométrica.

Foram encontrados 41 reações e 44 metabólitos e a matriz foi formada com esses componentes. O sARCASM cria duas abas, a primeira contem a matriz e a segunda contém os genes relacionados a cada reação. Ao todo foram encontrados 15 genes que estão relacionados as 41 reações mostrados na Tabela 2.

Tabela 2: Genes relacionados com suas respectivas reações

Número da reação	Id. da reação	Genes
1	R07764	fabZ or faz
2	R04969	fabK or fabI or fas or fabV
3	R04966	fabK or fabI or fas or fabV
4	R04724	fabK or fabI or fas or fabV
5	R04961	fabK or fabI or fas or fabV
6	R04958	fabK or fabI or fas or fabV
7	R04955	fabK or fabI or fas or fabV
8	R04429	fabK or fabI or fas or fabV
9	R07765	fabI or fas or fabK
10	R07763	fabG or faz
11	R04543	fabG or FAS2 or FASN or faz
12	R04566	fabG or FAS2 or FASN or faz
13	R04964	fabG or FAS2 or FASN or faz
14	R04534	fabG or FAS2 or FASN or faz
15	R04536	fabG or FAS2 or FASN or faz
16	R04953	fabG or FAS2 or FASN or faz
17	R04533	fabG or FAS2 or FASN or faz
18	R04969	fabI
19	R04970	fabL or fabI or FAS1 or FASN
20	R04966	fabI
21	R04967	fabL or fabI or FAS1 or FASN
22	R04724	fabI
23	R04725	fabL or fabI or FAS1 or FASN
24	R04961	fabI
25	R04962	fabL or fabI or FAS1 or FASN
26	R04958	fabI
27	R04959	fabL or fabI or FAS1 or FASN
28	R04955	fabI
29	R04956	fabL or fabI or FAS1 or FASN
30	R04429	fabI
31	R04430	fabL or fabI or FAS1 or FASN
32	R01626	fabD or FAS1 or FASN or faz
33	R10707	fabH or fabY
34	R07762	fabF or fas or fabB
35	R04968	fabF or fabB or FAS2 or FASN or faz
36	R04726	fabF or fabB or FAS2 or FASN or faz
37	R04963	fabF or fabB or FAS2 or FASN or faz
38	R04960	fabF or fabB or FAS2 or FASN or faz
39	R04957	fabF or fabB or FAS2 or FASN or faz
40	R04952	fabF or fabB or FASN or FAS2 or faz
41	R04355	FAS2 or FASN or fas or fabB or fabF

Fonte: O autor

4.2 Geração do Arquivo no formato SBML

Após a obtenção da matriz estequiométrica foi feita a curadoria da via baseando-se no estudo de outros trabalhos publicados (Carneiro et al., 2012; Kawasaki et al., 2016) para obter-se os metabólitos externos que ajudam a equilibrar a função e tornar possível o cálculo do FBA mais a frente.

Após essa etapa foi utilizado o módulo `sARCASM_CtS.py` do *script*, que tem como função criar uma versão em formato SBML da matriz de interesses. Além de criar o modelo ainda exibe informações como: todos os metabólitos, todas as reações e suas reversibilidades e os genes com todas as reações que estão relacionado com eles.

4.3 Cálculo e análise do FBA

Os resultados obtidos não foram comparados com outras ferramentas do mesmo tipo pelo fato de ter sido utilizado módulos da biblioteca já existente para o cálculo do FBA, o COBRApy (Ebrahim, 2013). Os resultados obtidos foram comparados com o trabalho de Kawasaki (2016) que utilizou também um método semiautomático, mas apenas no cálculo do FBA.

Os resultados ficaram muito próximo do esperado e pelo fato de ser uma análise estatística não influenciam na análise do organismo em estudo, as Tabela 3 e Tabela 4 mostram a saída do `sARCASM_FBA.py` da bactéria *E. Antarticum* B7, contém as entradas e saídas da via metabólica simulada e seus objetivos e valores respectivos de cada reação.

Tabela 3: Fluxos de entrada e saída do FBA.

