



UNIVERSIDADE FEDERAL DO PARÁ  
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS  
FACULDADE DE COMPUTAÇÃO  
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

WENDEL RENAN MACEDO DOS SANTOS

**MODELAGEM DA VIA DE REGULAÇÃO GÊNICA DE ÁCIDOS GRAXOS DA  
BACTÉRIA *Escherichia coli* K12 UTILIZANDO REDES BOOLEANAS**

BELÉM  
2018

WENDEL RENAN MACEDO DOS SANTOS

**MODELAGEM DA VIA DE REGULAÇÃO GÊNICA DE ÁCIDOS GRAXOS DA  
BACTÉRIA *Escherichia coli* K12 UTILIZANDO REDES BOOLEANAS**

Trabalho de Conclusão de Curso  
apresentado a Universidade Federal do  
Pará (UFPA), para obtenção do grau de  
Bacharel em Ciência da Computação.

Orientadora: Profa. Dra. Regiane Silva  
Kawasaki Francês

BELÉM  
2018

Wendel Renan Macedo dos Santos

**MODELAGEM DA VIA DE REGULAÇÃO GÊNICA DE ÁCIDOS GRAXOS DA  
BACTÉRIA *Escherichia coli* K12 UTILIZANDO REDES BOOLEANAS**

Trabalho de Conclusão de Curso  
apresentado a Universidade Federal do  
Pará (UFPA), para obtenção do grau de  
Bacharel em Ciência da Computação.

Orientadora: Profa. Dra. Regiane Silva  
Kawasaki Francês

Data da defesa: Belém-PA, 1º de Março de 2018

Conceito: Excelente

---

**Profa. Dra. Regiane Silva Kawasaki Francês** – Orientadora  
Faculdade de Computação – ICEN/UFPA

---

**Profa. Dra. Danielle Costa Carrara Couto** – Membro  
Faculdade de Geoprocessamento – CAMPUS ANANINDEUA/UFPA

---

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

*Dedico este trabalho à minha mãe, ao meu avô Rodrigo, ao meu padrasto Erlando, às minhas tias Mara, Marléa e Telma, e ao meu tio Pedro que sempre apoiaram meus estudos.*

## **AGRADECIMENTOS**

À minha orientadora professora Dr<sup>a</sup> Regiane Kawasaki pela excelente orientação e pela oportunidade que me foi concedida de fazer parte do Laboratório de Bioinformática e Computação de Alto Desempenho (LaBioCAD). Agradeço pela paciência e por me apresentar esse incrível campo de pesquisa chamado Biologia de Sistemas.

Aos membros do LaBioCAD Msc. Alex Raniere e Msc. Renato Oliveira por reservarem parte do tempo em suas pesquisas para ministrar minicursos aos integrantes do laboratório nos ensinando conceitos fundamentais da bioinformática e da biologia para que pudéssemos ter um nivelamento e uma base mais sólida no assunto, e também, agradeço a todos os membros do laboratório que contribuíram direta ou indiretamente para este trabalho.

Aos professores da Faculdade de Computação da UFPA pelas lições técnicas e de vida que aprendi durante esses anos de graduação.

Aos amigos que fiz na graduação Angelo Oliveira, Cleidyr Lima, Fábio Malcher, Leonardo Formento, Marissa Brasil, Michell Cruz, Matheus Freire, Nicole Souza, Raíssa Lorena, Ramiro Zahlouth, Ronald Assunção e Yuri Souza. Agradeço pela ajuda e pelo companheirismo durante toda a graduação..

Aos meus amigos Daniel Nascimento, Jéssika Costa, Raquel Souza, Renato Andrade e Pablo Pantoja por me acompanharem nessa caminhada e me apoiarem em momentos difíceis me dando forças para continuar – muito obrigado.

À minha namorada Danniela pela força, paciência e compreensão durante a maratona de estudos e trabalhos que me ocuparam durante vários finais de semana – muito obrigado.

*“Em algum lugar, alguma coisa incrível está esperando para ser descoberta.”*

*(Carl Sagan)*

## RESUMO

Biologia de sistemas é a ciência que descreve o comportamento de um sistema por meio da modelagem matemática e computacional. É uma área multidisciplinar e recente na bioinformática. Utilizando uma abordagem holística, ela busca entender a complexidade biológica dos seres vivos. Em um nível biológico, podemos representar nosso organismo como um conjunto de redes metabólicas integradas que formam o todo. A biologia de sistemas estuda essas redes metabólicas para entender o comportamento do organismo e propor hipóteses, modelos computacionais e contribuir com o estudo das funções biológicas. Neste trabalho, é estudado a rede de regulação da expressão gênica de ácidos graxos do organismo modelo *Escherichia coli* k12. Regulação da expressão gênica é um mecanismo de controle celular que garante que os genes em um organismo sejam expressos de acordo com a necessidade, governando a capacidade do indivíduo de responder a estímulos e se adaptar ao ambiente. Podemos simplificar esse processo afirmando que a regulação da expressão gênica controla a ativação ou repressão dos genes. A partir desse princípio, foi detectada uma similaridade entre o processo de regulação gênica e uma técnica computacional de modelagem chamada Redes Booleanas. Estas redes trabalham com os valores binários 0 e 1. Na modelagem booleana utilizamos as operações básicas AND, OR e NOT para representar as transições que ocorrem no modelo biológico. Essas operações agrupadas formam as redes booleanas. Neste trabalho utilizamos este princípio para modelar a via metabólica do organismo modelo, partindo de informações contidas no genoma disponíveis em bancos de dados biológicos. Os resultados obtidos apresentam as relações entre os genes da via dos ácidos graxos de acordo com a base de dados KEGG na forma de redes booleanas.

**Palavras-chave:** redes booleanas, regulação booleana, análise de rede booleana

## **ABSTRACT**

Systems Biology is the science that describes the behavior of a system through mathematical and computational modeling. It is a multidisciplinary and recent area in bioinformatics. Using a holistic approach, it seeks to understand the biological complexity of living things. On a biological level, we can represent our organism as a set of integrated metabolic networks that form the whole. Systems biology studies these metabolic networks to understand the behavior of the organism and to propose hypotheses, computational models and contribute to the study of biological functions. In this work, the regulation network of the gene expression of fatty acids of the model organism *Escherichia coli* k12 is studied. Gene expression regulation is a mechanism of cellular control that ensures that genes in an organism are expressed according to need, governing the individual's ability to respond to stimuli and adapt to the environment. We can simplify this process by stating that the regulation of gene expression controls the activation or repression of genes. From this principle, a similarity was detected between the process of gene regulation and a computational modeling technique called Boolean Networks. These networks work with binary values 0 and 1. In Boolean modeling we use the basic operations AND, OR and NOT to represent the transitions that occur in the biological model. These grouped operations form Boolean networks. In this work we use this principle to model the metabolic pathway of the model organism, based on information contained in the genome available in biological databases. The results obtained show the relationships between the genes of the fatty acid pathway according to the KEGG database in the form of Boolean networks.

## LISTA DE ABREVIATURAS E SIGLAS

BN – Rede Booleana (do inglês, *Boolean Network* )

DNA – Ácido Desoxirribonucleico.

*E. coli* – *Escherichia coli*

EDO – Equações Diferenciais Ordinárias.

KEGG – Kyoto Encyclopedia of Genes and Genomes.

MIT – Instituto Tecnológico de Massachusetts (do inglês, Massachusetts Institute of Technology)

mRNA – RNA mensageiro.

PBN – Rede Booleana Probabilística (do inglês, *Probabilistic Boolean Network*)

rFBA – Análise de Balanço de Fluxo regulatório (do inglês, *regulatory Flux Balance Analysis*)

RNA – Ácido Ribonucleico.

## LISTA DE FIGURAS

FIGURA 1 - CÉLULA EUCARIOTA X CÉLULA PROCARIOTA	21
FIGURA 2 - ORGANIZAÇÃO DOS OPERONS	24
FIGURA 3 - ESQUEMA DA RELAÇÃO ENTRE OS GENES V1, V2 E V3.	29
FIGURA 4 - ESTADOS REPRESENTANDO AS TRANSIÇÕES E OS ATTRACTORS	31
FIGURA 5 - EXEMPLO DE UMA PBN	32
FIGURA 6 - EXEMPLO DE MODELAGEM MODULAR	34
FIGURA 7 - REDE DE PETRI	35
FIGURA 8 - COMPARAÇÃO ENTRE AS REDES	36
FIGURA 9 - FLUXOGRAMA DA METODOLOGIA ABORDADA	37
FIGURA 10- VIA DE REGULAÇÃO DA BIOSÍNTESE DE ÁCIDO GRAXO	46
FIGURA 11 - REDE DOS ÁCIDOS GRAXOS REPRESENTADA PELOS GENES ATIVOS	47
FIGURA 12 - REDE BOOLEANA	48
FIGURA 13 - MAPA DA REDE NO SOFTWARE BOOLESIM	49
FIGURA 14 - EXECUÇÃO NO BOOLESIM COM TODOS OS GENES RECEBENDO VALOR 1	50
FIGURA 15 - MAIORIA DOS GENES INIBIDORES DESATIVADOS	51
FIGURA 16 - SIMULAÇÃO COM OS INIBIDORES DESATIVADOS	52
FIGURA 17: REGRAS BOOLEANAS E SUAS APRESENTAÇÕES	53
FIGURA 18 – PORTAS PARA CADA REGRA DOS GENES	53
FIGURA 19: REPRESENTAÇÃO DAS ENTRADAS LÓGICAS	51

## LISTA DE TABELAS

TABELA 1 - TRANSIÇÃO DE ESTADOS	29
TABELA 2 - RELAÇÃO ENTRE O REPRESSOR E UM GENE	47

## SUMÁRIO

1 INTRODUÇÃO .....	13
1.1 CONTEXTO .....	13
1.2 BANCO DE DADOS .....	14
1.3 MOTIVAÇÃO .....	16
1.5 OBJETIVOS .....	17
1.6 METODOLOGIA .....	19
2 REFERENCIAL TEÓRICO.....	19
2.1 EXPRESSÃO GÊNICA .....	22
2.2 BIOLOGIA SINTÉTICA .....	35
3 MATERIAIS E MÉTODOS .....	41
3.1 ELABORAÇÃO DAS REDES BOOLEANAS.....	42
4 RESULTADOS E DISCUSSÃO .....	47
4.2 DEFININDO PORTAS LÓGICOS A PARTIR DAS REGRAS BOOLEANAS .....	40
5 CONCLUSÃO .....	55
6 CONSIDERAÇÕES FINAIS .....	55
REFERÊNCIAS .....	57

# 1 INTRODUÇÃO

## 1.1 CONTEXTO

O desenvolvimento tecnológico tem contribuído para o avanço da bioinformática e o surgimento de novas técnicas computacionais como os sequenciadores de nova geração e software de visualização mais robustos, esse desenvolvimento vêm gerando uma considerável quantidade de dados ômicos, levando ao desenvolvimento de diversas áreas relacionadas à Biologia Computacional.