Fluxo de Entrada	Valores	Fluxos de saída	Valores
Acetyl-CoA;	1000	ADP;	933
Acyl-carrier...	1000	Orthophosphate;	933
H+;	1000	CoA;	600
ATP;	933	CO ₂ ;	533
HCO ₃ -;	933	H ₂ O;	533
NADH;	533	NAD+;	533
NADPH;	533	NADP+;	533
-----	-----	Malonyl-CoA;	400
-----	-----	Octadecanoyl...	66.7

Fonte: O autor

Tabela 4: Objetivos e suas devidas porcentagens obtidas no cálculo do FBA.

Objetivos	Valores	Objetivos	Valores
R1	-66.7	R4	-66.7
R10	66.7	R40	66.7
R11	66.7	R41	66.7
R12	0	R42	66.7
R13	66.7	R43	66.7
R14	66.7	R44	66.7
R15	66.7	R45	0
R16	66.7	R46	1000
R17	66.7	R47	-400
R18	66.7	R48	0
R19	66.7	R49	66.7
R2	-66.7	R5	-66.7
R20	0	R50	0
R21	-66.7	R51	933
R22	0	R52	933
R23	-66.7	R53	533
R24	0	R54	1000
R25	-66.7	R55	0
R26	0	R56	0
R27	-66.7	R57	933
R28	0	R58	933
R29	-66.7	R59	533
R3	-66.7	R6	-66.7
R30	0	R60	533
R31	-66.7	R61	533
R32	0	R62	533
R33	-66.7	R63	933
R34	533	R64	1000
R35	933	R65	600
R36	0	R66	533
R37	66.7	R7	0
R38	66.7	R8	66.7
R39	66.7	R9	-66.7

Fonte: O autor

O metabólito principal de entrada é o Acetil-CoA. Para melhorar a convergência, os limites superior e inferior foram como [0,1000] para reações irreversíveis e [-1000,1000] para reação reversível.

Os resultados foram então comparados com o trabalho de referência de Kawasaki (2016) e as Figura 14 e Figura 15 mostram todas as reações contidas nos dois trabalhos comparados, as reações

não estão na mesma ordem pelo fato de que o sARCASM cria a Matriz com suas devidas reações na ordem em que elas aparecem no KGML não levando em consideração a ordem real dos metabólitos, porem a ordem dos fatores não altera o resultado para essa análise.

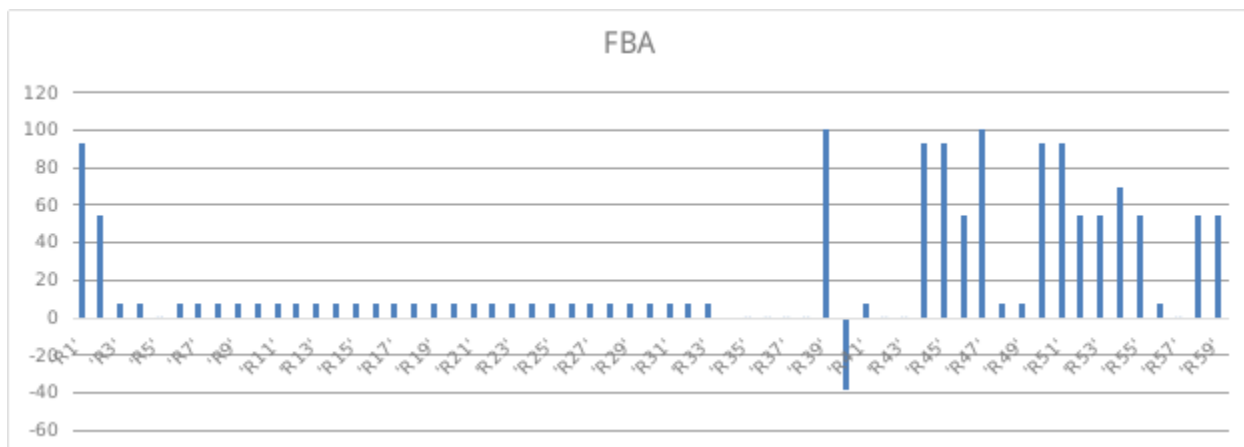


Figura 14: Resultado retirado de Kawasaki (2016) .