A Biologia de Sistemas, ciência que descreve o comportamento de um sistema por meio de modelagem matemática e computacional, busca interpretar os dados que foram gerados pelas tecnologias de sequenciamento, e, por meio deles, elucidar os mecanismos que realizam a integração dos sistemas biológicos para melhor entendê-los e também descobrir novos dados e modelos biológicos (RIGOUSTSOS; STEPHANOPOULOS, 2007). A integração de sistemas biológicos já foi estudada anteriormente, porém não havia tecnologia suficiente para detalhar esse estudo e entender essa integração (VEMURI, NIELSEN, 2002).

A abordagem da Biologia de Sistemas diferencia do padrão adotado pelas técnicas reducionistas que ganharam força após a publicação do dogma central da biologia por Crick (1970). Há algumas décadas, essa técnica utiliza experimentos com foco no estudo de pequenas moléculas como genes, algumas macromoléculas e pequenos grupos de metabólitos. O foco da Biologia de Sistemas não é somente para o entendimento dos genes e proteínas, mas também para toda a estrutura e dinâmica do sistema (KITANO, 2002).

Segundo RIGOUSTSOS e STEPHANOPOULOS (2007), Biologia de sistemas descreve o campo de pesquisa que trabalha com uma visão geral com o objetivo de melhor compreender e destacar os mecanismos e processos biológicos. Diante do exposto, eles definem Biologia de Sistemas como uma integração de conhecimentos que irão explicar as conexões entre diferentes moléculas dentro de processos e sistemas celulares.

A Biologia de Sistemas possui diferentes abordagens para modelar um sistema biológico, uma delas é a modelagem booleana. Essa modelagem é aplicada no conceito biológico de rede de regulação gênica, e foi um conceito introduzido e

aplicado por Kauffman (1971) e tem como objetivo representar as interações em redes de regulação gênica por meio de grafos de tal forma que seja possível determinar, via regras booleanas, se um gene está ativo ou inativo. A análise booleana é abstrata, ou seja, é uma abordagem de alto nível que busca a simplicidade nas representações.

A aplicação da modelagem das redes booleanas é um método para análise de vias metabólicas que não substitui em hipótese alguma os testes feitos em laboratório, pois ela utiliza como parâmetro os resultados obtidos *in vivo*. Com esses resultados em mãos, é possível construir e simular cenários utilizando as operações booleanas e também utilizar técnicas de inferências gênicas (FACHINI, 2016).

Utilizando métodos computacionais, conceitos biológicos e modelos matemáticos, a regulação gênica por essa técnica booleana tem sido usada não somente com os organismos procaríotos que têm uma regulação baseada nos *operons* (estruturas controladoras da expressão gênica em procaríotos) como também nos eucariotos onde a regulação é mais complexa (SAADATPOUR, 2013).

Entre as diversas vias possíveis para realizar o estudo, foi escolhida via de regulação gênica dos ácidos graxos que são componentes lipídicos celulares e se localizam na membrana celular (WELCH, 1991). Eles têm participação importante no metabolismo dos lipídios, pois grande parte desses lipídios possuem ácidos graxos em sua estrutura. Entre as funções dos lipídios estão a comunicação celular, interação enzimática, reserva de energia, além de participarem de reações importantes para produzir ações fisiológicas nos seres vivos (BROADHURST et al., 2002). Devido sua importância, foi escolhida a via de regulação dos ácidos graxos para realizar a modelagem neste trabalho para compreender o processo de regulação gênica dessa via e colaborar com os estudos.

Baseado nisto, este trabalho propõe utilizar a modelagem booleana para descrever os genes que participam do metabolismo da via de regulação da expressão gênica de ácidos graxos do micro organismo modelo *Escherichia coli* k12,

## 1.2 BANCO DE DADOS

Segundo LAL et al. (2017), os bancos de dados biológicos consistem de informações e conhecimentos coletados de pesquisas científicas, publicações e

análises estatísticas de diferentes tipos de dados, ficando espalhados em diferentes continentes e mantidos por várias organizações.

Bancos de dados biológicos têm sido criados para armazenar os dados gerados pelas tecnologias de sequenciamento, proteínas e as informações referentes à elas, estruturas moleculares, taxonomia e bibliografia da área biológica. Essas informações podem ser usadas para referenciar conjuntos de dados e também para a reconstrução de vias metabólicas, entre outras redes moleculares (KANEHISA et. al., 2015).

### 1.2.1 Kyoto Encyclopedia of Genes and Genomes (KEGG)

A Enciclopédia de Genes e Genomas de Kyoto (do inglês, Kyoto Encyclopedia of Genes and Genomes - KEGG) é uma base de conhecimentos para análises sistemáticas de funções gênicas em termos de redes de genes e moléculas. Além disso, a base de dados KEGG mantém o banco de dados de genes para o catálogo de genes de todos os organismos com genomas completos e alguns selecionados com genomas parciais que são continuamente anotados (OGATA et al., 1999).

Existe uma base de dados no KEGG chamada LIGAND. Esta base de dados contém os conhecimentos obtidos no universo das substâncias químicas e reações que são relevantes para a vida. A base de dados LIGAND possui compostos, reações, enzimas, entre outros dados necessários para uma melhor análise do organismo (OGATA et al., 1999).

O KEGG oferece, também, ferramentas gráficas em Java para apresentar, comparar e manipular expressões. Além disso, ele possui ferramentas computacionais para comparação de sequências, grafos e computação de vias metabólicas (KANEHISA; GOTO, 2000).

### 1.2.2 REGULON DB

O banco de dados REGULON DB organiza uma grande quantidade de dados de um dos organismos mais estudado na bioinformática, a *Escherichia coli* k-12. Devido a isso, ele é uma das ferramentas mais importantes nas pesquisas com regulação gênica bacteriana. Ele oferece resumos sobre fatores de transcrição da *E. coli* k-12. Esses textos envolvem várias informações como a estrutura, mecanismos e propriedades fisiológicas sendo atualizados frequentemente. (CRUZ ET. AL., 2017).

O REGULON DB oferece informações sobre a organização dos operons (estruturas controladoras da regulação gênica em procariotos) e suas decomposições em unidades de transcrição. Esse banco de dados possui uma interface de busca simples onde permite especificar o tipo de dados como gene, operon, regulon, sigmulon, sRNA e condição de crescimento (REGULON DB website, 2017). Esse banco de dados foi utilizado neste trabalho para coletar informações de cada um dos genes e seus respectivos ativadores e inibidores. Além disso, ele foi utilizado para comparar os dados presentes no KEGG e confirmar os genes presentes na via metabólica estudada.

### 1.2.3 UniProt (Universal Protein Resource)

Segundo informações do trabalho do Laboratório Europeu de Biologia Molecular (do inglês, *European Molecular Biology Laboratory - EMBL*), a base de dados universal de proteínas (do inglês, Universal Protein Resource - UniProt) (UniProt website, 2017) disponibiliza dados estáveis, compreensíveis e de acesso livre sobre sequências de proteínas e anotação funcional. Ele é um banco de dados colaborativo entre o Instituto Europeu de Bioinformática (do inglês, *European Bioinformatics Institute - EBI*), Recurso de informação Proteica (do inglês, *Protein Information Resource - PIR*) e o Instituto de Bioinformática Suíço (do inglês, *Swiss Institute of Bioinformatics - SIB*).

Ainda de acordo com a pesquisa citada no parágrafo anterior, entre as principais atividades desse centro de pesquisa estão as curadorias manuais de sequências de proteínas acompanhada de análises computacionais, arquivamento de sequências, e também cuidados com o website para que ele tenha uma interface amigável. De modo geral, esse banco de dados possui uma base de conhecimentos, clusters de referências, arquivos e uma unidade para a metagenômica.

Neste trabalho, o UniProt foi usado para obter uma referência entre o gene e as proteínas com o objetivo de entender o que seria afetado na ativação ou inativação de um determinado gene.

## 1.3 MOTIVAÇÃO E JUSTIFICATIVA

O desenvolvimento deste trabalho é motivado pelos desafios da interdisciplinaridade que envolve o mesmo e o sincronismo que deve existir entre

os conhecimentos de biologia, química, matemática e computação. Além disso, ele é motivado pela possibilidade de alcançar bons resultados para o estudo da regulação gênica em procariontes por meio da modelagem booleana, que tem como principais características, a simplicidade em sua representação e abstração dos conceitos biológicos. Além disso, este trabalho é motivado pela possibilidade de continuação dos estudos nessa área na UFPa e aprimoramento dessa técnica de modelagem para utilização em futuros trabalhos.

#### 1.4 OBJETIVOS

- Objetivo Geral

Desenvolver um modelo booleano para descrever o comportamento da via de biossíntese de ácidos graxos da *E. coli* k-12 capaz de representar de forma fidedigna a relação entre os genes presentes nesta via.

- Objetivos Específicos

- Identificar as bases de dados para extrair informações sobre a via;
- Identificar os genes presentes na via metabólica de ácidos graxos no KEGG;
- Extrair os genes da via metabólica de ácidos graxos.
- Estudar as relações entre os genes presentes na via.
- Definir as regras booleanas entre os genes;
- Elaborar as tabelas-verdade;
- Construir a rede de regulação;
- Analisar o modelo utilizando informações dos bancos de dados;
- Validar o modelo comparando com as informações extraídas do banco de dados KEGG e REGULONDB.

## 1.5 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em 5 seções:

**INTRODUÇÃO:** Apresenta o trabalho com uma ideia geral do conteúdo abordado assim como a motivação, justificativa e objetivos geral e específico.

**REFERENCIAL TEÓRICO:** Apresenta os conceitos e técnicas utilizadas para o desenvolvimento deste trabalho.

**MATERIAIS E MÉTODOS:** Apresenta todas as etapas necessárias para realizar a modelagem booleana proposta neste trabalho como a análise do organismo, coleta dos dados, definição das regras e relação entre os genes.

**RESULTADOS:** Apresenta a relação gênica da via de expressão gênica extraída do KEGG reduzida à relações booleanas, sua tabela-verdade, a representação lógica e a discussão sobre o assunto.

**CONSIDERAÇÕES FINAIS:** Apresenta considerações sobre o trabalho e trabalhos futuros.

## 2 REFERENCIAL TEÓRICO

### 2.1 CONCEITOS BIOLÓGICOS

Bonato e Nakaya (2015) descrevem que a Biologia de Sistemas tem como objetivo elucidar como comportamentos complexos de sistemas biológicos emergem de propriedades dos componentes e interações nos sistemas. São utilizadas técnicas computacionais e experimentais para que seja possível realizar a análise de um componente e identificar sua função no sistema como um todo. Eles afirmam ainda que a capacidade de análise de dados é muito grande, gerando um alto *throughput* e, em consequência disso, o número de análises das atividades que podem ser feitas no organismo também aumenta devido à alta atividade dos diversos componentes da célula.

O estudo das “ciências ômicas” como genômica, proteoma, metabolômica, interactoma, entre outras, tem proporcionado uma análise mais específica de partes do organismo, por exemplo, a detecção de mRNA (RNA mensageiro) através das técnicas de microarray ou RNA-seq (BONATO E NAKAYA, 2015). Ainda nos estudos destes autores, outro exemplo pode ser encontrado na genômica com o desenvolvimento dos sequenciadores que, inclusive, contribuíram para as outras ômicas. Essa combinação nos avanços do estudo da biologia molecular junto com as tecnologias para o processamento dos dados gera uma rica fonte de entendimento dos mecanismos dos processos biológicos. No entanto, é necessário realizar um processo de modelagem nesses dados para que seja gerada uma hipótese testável.

Bonato e Nakaya (2015) avaliam dois tipos de abordagem na biologia de sistemas: *top-down* e *botton-up*. A primeira abordagem parte do todo para o específico. As análises e modelagens são feitas com o processamento dos dados ômicos na tentativa de decifrar as camadas que compõem o sistema e informações biológicas sobre o processo biológico. A abordagem *botton-up* modela como componentes interagem para obter conhecimento sobre os processos biológicos. Esses componentes são de uma escala menor para uma maior. Neste caso, análise é feita a partir dos genes, proteínas e metabólitos a fim de entender a participação de cada um nos processos biológicos.

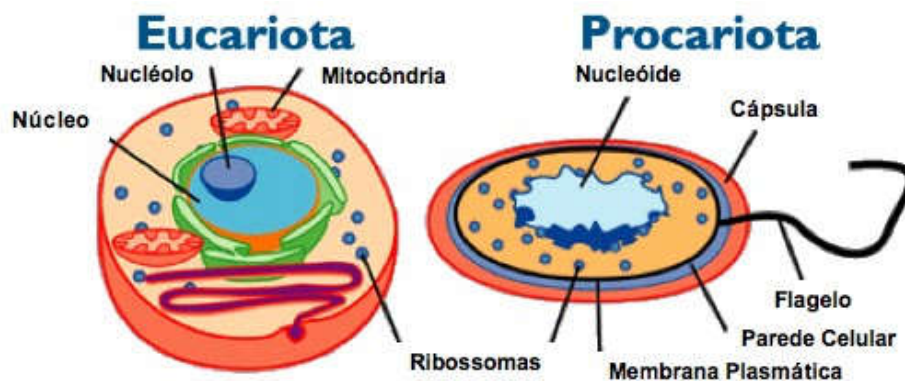
Para finalizar as ideias de Bonato e Nakaya (2015), é discutido que este tipo de abordagem começa com hipóteses que são geradas em laboratórios através de observações do organismo estudado. O modelo dinâmico contínuo, um dos métodos

mais utilizados na abordagem *botton-up*, necessita de alguns parâmetros chaves como a síntese e a taxa de degradação do organismo, o que torna esse método bem prático para modelos com dez ou menos componentes, o que é considerado pouco, porém eficiente. No modelo dinâmico discreto como as redes booleanas, modelos de dados multi-valorados e redes de Petri é possível determinar descrições qualitativas dos dados do sistema biológico.

### 2.1.1 BIOLOGIA CELULAR

A célula é a menor e mais organizada forma de vida básica na qual todos os organismos são construídos, assim a célula pode ser chamada de “bloco de construção da vida” (LIIVING, et. al. 2011). De uma forma geral, a célula pode ser dividida em dois tipos, eucariotes e procariotes, sendo diferenciadas pela presença ou não da carioteca envolvendo o conteúdo genético. As células são organizadas em compartimentos, onde dentro deles existem as organelas responsáveis por suas atividades básicas.

FIGURA 1: CÉLULA EUCARIOTA X CÉLULA PROCARIOTA



FONTE: <http://wikiciencias.casadasciencias.org/wiki/index.php/Procariote>, 2010

Neste trabalho, será analisado um organismo procarioto, a bactéria *Escherichia coli*. Apesar de milhares de células pertencerem a um mesmo organismo, elas não realizam a mesma função. As células são diferenciadas para atuarem em várias funções vitais e possuem diversas reações.

O metabolismo é um conjunto de reações que juntas produzem ou degradam substâncias para regular as funções vitais do organismo (LIIVING et. al., 2011). Ele é dividido em duas categorias: anabolismo e catabolismo. O anabolismo é responsável pelo crescimento, regeneração e manutenção dos tecidos e órgãos presentes no organismo. Para isso, ele demanda substratos e energia para usar em suas reações. Já o catabolismo é usado para a obtenção de energia através da utilização das reservas liberando aminoácido e glicose para serem consumidos.

Todas essas reações dos organismos são organizadas em vias metabólicas onde podemos encontrar reações, genes, proteínas, entre outros, sendo transformados através de enzimas atuando na regulação do ser vivo. Uma única via metabólica é formada por muitas reações ocorrendo de forma dinâmica para o balanceamento de energia, controle da expressão gênica, entre outros fatores.

Como as reações químicas são relativamente lentas, em alguns casos, as enzimas são acionadas para acelerar o processo. As enzimas são proteínas com propriedade de acelerar reações químicas sem serem consumidas. Elas têm importância fundamental no processo metabólico. As enzimas que provocam diminuição na atividade ou na produção de outra enzima são chamadas de inibidoras. Por outro lado, aquelas que contribuem para a produção de uma determinada substância são chamadas de ativadoras.

Uma das substâncias sintetizadas no citosol (compartimento celular) nas vias metabólicas é o ácido graxo. O ácido graxo é um dos maiores componentes dos triglicerídeos que é uma forma de lipídio utilizado para armazenar energia. Eles são importantes, pois triglicerídeo é a segunda fonte de energia em caso de não haver açúcar disponível.

Um dos dogmas centrais da biologia afirma que o DNA é transcrito em RNA e traduzido para proteína (CRICK, 1970). Esse processo ocorre quando há a expressão gênica, por exemplo. Com o objetivo principal de economizar energia, os organismos não expressam todos os genes ao mesmo tempo. A expressão gênica é o processo pelo qual a informação genética será sintetizada para formar produtos (HALL, 2017). Hall também afirma que esses produtos são normalmente as proteínas que irão executar funções vitais como a ativação ou inibição de um gene formador de outra proteína, por exemplo. Existem também, os genes que não codificam para proteínas, mas que atuam de forma significativa nos processos regulatórios e pós-transcricionais. Para o controle da expressão gênica as células executam mecanismos

de controle para determinar o tempo e a quantidade de material genético que será transcrito.

### 2.1.2 EXPRESSÃO GÊNICA

O DNA dos seres vivos é formado por milhares de nucleotídeos. Esse DNA possui regiões que podem ser expressas (ou não) resultando em atividades específicas no processo de regulação dos genes. Essa regulação ocorre devido ao processo de transcrição do DNA em RNA e sua tradução em proteína (ALBERTS et. al., 2010).

Nos procariotos a regulação gênica ocorre por meio dos *operons* (Figura 2) que são estruturas reguladoras presentes na fita do DNA de organismos procariotos. Eles são basicamente divididos em três partes: região controladora, gênica/codificadora e terminadora.

Osborn e Field (2009) definem o operon em procariotos como uma região genética dividida em região controladora formada por um promotor, um operador, uma região gênica onde estão localizados os nucleotídeos que serão transcritos, e um terminador. O promotor é a região onde a RNA polimerase se liga para percorrer o DNA e realizar a transcrição. O operador funciona como um *plug* de liga e desliga, uma espécie de interruptor genético, pois dependendo da substância que está ligada a ele é possível que ocorra ou não a transcrição.

Berg et.al descreve o processo de funcionamento do operon. Ele explica que quando uma proteína inibidora se liga ao operador, ela impede a passagem da mRNA polimerase, logo não ocorre a expressão desse gene. Caso contrário, quando um ativador se liga nessa região, ele muda o formato da proteína bloqueante e promove o desligamento dela da região, deixando assim, o caminho livre para que a polimerase percorra os nucleotídeos realizando a expressão dos genes. A região gênica contém os genes que serão expressos.

Normalmente, nos procariotos, essa região contém mais de um gene e eles possuem funções semelhantes atuando em conjunto em determinado processo, ou seja, irão produzir proteínas de uma mesma via metabólica. E, por último, uma sequência terminadora indicando que naquele ponto encerra a transcrição.

FIGURA 2: ORGANIZAÇÃO DOS OPERONS



**FONTE:** Regulação da expressão gênica em procaríotos, Sérgio Araújo, 2013

a) **Região controladora:** formada pelo promotor e pelo operador. O promotor é o local onde a polimerase irá se ligar para promover o processo de transcrição. O operador funciona como uma chave liga/desliga. Ele apresenta dois estados, liberado ou bloqueado. O estado liberado ocorre quando não há uma proteína ligada ao operador impedindo a passagem da RNA polimerase. Neste caso, ocorre a transcrição. O estado bloqueado ocorre quando existe uma proteína ligada ao operador. Essa proteína impede a passagem da RNA polimerase, logo, não é possível realizar a transcrição dos genes.

b) **Região Gênica:** são os genes do *operon*. Cada *operon* é formado por um conjunto de genes relacionados entre si (área verde da figura), isto é, que atuam na mesma proteína, inibindo ou estimulando.

c) **Região Terminadora:** é a área do DNA que indica o final do *operon*. Essa região indica para a polimerase que neste ponto termina a transcrição.