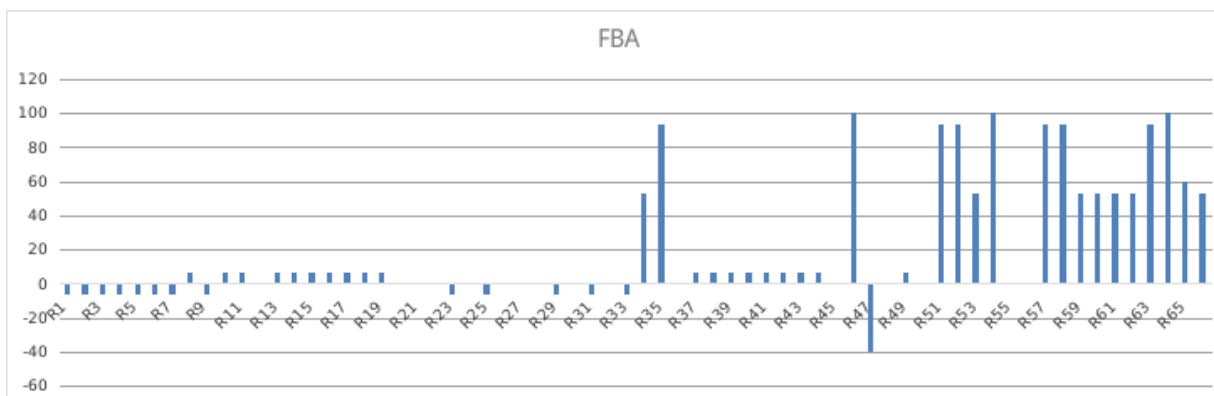


Figura 15: Resultado obtido pelo sARCASM.

Para a melhor compreensão dos resultados foi retirado apenas as reações mais relevantes nos dois trabalhos.

Observa-se que apesar dos resultados de Kawasaki (2016) serem mais discretos tem a mesma relevância pelo fato de serem estatísticos e os valores serem bem aproximados.

sARCASM	Reações	Resultados (Kawasaki, 2016)
100	Acetyl-CoA_ext => Acetyl-CoA	100
100	H_ext => H+	100
93,3	ATP + Acetyl-CoA + HCO3- => ADP + Orthophosphate + Malonyl-CoA	92,308
93,3	ATP_ext => ATP	92,307
93,3	HCO3-_ext => HCO3-	92,3078
93,3	ADP => ADP_ext	92,307
93,3	Orthophosphate => Orthophosphate_ext	92,3078

Tabela 5: Tabela de comparação de resultados de FBA.

Então esses resultados comprovam que o processo semiautomático do sARCASM é válido, poupa tempo, é confiável e de fácil manipulação dos dados antes da simulação do FBA.

Além disso, padroniza o modelo para o formato SBML tornando possível ser utilizado por outras ferramentas e também permite que se utilize um arquivo SBML de qualquer outra ferramenta para efetuar o cálculo do FBA.

Lembrando sempre que foram utilizadas ferramentas já publicadas e comprovadas tanto para a criação do modelo SBML quanto para o FBA, para garantir um bom resultado e garantir a eficácia da ferramenta.

5. CONSIDERAÇÕES FINAIS

A Biologia de sistemas vem atuando como peça fundamental para a biologia permitindo a união entre diversas ciências: matemática, física, biologia e computação.

Com a grande geração de informações por parte da biologia, como dados genômicos, transcriptômicos, proteômicos e metabolômicos, tornou-se necessário saber como interpretar esses dados e extrair informações úteis dos mesmos, sendo esta, uma das tarefas essencial na biologia de sistemas.