### 2.1.3 BIOLOGIA SINTÉTICA

O termo biologia sintética foi usado inicialmente por Barbara Hobom em 1980. O termo foi usado para descrever as bactérias que foram modificadas geneticamente através das tecnologias de recombinação de DNA. Segundo Hobom (1980), as bactérias são sistemas vivos que sofrem intervenções humanas (sintética).

Inicialmente, biologia sintética era amplamente usado como sinônimo de bioengenharia.

Keshava et al. (2017) descreveu biologia sintética como moléculas orgânicas artificiais que funcionam em organismos vivos. Nesse contexto, considerando um escopo mais amplo, biologia sintética é usada como referência para “redesenhar a vida”. A biologia sintética pode ser entendida como uma forma de estudar sínteses orgânicas e criar moléculas artificialmente que imitem o comportamento biológico natural do organismo (BENNER; SISMOUR, 2005). Ainda sobre os estudos de Benner e Sismour(2005), um dos principais objetivos da criação de moléculas artificiais é ajudar cientistas a entender o funcionamento biológico para desenvolver técnicas que futuramente irão colaborar com a medicina, biologia, biotecnologia, entre outras áreas.

Segundo estudos da comissão europeia (CE) (2014) de biologia sintética, a definição para melhor explicar esta nomenclatura é “a aplicação da ciência, tecnologia e engenharia para facilitar e acelerar a modelagem, formação e/ou modificação de materiais genéticos em organismos vivos”. Portanto, esta definição tem algumas características em comum com as definições anteriores como a modelagem de sistemas biológicos para melhor entender seu funcionamento e de sistemas inexistentes para observar consequências após determinada modificação.

Ainda de acordo com a CE, esses estudos tecnológicos na área da biotecnologia terão grande impacto na economia europeia e outras grandes economias mundiais por oferecer oportunidades para a criação de novas indústrias especializadas. Apesar dos inúmeros benefícios prometidos por essa recente linha de pesquisa, há muitas incertezas sobre o desenvolvimento de vidas sintéticas, principalmente sobre seus impactos referentes à perda da diversidade biológica por falta de controle, aos componentes utilizados para manipulação e as consequências dos resultados obtidos a longo prazo.

O primeiro resultado obtido nesse campo de engenharia sintética foi de Komaroff et al. (1978) que realizou um trabalho sobre a detecção em larga escala da produção de insulina na bactéria modelo *Escherichia coli*. O desenvolvimento da tecnologia do DNA recombinante demonstrou a partir da década de 70 que os genes poderiam ser sintetizados artificialmente. Isso permitiu que fossem feitas manipulações naturais ou artificiais nas células (ALBERTS et al., 1989).

## a) Áreas da Biologia Sintética

### A.1 Genômica Sintética e Síntese de DNA

Um dos principais métodos da genômica sintética é a sintetização de DNA e a capacidade de montagem e construção de material genético para introduzir em organismos vivos (ANEMAET et al., 2010). Essa área da biologia sintética visa não somente uma parte do sistema biológico ou uma única via metabólica, mas sim uma escala genômica, ou seja, uma busca para um organismo completamente sintético ou a capacidade de reproduzir partes desse organismo capazes de substituir a natural caso seja necessário. A tendência, segundo essa linha de pesquisa, é que toda aplicação prática de biologia sintética siga os requerimentos de genômica sintética (MONTAGUE et. al., 2012).

### A.2 Engenharia Metabólica

Engenharia molecular inclui métodos para o projeto e síntese de novas moléculas com propriedades físicas ou funcionalidades desejáveis (NATURE, 2017). Um aspecto importante na engenharia metabólica é a capacidade de sintetizar moléculas e montá-las. Sandoval (2012) afirma que comparando com abordagens como gene-gene a montagem molecular em escala genômica pode ser fundamental para transformar o tempo e os custos necessários hoje para o desenvolvimento de um produto, e esses produtos podem dar origem para novas fontes de energia menos poluentes.

### A.3 Biosistemas Ortogonais

Esse tipo de biosistema é mais um dos grandes desafios da biologia sintética. Basicamente, todos os estudos feitos em organismos vivos utilizam o DNA como fonte primária para coletar informações acerca do organismo. Tentando reutilizar componentes bioquímicos, engenheiros começaram um novo projeto para servir como alternativa para o DNA. Esse estudo (HERDEWIJN et al., 2008) sugeriu que a descoberta de uma alternativa ao DNA pode ser responsável pela produção de novas proteínas capazes de fortalecer células e colaborar com a indústria farmacêutica e interessados de um modo geral.

### A.4 Protocélulas

As protocélulas visam a replicação própria da célula para manter os organismos artificialmente e encapsular sistemas sintéticos inteiros. No entanto, possui o objetivo geral comum com as áreas anteriores, pois sua abordagem é da parte para o todo até formar sistemas mais complexos (CHANNON, 2008).

Diante do exposto, apesar de atuarem com problemas específicos, todas as áreas da biologia sintética utilizam a construção de “blocos básicos da célula”. Esses “blocos” são modificados artificialmente e reinseridos no organismo para analisar as consequências para futuro uso no meio medicinal. O estudo de circuitos artificiais mostrou que é possível, em alguns casos, representar certos funcionamentos da célula através de circuitos lógicos, como os utilizados na eletrônica (COMISSÃO EUROPEIA, 2014).

Para auxiliar nesse processo de representação computacional através de circuitos, linguagens de programação estão sendo desenvolvidas e adaptadas para a programação celular. O MIT, em 2016, criou uma plataforma chamada CELLO (*Cell Logic*, disponível em [www.cellocad.org](http://www.cellocad.org)). CELLO é um *framework* que funciona como um compilador da linguagem de programação verilog (CELLO website), e, de outros parâmetros de entrada, consegue convertê-los em circuitos digitais.

Esta plataforma utiliza os sinais digitais gerados pelas operações lógicas NOT e NOR para gerar um circuito otimizado que melhor complete e represente a sequência de DNA em análise. Além disso, o CELLO representa um avanço em busca desse novo desafio que é representar células através de circuitos lógicos. No entanto, alguns dados de entrada precisam ser coletados *in vivo* como é o caso do dado de entrada RPU (do inglês, *Rate Promoter Unit*). A coleta desses dados de entrada *in vivo* foi um impedimento para que esta plataforma fosse utilizada neste trabalho.

## 2.2 REDES BOOLEANAS

As redes booleanas que foram, a princípio, propostas como modelos genéricos de regulação gênica, são um caso especial de modelos discretos contínuos e aplicados inicialmente por Kauffman (1969). A rede booleana é formada por uma série de nós cujos estados são representações binárias que são determinadas por outros nós através de funções booleanas. Em termos de complexidade, as redes booleanas são esquemas de representação simples e

abstratas, porém essa simplicidade é a vantagem do modelo, pois tenta representar de maneira simples o processo de regulação gênica. Essas características, no entanto, são limitadas a um estudo topológico da via, pois seus resultados mostram ligações diretas entre os genes estudados sem o nível de detalhamento das reações químicas.

Com base nos trabalhos de Vicente (2016) sobre as redes booleanas, podemos inferir que elas podem ser usadas para descrever o comportamento em um determinado instante no tempo definido no sistema e entender como perturbações internas ou externas podem influenciar esse comportamento. Desse modo, as redes booleanas também podem ser usadas para testar hipóteses que poderão ser confirmadas posteriormente em laboratório, esse processo requer menos tempo para ser executado e pode gerar resultados fidedignos. Na sequência, será apresentada uma visão geral da metodologia aplicada para melhor exemplificar o processo de construção de uma rede booleana.

Uma variável booleana assume apenas dois valores: 0 (zero) e 1 (um). Estes valores fazem referência às operações “Falsa” e “Verdadeira”, respectivamente (Wang, 2012). Na análise biológica, esses dois valores referem-se à presença ou não de uma determinada substância no meio, como a ativação ou não de um gene. Através de funções booleanas é possível transitar entre os estados do sistema. Wang (2012) afirma que as funções booleanas herdam as propriedades da lógica booleana composta das operações básicas AND, OR e NOT.

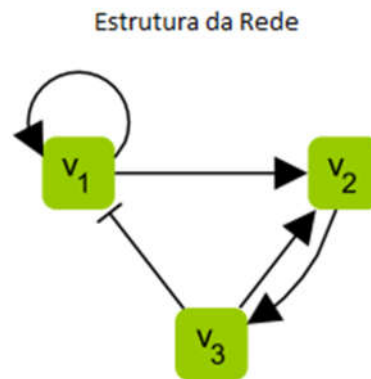
Essas operações podem ser combinadas para gerar a saída do sistema ou mudança entre estados, por exemplo, a função  $D = (A \text{ OR } B) \text{ AND NOT } C$  é formada por quatro variáveis (A, B, C e D) e determina que o estado D depende do resultado das operações entre A, B e C. A entrada de um sistema booleano normalmente se dá por uma tabela-verdade que é formada por todas as variáveis e todas as combinações binárias possíveis entre elas onde cada linha apresenta uma combinação de valores (estados) entre as variáveis de entrada e os valores de saída associados a elas.

Ainda segundo Wang (2012), uma rede booleana consiste em um conjunto de variáveis booleanas ( $a_1, a_2, a_3 \dots a_n$ ) cujo valor é determinado por outra variável na rede através de um conjunto de funções booleanas  $b=(b_1, b_2, b_3, \dots, b_n)$  associadas com cada variável. A atualização dos estados no sistema pode ocorrer de forma síncrona, onde todos os estados são alterados simultaneamente assim que ocorra

uma modificação na entrada do sistema. Nesse caso, o trabalho é feito com instantes de tempo onde cada resultado mostrado em um instante  $t+1$  é determinado por um estado no instante de tempo  $t$ .

Em uma atualização assíncrona, um estado pode ser atualizado ou não, depende da análise do tempo que se está fazendo em um determinado estado. As redes booleanas podem ser projetadas em forma de grafo  $(V, E)$ , onde  $V$  significa os vértices e  $E$  as arestas. Esse grafo é gerado observando a relação entre os dados envolvidos na análise. Exemplificando, caso haja uma relação onde um gene  $R$  ativa ou participa da formação do gene  $T$  uma seta será direcionada de  $R$  para  $T$  ( $R \rightarrow T$ ). Esse processo é feito até que todas as dependências sejam estabelecidas. A figura abaixo mostra uma relação entre os genes  $v_1$ ,  $v_2$  e  $v_3$ .

FIGURA 3: ESQUEMA DA RELAÇÃO ENTRE OS GENES  $v_1$ ,  $v_2$  E  $v_3$ .



FONTE: Albert e Wang (2009).

Nessa figura, pode-se perceber a estrutura de um grafo para indicar a relação entre as substâncias envolvidas. Entretanto, um detalhe é mostrado entre os genes  $v_1$  e  $v_3$ . A seta em forma de T é usada para indicar que um gene inibe o outro, neste exemplo,  $v_3$  inibe  $v_1$ . As funções booleanas que podem ser usadas para representar a imagem da figura são apresentadas a seguir:

$$V_1 \leftarrow V_1 \text{ OR } (\text{NOT } V_3)$$

$$V_2 \leftarrow V_1 \text{ AND } V_3$$

$$V_3 \leftarrow V_2$$

Essas são as regras que irão determinar a saída para o estado seguinte. Podemos exemplificar essas funções com a regulação gênica. A operação AND indica que ambos devem estar presentes para que provoquem uma mudança de estado no gene alvo. A operação OR significa que pelo menos uma das substâncias deve estar ativa para realizar uma reação no gene alvo. A operação NOT ocorre quando um gene inibe outro gene. Nesse caso, a operação NOT é representada no grafo com uma seta em forma de T. A tabela a seguir apresenta a tabela-verdade para análise de diferentes instantes de tempo.

TABELA 1: TRANSIÇÃO DE ESTADOS

T				T +1		
V1	V2	V3		V1	V2	V3
0	0	0		1	0	0
0	0	1		0	0	0
0	1	0	Aplicando as equações 1, 2 e 3, obtemos:	1	0	1
0	1	1		0	0	1
1	0	0		1	0	0
1	0	1		1	1	0
1	1	0		1	0	1
1	1	1		1	1	1

FONTE: criada pelo autor

Com a tabela-verdade gerada é possível aplicar essas regras e determinar a situação de cada estado do sistema em um instante  $t$  para um instante  $t+1$ . As regras são aplicadas nas linhas da tabela e uma nova tabela é criada para representar o instante de tempo posterior. O número de estados no sistema é determinado pelo número de componentes de entrada onde a definição é dada pela função exponencial  $2^n$ , onde  $n$  é o número de entrada. Neste exemplo,  $n$  é igual a três, pois existem três variáveis (V1, V2 e V3), logo, temos um total de 8 estados.

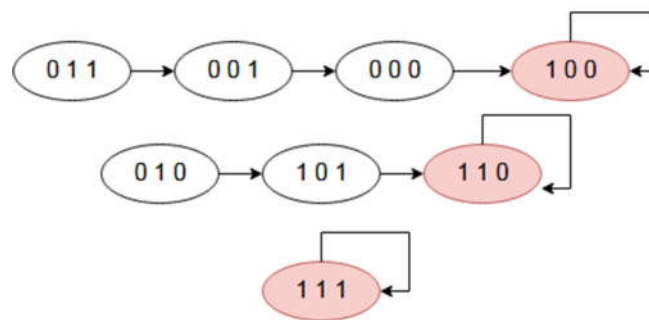
Para aplicar as mudanças de estado na tabela acima realizamos o seguinte processo. Primeiramente, criamos a tabela com as 3 entradas e todas as combinações possíveis. A seguir, para a criação do lado direito da tabela (T+1) utilizamos as regras. Logo, a primeira coluna V1 é o resultado das operações V1 OR

(NOT V3) no tempo  $t$ . A segunda coluna V2 é resultado da operação V1 AND V3, e por fim, a coluna V3 recebe o valor de V2 do tempo  $t$  seguindo a regra  $V3 \leftarrow V2$ .

Com isso, comparamos linha por linha, afirmando, por exemplo, que do estado 000 o sistema vai para o estado 100, do 001 ele muda para 000, e assim sucessivamente.

Após a aplicação das regras, o sistema segue um fluxo contínuo até encontrar um ponto fixo ou um loop indicando o final do processo para a amostra. Esse estado final é conhecido como *attractor*, destacados na cor rosa como apresentado na figura 4 e cada sequência abaixo é chamada de *bacia*. Essa sequência representa os estados de transição da via de acordo com a tabela verdade. Portanto, com 3 sequências diferentes, temos 3 maneiras para analisar o fluxo dessa via.

FIGURA 4: ESTADOS REPRESENTANDO AS TRANSIÇÕES E OS ESTADOS FINAIS.



FONTE: criada pelo autor.

As ligações diretas nos grafos podem ser diretamente relacionadas com o diagrama das vias metabólicas do sistema regulatório biológico, pois cada vértice pode indicar um gene, metabólito ou uma proteína que pode estar ativo ou em grande quantidade (*TRUE*) ou inativo ou em pequena quantidade no ambiente (*FALSE*).

### 2.2.1 Redes Booleanas Probabilísticas

As redes booleanas probabilísticas (PBN, do inglês *Probabilistic Boolean Networks*) apresentam transações para mais de um estado. Ela funciona como um acréscimo para a rede booleana anterior, pois além dos estados prováveis de uma BN, a rede probabilística também possui transações para estados menos prováveis. As arestas possuem pesos indicando a probabilidade de um determinado estado ir para os seguintes. A função de entrada é definida por Li et al (2004) e Zang et al (2006) como:

Equação 1: Função Probabilística de entrada

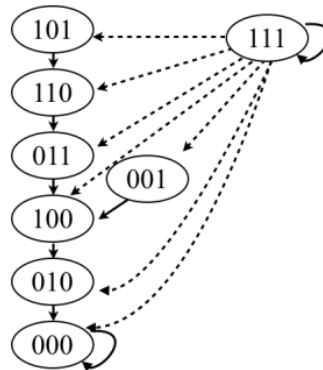
$$Y_i(t) = \sum_j a_{j,i} \cdot x_j(t)$$

Onde  $a_{i,j}$  é 1 se o gene  $j$  ativa o gene  $i$ , -1 se inibe e 0(zero) se não há relação entre eles, e o 'x' é o valor de entrada da função

A principal característica de uma rede booleana probabilística é a possibilidade de mudança de um estado para qualquer outro. Vicente (2016) afirma que “não se espera em um sistema como o sistema biológico as transições entre estados sejam determinísticas. Portanto, é importante supor que há uma chance (ainda que mínima) de transições para quaisquer estados”. Os pesos são definidos durante a modelagem. Por padrão, os valores iniciais são os mesmos, divididos de maneira uniforme entre as arestas, porém, podem ser manipulados e alterados.

A imagem a seguir exemplifica as PBN:

FIGURA 5: EXEMPLO DE UMA PBN



FONTE: Vicente, F. (2016).

Nesse exemplo existem transições entre todos os estados. Para facilitar a visualização, foram adicionadas apenas todas as transições para um dos estados.

### 2.2.2 Redes Gênicas Probabilísticas

Como explicado anteriormente, a transcrição gênica nos procariotos ocorre através dos operons. A tradução ocorre logo após a transcrição. Considerando esses conceitos, a transcrição de um gene para ser traduzido para proteína depende de outra proteína que já foi expressa em um estado de tempo anterior. Dessa mesma

forma, quando ocorre a tradução a proteína gerada (codificadora ou não) irá atuar em outro processo de regulação como ativadora ou inibidora.

As redes gênicas probabilísticas utilizam esses conceitos para modelar de forma discreta e dinâmica através de instantes de tempo onde o estado seguinte depende do estado anterior. Barreira et al (2007) formalizou o modelo utilizando algumas regras e conjuntos onde os genes podem ser representados como sub-expresso (-1), transcrição basal (0) e super-expresso (1). Os genes superexpressos são aqueles que possuem um nível de expressão alto na via em determinado momento de tempo. A transcrição basal ocorre quando o organismo está com seus níveis de expressão em condições normais, e a subexpressão ocorre quando um determinado a expressão de um gene diminui.

### 2.3 Grafos e a Dinâmica de Redes Bioquímicas

Como apresentado anteriormente, a análise booleana é uma análise abstrata que tenta explicar de forma simples e eficiente o comportamento celular. Realizando uma análise mais biológica, sabemos que todo o funcionamento celular ocorre através de reações químicas e essa análise dinâmica não é algo tão trivial. Esse dinamismo das reações químicas inspirou o estudo dessas vias metabólicas através de grafos, pois eles podem representar estados e possuir várias saídas para diferentes estados.

Outra técnica para modelar vias metabólicas é o uso de equações diferenciais ODE (do inglês, *Ordinary Differential Equation*) que pode ser representada como na figura a seguir:

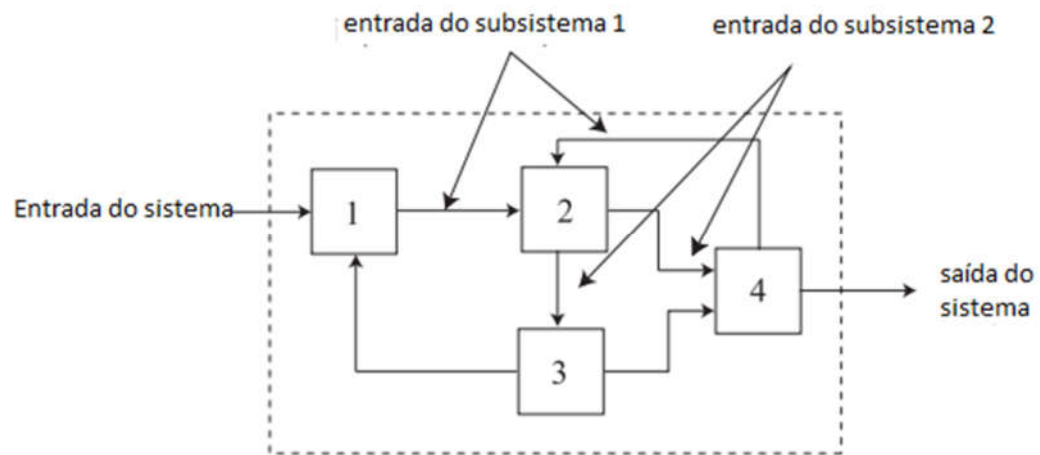
#### EQUAÇÃO 2: REPRESENTAÇÃO DAS EQUAÇÕES DIFERENCIAIS

$$\begin{aligned} \dot{x}_n(t) &= \frac{dx_n(t)}{dt} = f_n(x_1(t), \dots, x_n(t), \dots, u_m(t)) \\ y(t) &= (x_1(t), \dots, x_n(t)) \end{aligned}$$

Onde  $x_n(t)$  e  $y = h(t)$  representam a concentração química (proteínas, m RNA, metabólitos),  $f_n$  e  $h$  representação o vetor para armazenar as entradas e  $x(t)$  são entradas (estímulos, sinais externos, controles, funções). As representações das redes de reações químicas são feitas através de grafos bipartidos.