Neste trabalho, foi abordado um pouco mais sobre os princípios da biologia de sistemas, que estuda principalmente os genomas e suas funções interagindo dentro do mesmo ambiente, e onde busca-se entender o funcionamento completo de uma célula de um organismo podendo até mesmo prever o que pode acontecer com esse organismo em um ambiente simulado poupando tempo de bancada e insumos nos testes.

A análise e reconstrução é feita de diversas maneiras tanto manual e também por diversas outras ferramentas como o KEGG, pyFBA (Cuevas et al., 2016) e COBRApy (Ebrahim, 2013), porém não existe uma integração automática entre eles e pouca padronização.

Partindo dessas dificuldades foi desenvolvido o sARCASM, uma ferramenta semi-automática para análise e reconstrução de matrizes estequiométricas.

O método apresentado neste trabalho também foi executado na análise do genoma de um organismo, o *Exiguobacterium antarcticum* B7, utilizando um conjunto de dados contendo 2.928 genes dos quais apenas os genes relacionados a via metabólica de biossíntese de ácidos graxos e

teve seus resultados validados ao compará-los com os resultados obtidos no trabalho de Kawasaki (2016), utilizando-se o mesmo conjunto de dados.

Concluiu-se a partir dos resultados obtidos pelo algoritmo desenvolvido neste trabalho foi que:

- A ferramenta tem um tempo relativamente baixo na construção da matriz estequiométrica, aproximadamente 7 minutos para filtrar apenas os genes referentes a via de interesses no genoma e criar a matriz estequiométrica. A geração de arquivo SBML e cálculo do FBA, pelo fato de não precisar de uma conexão com internet e por fazerem pouco esforço computacional, são executados em questão de segundos.
- Apesar dos resultados não serem exatamente iguais com o trabalho de referência são muito próximos e pelo fato de serem análises estatísticas não influenciam no resultado final.
- A semi-automatização do processo facilita o processo total além de poupar tempo com a modelagem.
- Facilidade de utilizar a ferramenta em qualquer sistema operacional.

5.1 Trabalhos futuros

Muitas melhorias podem ser aplicadas à técnica desenvolvida neste trabalho. Entre elas permitir que o usuário possa usar como entrada várias vias metabólicas para que a partir delas seja feita a filtragem e geração da matriz com as diversas vias. Isso permitiria fazer a modelagem completa do organismo com todas as suas vias ou apenas as vias de interesse de maneira muito mais rápida e eficiente pois nesse cenário iriam existir muitas reações envolvidas o que tornaria muito difícil a construção manual.

Outra melhoria seria adicionar mais métodos de análises como FVA (*Flux variability analysis*) computacional utilizada para avaliar o alcance mínimo e máximo de cada fluxo de reação (Vlasov), MOMA (*method of minimization of metabolic adjustment*) que se baseia nas mesmas restrições estequiométricas que o FBA, mas relaxa a suposição de fluxo de crescimento ótimo para deleções genéticas (Sagre, 2002) e MFA (*Metabolic flux analysis*) que é uma técnica analítica importante para quantificar fluxos metabólicos intracelulares como consequência de todas as interações catalíticas e transcricionais (Lee, Park, Kim,. 2011).

REFERÊNCIAS

ALBERTS, Bruce et al. Fundamentos da biologia celular. Artmed Editora, 2006.

ALTSCHUL, Stephen F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic acids research, v. 25, n. 17, p. 3389-3402, 1997.

ANACONDA, Download Anaconda Distribution. Disponível em: <<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>>. Acesso em: 24 nov. 2017.

BEAUTIFUL SOUP, Beautiful Soup Documentation. Disponível em: <<https://www.crummy.com/software/BeautifulSoup/bs4/doc/>>. Acesso em: 24 nov. 2017.

BRUGGEMAN, Frank J.; WESTERHOFF, Hans V. The nature of systems biology. TRENDS in Microbiology, v. 15, n. 1, p. 45-50, 2007.

COX, Michael M.; NELSON, David L. Lehninger Principles of Biochemistry. 2008.

CUEVAS, Daniel A. et al. From DNA to FBA: how to build your own genome-scale metabolic model. Frontiers in microbiology, v. 7, p. 907, 2016.