A abordagem modular é baseada na visualização de redes grandes como sendo feitas de pequenas sub-redes conectadas à variáveis de entrada e saída (ANGELI; SONTAG, 2011). Eles afirmam também que essa abordagem é muito boa para sistemas de pouca complexidade, principalmente se ela for usada para abstrair reações. Um modelo para a abordagem modular pode ser visto na figura 6:

FIGURA 6: EXEMPLO DE MODELAGEM MODULAR



FONTE: Angeli e Sontag (2011).

Na figura 6, os números 1 e 2 representam o primeiro subsistema e os números 3 e 4 representam o segundo subsistema. Cada subsistema possui funções específicas, porém se comunicam entre si. A saída do subsistema 1 é a entrada do subsistema 2, e a saída do subsistema 2 pode ser a saída do sistema por completo, ou pode servir de entrada para o subsistema 1 realizar outro processamento.

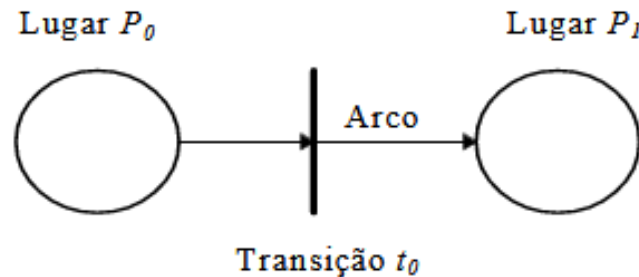
#### 2.4 Redes de Petri

A modelagem utilizando redes de Petri foi utilizada pioneiramente por Reddy et al (1993). As redes de Petri são modelos matemáticos e gráficos normalmente estáticos que são aplicados em sistemas de telecomunicação e outros sistemas não-biológicos, e hoje, também são utilizados para representar processos biológicos normalmente em redes assíncronas, ou seja, onde a atualização de um estado não ocorre no mesmo instante de tempo para todos.

A representação visual de uma rede de Petri é formada por um ativo e um passivo chamados de transição (barra) e lugar (círculo), respectivamente. Os estados

são representados pelos lugares e as transições representam as funções e ações realizadas pelo sistema. Eles são conectados através de arestas (arcos) que podem ser simples ou múltiplos (FRANCÊS, 2003).

FIGURA 7: REDE DE PETRI



FONTE: Francês (2003).

As redes de Petri também são usadas para modelagem de regulação gênica (SIMÃO, 2006) e também na simulação de processos biológicos (TOMIYAMA, 2007). A rede proposta por Reddy (1993) trabalha com estados nos lugares e funções nos arcos. Ela funciona como uma sequência de passos que o processo biológico deve seguir, trabalhando, assim, de maneira sequencial. As redes de Petri têm o conceito de marcação. As marcações são feitas nos lugares (vértices) e indicam a situação desse lugar em um determinado momento. Essa marcação é modificada a cada transição que ocorre no vértice.

As redes de Petri são classificadas como Ordinárias e Não-Ordinárias. As ordinárias são divididas em **binárias e lugar-transição**. As binárias, são redes mais simples que permitem apenas um valor (*token*) em cada *lugar* e valores unitários nos *arcos*. As redes lugar-transição, permitem mais de um valor no *lugar* e valores não unitários nos *arcos*. As redes não-ordinárias não são uma forma de maximizar o modelo, porém elas podem trabalhar com a ideia de abstração tornando o modelo mais ou menos abstrato devido os maiores valores permitidos para os lugares e arcos.

Todas as redes, de um modo geral, são formadas pela ligação e pequenos módulos. Esses módulos são chamados de módulos elementares. Entre os módulos básicos definidos por REDDY (1993), temos:

1) Módulo de sequenciamento: trabalha com o conceito condicional e sequencial onde uma condição necessita ser atendida para que o arco siga para o lugar seguinte.

2) Módulo distributivo: parte do princípio que um lugar permite dos arcos de saída dando origem a dois processos paralelos. Assim temos o arco origem (pai) e os arcos paralelos surgidos a partir dele (filhos).

3) Junção: processo inverso do modelo distributivo. Quando dois processos convergem para um mesmo lugar chamamos esse processo de junção.

4) Módulo não-determinístico: Este processo é um modelo distributivo, porém não existe uma possibilidade de escolha. Por não haver essa possibilidade de escolha, a situação gera um momento de conflito, pois o lugar pai precisa decidir para qual lugar filho será enviada a transição. Esse mecanismo precisa ser resolvido através das funções de transição, sendo isto, um dos desafios das redes de Petri.

A imagem, figura 8, apresenta um comparativo entre as redes apresentadas anteriormente:



FONTE: (Wang, adaptada), 2012.

## 2.5 Trabalhos Relacionados

Alguns trabalhos recentes demonstram o estado da arte sobre o estudo das redes booleanas para modelagem das vias de regulação gênica Giri et al., (2017)

realizou um trabalho para modelar a regulação gênica da proteína P53. Esta proteína, é uma proteína supressora de tumores localizada no 17º cromossomo (GIRI, et al., 2017). O estudo de Giri et al.(2017) realizou a modelagem da via utilizando redes booleanas. Essa modelagem deu origem aos circuitos lógicos que representam o comportamento da via no organismo e tem como objetivo contribuir para o tratamento do câncer.

Outro trabalho importante na área é o de Bloomingdale et al. (2018) que estuda como os componentes presentes em remédios afetam o dinamismo do sistema biológico. Para a construção de uma rede simples que não necessitasse de muitos parâmetros, eles optaram por usar a modelagem booleana como ponto de partida em suas pesquisas.

Para finalizar, o trabalho de Kobayashi e Hiraishi (2017) estuda um aperfeiçoamento das redes booleanas para as redes probabilísticas. O problema das redes probabilísticas é a condição para definir as probabilidades para decidir como ficará o percurso até chegar a estado-final. A aplicação desse estudo em futuros trabalhos é um dos objetivos deles.

Com isso, percebemos que o estudo das redes booleanas está sendo usada, também, para buscar resultados para a área da saúde e dando sua contribuição para a melhora dos estudos biológicos. No tópico seguinte, são apresentados passos importantes para a elaboração de uma rede booleana.

## **2.6 ELABORAÇÃO DAS REDES BOOLEANAS**

Um dos principais objetivos dessa modelagem é simplificar as relações gênicas para identificar o comportamento dos genes em determinadas situações. Segundo Albert e Wang (2009), existem seis passos principais para a elaboração das redes booleanas:

### **2.6.1 Síntese da Rede**

Durante a preparação dos estudos utilizando redes booleanas é importante realizar bem a preparação dos dados que inclui o estudo da literatura, os dados experimentais e o software NET-SYNTHESIS (KASHALO et al., 2007) utilizado nesta referência. A análise da literatura é basicamente o estudo do organismo que se deseja estudar e como ocorrem os fenômenos biológicos dele. O estudo das vias

biológicas é um exemplo para que se possa conhecer quais as funções químicas, ou quais os genes ou proteínas que fazem parte da rede gênica do organismo.

Como nem sempre a literatura mostra uma informação completa, ou um estudo de tudo aquilo que podemos analisar, é realizado o complemento desta pesquisa por meio de observações em laboratório para que possam ser identificados atividades, taxas de concentração, perturbações dos meios internos, externos e testes.

Com isso, uma mutação ou a super expressão de um gene pode apresentar evidências de qual reagente provoca determinada mudança. Atividades enzimáticas, interação proteína-proteína, e interações de transcrição oferecem provas bioquímicas da relação direta entre dois componentes.

Essas relações entre os componentes podem ser representadas na forma componente-componente com A promovendo B ( $A \rightarrow B$ ) ou A inibindo B ( $A \dashv\vdash B$ ) que corresponde à uma ligação direta na representação com grafos. O principal objetivo dessa fase é levantar informações e dados dos componentes a serem estudados e construir uma relação entre eles na forma booleana.

### 2.6.2 Transferência da Função Booleana

Com os dados recolhidos, para que possamos entender o comportamento do sistema é preciso entender a relação de dependência dos componentes entre si determinando quais funcionam como ativadores e inibidores, pois todas as funções e grafos gerados dependem de um conhecimento prévio da via metabólica. Entendendo essa relação, é possível criar as regras booleanas e conseqüentemente os nós para representar os estados e as relações de suas transações. A função de transferência é simplesmente a passagem de um estado para outro, por exemplo, a relação  $A^* = L$  indica que o próximo estado de A será o estado corrente de L, o asterisco indica que A será modificado em um instante t futuro (t+1).

Nessa fase, pode ser feito um comparativo com os dados biológicos, por exemplo, quando existe um inibidor ( $A \dashv\vdash B$ ), neste caso A inibe B, assumimos que quando há uma alta concentração de A, B é inativado, pois ele é inibido pela presença de A, ou seja, em caso de inibição as concentrações são inversamente proporcionais. No entanto, quando ocorre uma ativação, as concentrações são diretamente

proporcionais. Supondo um segundo exemplo onde C ativa B ( $C \rightarrow B$ ), quando o percentual de C estiver alto, o de B estará também.

Quando a operação requer vários componentes são feitas combinações com os três operadores básicos da lógica booleana (AND, OR, NOT) a fim de determinar a ativação ou inibição de um composto, como apresentado na equação 1. Portanto, com o término dessa segunda fase já é possível entender quais os compostos que serão analisados e quais as relações entre eles.

### 2.6.3 Transição de Estados

Dada a estrutura topológica e as funções de transição, a passagem de um estado para o outro pode ser dada de diversas formas. Em um sistema síncrono, essa passagem é feita atualizando todos os estados simultaneamente. A equação a seguir mostra essa transição:

$$X_i(t+1) = F_i(X_1(t), x_2(t), \dots, X_n(t)) \quad (\text{Equação 1})$$

A vantagem dos modelos síncronos é que a sequência do estado dinâmico do sistema é determinística, isto é, ela leva para um mesmo *attractor* em diferentes simulações. Os processos biológicos são complicados, pois podem variar muito em relação à concentração e ao tempo em que acontecem. O modelo booleano assume que há um perfeito sincronismo entre os estados dos componentes e suas transições, dessa forma é usada uma abstração para que seja feita a simulação, não levando em consideração os detalhes que ocorrem na célula do organismo. Para aprimorar essa precisão, o modelo assíncrono não faz a atualização de todos os nós simultaneamente, nesse caso, apenas um ou alguns nós são atualizados em cada instante de tempo sempre garantindo que um nó seja modificado apenas uma vez a cada unidade de tempo.

### 2.6.4 Analisando os Estados do Sistema

O número de estados em um sistema booleana é na ordem de  $2^n$ , onde  $n$  é o número de entradas. Sendo assim, existe um número limitado de estados formados e todas as sequências levarão para um ou mais estados finais ou para um ciclo.

Configurando os estados iniciais de forma diferente o comportamento final pode ser identificado. Sendo assim, *attractors* podem ser descritos também como uma combinação de entradas sinalizando um comportamento das vias metabólicas do organismo.

Esse estado final é mais interessante para estudo do que os estados intermediários, pois mostra uma solução para uma hipótese ou testa várias possibilidades de acontecimentos dadas as condições iniciais. Estas condições iniciais são as entradas binárias que podem ser arbitrárias ou aleatórias. Com isso, esta parte da análise contém as condições iniciais do sistema como dados biológicos que são transformados em amostras binárias e também uma análise dos estados finais do sistema. Algumas bibliotecas como BoolNet em R permitem a identificação dos *attractors* e a visualização da rede em forma de grafos.

#### 2.6.5 Teste de Correção do Modelo

A preparação dos testes do modelo visa identificar se o modelo é capaz de reproduzir de forma correta os modelos biológicos reais comparando suas entradas com as saídas desejadas e robustez do sistema, isto é, como eles se comportam com mudanças nas interações. Com essas observações pode-se inferir se o modelo precisa de mudanças ou se atende as necessidades dos dados observados em experimento ou na literatura. O teste do modelo pode, também, identificar se há uma modelagem errada, por exemplo, se um modelo sempre leva todos os estados para FALSE, sendo que nas observações o correto seria que alguns estados fossem positivos, devem-se revisar as funções que interligam o modelo para verificar se não existe uma operação trocada ou existe alguma faltando, e também a relação entre os reagentes para confirmar se um inibidor é realmente um inibidor, por exemplo.

#### 2.6.6 Implicações Biológicas e Predições

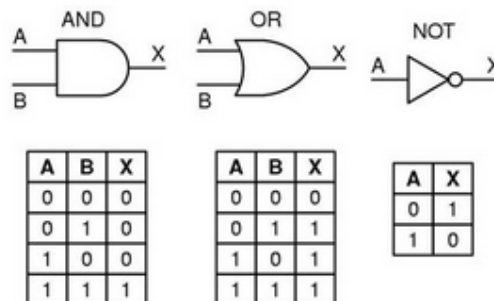
Segundo a definição de Albert (2009), o modelo discreto dinâmico nos permite integrar conhecimentos fragmentados em sistemas lógicos que posteriormente nos ajudam a interpretar e compreender o sistema em uma escala global. Uma grande vantagem desse método é a capacidade de predição, pois podemos simular diferentes entradas para o sistema. Entre as formas de perturbar o sistema para

realizar esses testes é através da inativação de determinadas entradas, setando todas para *FALSE* (*knockout*) ou tornando um gene, por exemplo, sempre ativo (*overexpression*). Essa técnica ajuda na compreensão da importância de determinados componentes para o funcionamento do sistema. Com isso, pode-se determinar o que acontecerá em longo prazo, isto é, indicar se determinada mudança tem fator vital para o organismo em questão, ou identificar qual a consequência para o organismo com o tempo.

## 2.7 DEFININDO PORTAS LÓGICAS A PARTIR DAS REGRAS BOOLEANAS

Considerando as regras apresentadas no item 4.1, podemos extrair as informações para a construção das tabelas-verdade e as portas lógicas para representar cada uma das tabelas. Para a construção das tabelas e circuitos foram utilizadas as seguintes definições:

FIGURA 9: REGRAS BOOLEANAS E SUAS REPRESENTAÇÕES



FONTE: Pereira, L. (2017).

Cada módulo abaixo representa uma interação entre os genes. A operação AND foi escolhida porque a ativação só ocorre quando as duas condições são satisfeitas. Quase todos podem sofrer uma inibição

### 3 MATERIAIS E MÉTODOS

#### 3.1 METODOLOGIA

Este trabalho é baseado no estudo de redes booleanas. Logo, nos capítulos iniciais são apresentados os principais conceitos e software utilizado na realização deste trabalho. O trabalho foi baseado nos estudos de Wang (2012), Karleback (2008) e Kauffman (1969). Estes trabalhos apresentam a metodologia para a elaboração de uma rede booleana. Com base nestes modelos, foi feita a síntese da rede (estudo do organismo), elaboração das funções booleanas, transição dos estados e análise do modelo como mostra o fluxograma abaixo.

FIGURA 10: FLUXOGRAMA DA METODOLOGIA ABORDADA



FONTE: Criada pelo autor.

Para este trabalho foi escolhido um procaríoto modelo (*Escherichia coli*), pois esta bactéria possui diversas referências e dados bem concretos para realizar a modelagem utilizando redes booleanas.

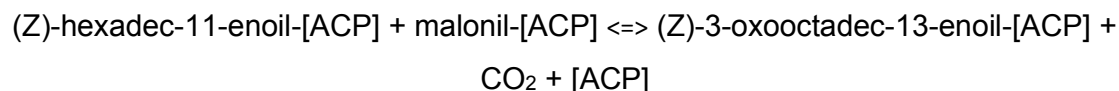
Os genes detectados para a via metabólica dos ácidos graxos, que farão parte da modelagem deste trabalho foram os genes: *fabH*, *fadD*, *fabB*, *fabF*, *fabG*, *fabZ*, *fabA*, *fabI*, *fadD*, *fadR*, *dksA*, *CPRX* e *arcA*. Os resultados dessa pesquisa indicarão a função e a relação entre eles.

A informação biológica foi retirada do banco de dados KEGG que apresenta as vias metabólicas do organismo, e também, foi usado o banco de dados RegulonDB que é um banco de dados dedicado a *Escherichia coli* k-12 e contém as informações dos genes de regulação e seus ativadores e inibidores. Para complementar essas informações, a base de dados UNIPROT foi utilizada para relacionar os genes com suas proteínas. Também foi utilizado o software Cytoscape (disponível em [www.cytoscape.org](http://www.cytoscape.org)) para a visualização da rede. Para a parte de biologia sintética foram usados dois softwares. O LogiSim (disponível em [www.cburch.com/logisim](http://www.cburch.com/logisim)) para o desenho do circuito utilizando as portas lógicas e o SimBoole (disponível em <https://omictools.com/boolean-network-simulator-tool>) para a simulação em forma de grafo definindo os genes como vértices.

Consultando os bancos de dados citados, foram extraídas informações sobre os genes destacados na via metabólica do KEGG. A seguir, são apresentadas as informações para cada um dos genes:

**a) *fabH*:** Responsável pela condensação de reações através da adição de um aceitador *acil* de dois carbonos do malonil-ACP. Ele também conduz o início das reações da via metabólica dos ácidos graxos e posteriormente governa a taxa de ácido graxo produzido. Além disso, esse gene possui uma especificidade por acetil-CoA. É essa especificidade que torna esse gene importante no percurso para a produção do ácido graxo. A proteína ligada a esse gene é a proteína transportadora de acil (ACP, do inglês, *Acyl Carrier Protein*) que é uma proteína universal e altamente conservada atuando no transporte de acil durante a síntese de ácidos graxos.

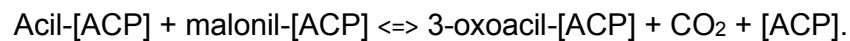
**b) *fabD*:** Assim como o *fabH*, esse gene atua de forma catalítica nas reações de condensação dos ácidos graxos. Sua reação é expressa pela equação:



**(Reação 2)**

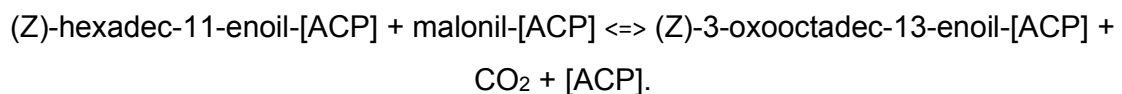
A proteína ligada ao fabD é a Malonil CoA-ACP transacilase (MCAT). Ela é uma enzima crítica responsável pelo transporte do malonil para a formação do malonil-ACP na via metabólica dos ácidos graxos. Essa enzima é considerada um importante fator na indústria farmacêutica no estudo de agentes antibacterianos (Zhang, et al., 2007).

**c) fabB:** Gene específico para o alongamento do C-10 para os insaturados C-16 e C-18. Sua atividade catalítica é apresentada pela equação:

**(Reação 3)**

Logo, a proteína ligada a esse gene é a 3-oxoacyl-[ACP].

**d) fabF:** esse gene tem preferência por substratos de cadeia curta e também atua suplementando outros substratos como o octanóico para a formação do ácido lipóico na biossíntese dos ácidos graxos. Esse ácido funciona como um anti oxidante e atua também nas funções energéticas, além disso, possui função importante tanto nas bactérias quanto no homem. Ele também tem relação direta como catalisador da enzima 3-oxoacil-[ACP]. A reação que representa a ação desse gene é a seguinte:

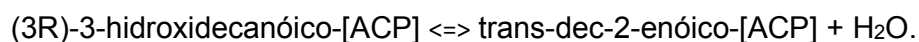
**(Reação 4)**

**e) fabG:** esse gene catalisa a redução do substrato NADPH para o produto beta-cetoacil-ACP. O primeiro passo redutivo é a redução do alongamento do ciclo da biossíntese dos ácidos graxos. Ele pode ser representado pela reação química:

(3R)-3-hidroxiacil-[ACP] + NADP<sup>+</sup>  $\rightleftharpoons$  3-oxoacil-[ACP] + NADPH e também tem ligação com a enzima 3-oxoacil-[ACP].