DUARTE, Natalie C. et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proceedings of the National Academy of Sciences, v. 104, n. 6, p. 1777-1782, 2007.

DUBITZKY, Werner et al. Encyclopedia of systems biology. Springer Publishing Company, Incorporated, p. 2014, 2013.

EBRAHIM, Ali et al. COBRApy: CONstraints-based reconstruction and analysis for python. BMC systems biology, v. 7, n. 1, p. 74, 2013.

EBRAHIM, Ali et al. COBRApy: CONstraints-based reconstruction and analysis for python. BMC systems biology, v. 7, n. 1, p. 74, 2013.

FOGAÇA, Jennifer. "Estequiometria de reações"; Brasil Escola. Disponível em <<http://brasilecola.uol.com.br/quimica/estequiometria-reacoes.htm>>. Acesso em 30 de outubro de 2017.

GONÇALVEZ, F. S. Dogma Central da Biologia. Disponível em: <<http://www.infoescola.com/biologia/dogma-central-da-biologia/>>. Acesso em: 26 setembro, 2017.

HARRINGTON, Eoghan D.; JENSEN, Lars J.; BORK, Peer. Predicting biological networks from genomic data. FEBS letters, v. 582, n. 8, p. 1251-1258, 2008.

HORGAN, Richard P.; KENNY, Louise C. ‘Omic’ technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist*, v. 13, n. 3, p. 189-195, 2011.

HUCKA, Michael et al. The Systems Biology Markup Language (SBML): language specification for level 3 version 1 core. Release 2, 2016.

IDEKER, Trey; GALITSKI, Timothy; HOOD, Leroy. A new approach to decoding life: systems biology. *Annual review of genomics and human genetics*, v. 2, n. 1, p. 343-372, 2001.

JETBRAINS. Download PyCharm. Disponível em: <<https://www.jetbrains.com/pycharm/download/#section=linux>>. Acesso em 24 de novembro. 2017.

KANEHISA, Minoru et al. KEGG as a reference resource for gene and protein annotation. *Nucleic acids research*, v. 44, n. D1, p. D457-D462, 2016.

KANEHISA, Minoru et al. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic acids research*, v. 45, n. D1, p. D353-D361, 2017.

KANEHISA, Minoru et al. The KEGG resource for deciphering the genome. *Nucleic acids research*, v. 32, n. suppl_1, p. D277-D280, 2004.

KEGG. KEGG: Kyoto Encyclopedia of Genes and Genomes Disponível em: <<http://www.genome.jp/kegg/>>. Acesso em 12 de abril. 2017.

KEGG. KGML (KEGG Markup Language). Disponível em: <<http://www.kegg.jp/kegg/xml/>>. Acesso em 12 de abril. 2017.

KOIDE, Tie; PANG, Wyming Lee; BALIGA, Nitin S. The role of predictive modelling in rationally re-engineering biological systems. *Nature reviews. Microbiology*, v. 7, n. 4, p. 297, 2009.

LEE, Sang Yup; PARK, Jong Myoung; KIM, Tae Yong. Application of metabolic flux analysis in metabolic engineering. In: *Methods in enzymology*. Academic Press, 2011. p. 67-93.

LOOIJEN, Rick C. Holism and reductionism in biology and ecology: the mutual dependence of higher and lower level research programmes. Springer Science & Business Media, 2012.

MÉDIGUE, Claudine; MOSZER, Ivan. Annotation, comparison and databases for hundreds of bacterial genomes. *Research in microbiology*, v. 158, n. 10, p. 724-736, 2007.

MICROSOFT. Excel 2016. Disponível em: <<https://products.office.com/pt-br/excel>>. Acesso em: 02 dez. 2017.

MORIYA, Yuki et al. KAAS: KEGG automatic annotation server. *Genome Informatics*, v. 5, p. 2005, 2005.

NCBI, National Center for Biotechnology Information Search database. Disponível em: <<https://www.ncbi.nlm.nih.gov/>>. Acesso em: 16 jan. 2018.

NELSON, D. L.; Cox, M. M. Princípios de Bioquímica de Lehninger. 5o ed. Porto Alegre: Artmed, p. 72-81, 2011.

NICHOLSON, Donald E. IUBMB-Nicholson metabolic pathways charts. Biochemistry and Molecular Biology Education, v. 29, n. 2, p. 42-44, 2001.

NUMPY, NumPy. Disponível em: <<http://www.numpy.org/>>. Acesso em: 24 nov. 2017.

OPENPYXL, Openpyxl - A Python library to read/write Excel 2010 xlsx/xlsm files. Disponível em: <<https://openpyxl.readthedocs.io/en/default/>>. Acesso em: 24 nov. 2017.

ORTH, J. D., THIELE, I., & PALSSON, B. Ø. What is flux balance analysis? Nature Biotechnology, 28(3), 245–248, 2010.

PALSSON, Bernhard. Representing Reconstructed Networks Mathematically. 2003.

PALSSON, Bernhard. Systems biology: Properties reconstructed networks. 2006.

PALSSON, Bernhard; PALSSON, Bernhard Ø. Systems Biology: Constraint-based Reconstruction and Analysis. Cambridge university press, 2015.

PAPIN, Jason A. et al. Metabolic pathways in the post-genome era. Trends in biochemical sciences, v. 28, n. 5, p. 250-258, 2003.

PUJOL, Albert et al. Unveiling the role of network and systems biology in drug discovery. Trends in pharmacological sciences, v. 31, n. 3, p. 115-123, 2010.

PYTHON, The ElementTree XML API. Disponível em: <<https://docs.python.org/2/library/xml.etree.elementtree.html>>. Acesso em: 24 nov. 2017.

PYTHON. Python 3.5.2. Disponível em: < <https://www.python.org/downloads/release/python-352/>>. Acesso em 24 de novembro. 2017.

SAUER, Uwe; HEINEMANN, Matthias; ZAMBONI, Nicola. Getting closer to the whole picture. Science(Washington), v. 316, n. 5824, p. 550-551, 2007.

SBML, .org. The Systems Biology Markup Language: Basic Introduction to SBML, 2012. Disponível em: <http://sbml.org/Basic_Introduction_to_SBML>. Acesso em: 05 dez. 2017.

SBML.org disponível em:<http://sbml.org/Basic_Introduction_to_SBML>. Acesso em 22 de Maio de 2017

SCHILLING, Christophe H.; LETSCHER, David; PALSSON, Bernhard Ø. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. Journal of theoretical biology, v. 203, n. 3, p. 229-248, 2000.

SCHUSTER, Stefan; FELL, David A.; DANDEKAR, Thomas. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature biotechnology*, v. 18, n. 3, p. 326, 2000.

SEGRE, Daniel; VITKUP, Dennis; CHURCH, George M. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, v. 99, n. 23, p. 15112-15117, 2002.

SZALLASI, Zoltan; STELLING, Jörg; PERIWAL, Vipul. *System modeling in cellular biology. From Concepts to*, 2006.

TAKAHASHI, N.S. Importância dos ácidos graxos essenciais. 2007. Artigo em Hypertexto. Disponível em: <http://www.infobibos.com/Artigos/2007_3/acidosgraxos/index.htm>. Acesso em: 27 fev. 2018.

THIELE, Ines; PALSSON, Bernhard Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, v. 5, n. 1, p. 93, 2010.

VLASOV, Vanja. Flux Variability analysis (FVA).

VOET, Donald et al. *Fundamentals of biochemistry: life at the molecular level*. 2013. UBUNTU, Ubuntu 14.04.5 LTS. Disponível em: <<http://releases.ubuntu.com/14.04/>>. Acesso em: 24 nov. 2017.

WATSON, James D. et al. *Biologia molecular do gene*. Artmed Editora, p. 33-41 2015.

WRZODEK, Clemens et al. Precise generation of systems biology models from KEGG pathways. *BMC systems biology*, v. 7, n. 1, p. 15, 2013.