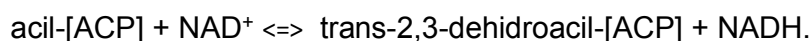
**(Reação 5)**

**f) fabA:** este gene é necessário para a introdução da instauração cis para os ácidos graxos. A função catalítica dele é na desidratação do (3R)-3-hidroxi-decanóico-ACP para E-2-decenóico-ACP. Um exemplo de uma reação química onde essa enzima está presente é a equação:



**(Reação 6)**

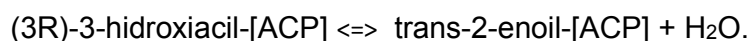
**g) fabI:** este gene está associado com o alongamento do ciclo da biossíntese dos ácidos graxos e pode ser apresentado na reação:



**(Reação 7)**

Ele é inibido pela enzima diazaborine e seus derivados.

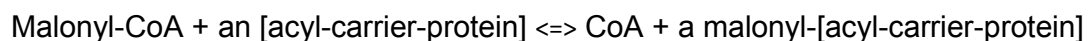
**h) fabZ:** este gene está envolvido principalmente com a biossíntese dos ácidos graxos insaturados. Ele também atua como catalisador da desidratação das cadeias curtas de beta-hidroxiacil-ACPs. Além disso, ele atua em cadeias longas e curtas saturadas de beta-hidroxiacil-ACPs.



**(Reação 8)**

**i) dksA:** esse é um fator de transcrição que atua diretamente se ligando à mRNA polimerase. Ele é um fator de transcrição mais externo à rede dos ácidos graxos atuando também na regulação de diversos amino ácidos. O dksA também atua na inibição de algumas enzimas e na reparação do DNA, podendo dessa forma, atuar em diversas reações químicas.

**j) fadD:** trabalha com longas cadeia de ácidos graxos e também na degradação ou incorporação deles com os fosfolípideos. Além disso, contribui para o crescimento da *E. coli* e fonte de energia.



### (Reação 9)

**k) fadR:** O fadR possui função de regulação multifuncional na biossíntese dos ácidos graxos. Ele atua tanto na repressão quanto na ativação de pelo menos 3 genes cujas funções estão relacionadas com o transporte de enzimas. Sua composição ao se conectar com o DNA é inibida por cadeia longas de acetil-COA que previnem o fadR de se ligar ao DNA.

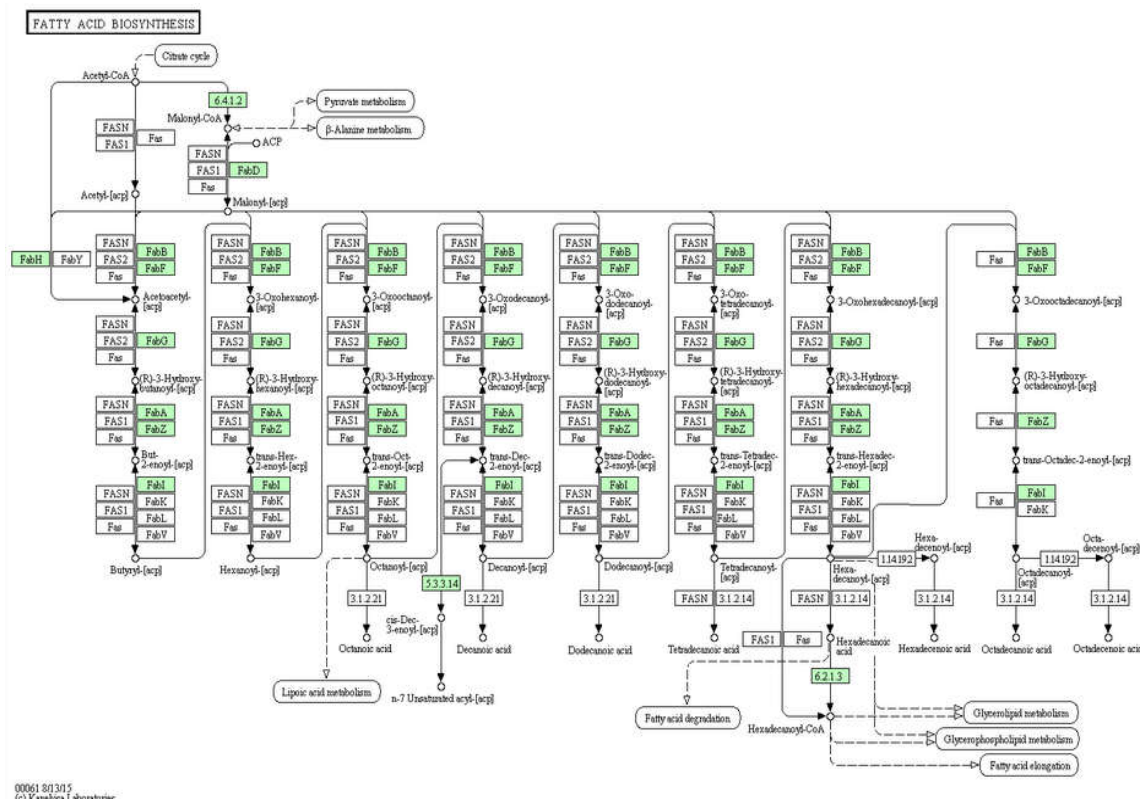
**l) gadE:** este gene é responsável por regular diversos genes envolvidos na resistência dos ácidos. Ele atua principalmente na regulação do gadA, porém pode influenciar em outros genes também.

**m) arcA:** responsável pelo controle aeróbico da célula. Ele é responsável por reprimir diversas enzimas em condições anaeróbicas, controlando assim, a resistência da *E. coli*. Sua ativação é dada pelo par do composto, o gene arcB.

**n) Cprx:** O cprx é um fator de transcrição que se liga ao DNA e atua de forma dupla tanto na inibição quanto na ativação de alguns genes. Um desses genes em que ele atua de forma ativa é o fabZ.

As características apresentadas de cada um dos genes acima podem facilitar nos resultados da simulação da modelagem, pois assim saberemos quais atividades estão sendo desempenhas em uma determinada condição inicial imposta para a simulação. A seguir a via metabólica utilizada para a modelagem booleana.

FIGURA 11: VIA DE REGULAÇÃO DA BIOSÍNTESE DO ÁCIDO GRAXO



FONTE: [http://www.genome.jp/kegg-bin/show\\_pathway?eco00061](http://www.genome.jp/kegg-bin/show_pathway?eco00061), 2017

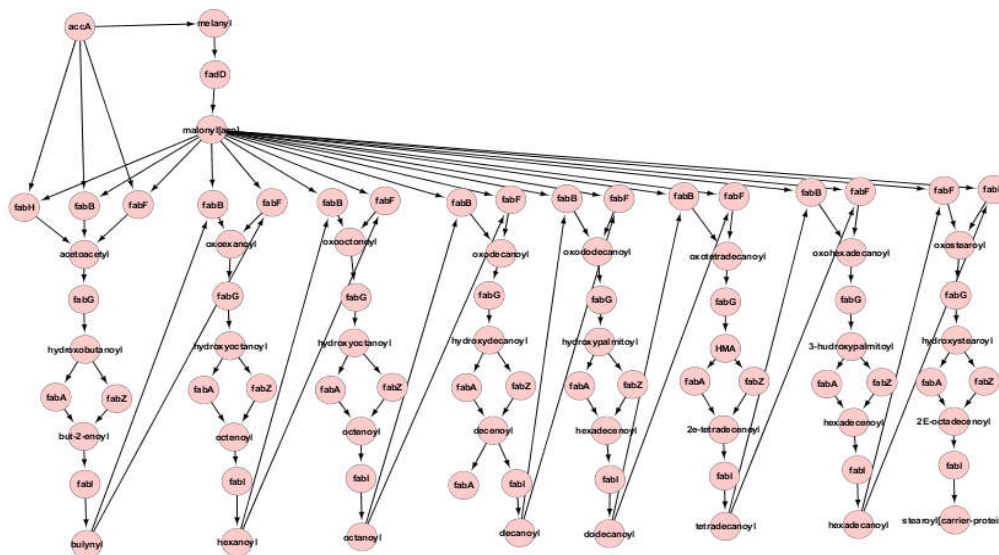
Na figura 11, são apresentados, em destaque, os genes ativos na biossíntese dos ácidos graxos. Realizando a importação dessa rede para um software de visualização (CytoScape), foram retirados os genes inativos na via que será apresentado no capítulo dos resultados.

## 4 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados os resultados obtidos para a via metabólica dos ácidos graxos da *E. coli* utilizando uma modelagem booleana.

### 4.1. UTILIZANDO A VIA METABÓLICA DO KEGG

FIGURA 12: REDE DOS ÁCIDOS GRAXOS REPRESENTADA PELOS GENES ATIVOS



FONTE: adaptada do KEGG, 2017

Vemos que os genes se repetem ao longo do percurso sendo ativados normalmente pelos mesmos genes através de diferentes enzimas. Os genes participantes do processo são fabH, fabD, fabB, fabF, fabG, fabZ, fabA e fabI. Com a utilização da base de dados RegulonDB, foram retiradas as relações de ativação e inibição para cada um deles. Considerando essas regras, foram obtidas as seguintes relações com suas respectivas regras booleanas:

- 1) fabH é ativado pelo fadR e inibido pelo dksA, logo,  $\text{fabH} \leftarrow \text{fadR NOT dksA}$ .
- 2) fabD é ativado pelo fadR e inibido pelo dksA, logo,  $\text{fabD} \leftarrow \text{fadR NOT dksA}$ .
- 3) fabB é ativado pelo fadR e inibido pelo fabR, , logo,  $\text{fabB} \leftarrow \text{fadR NOT fabR}$ .
- 4) fabF é ativado pelo fadR e inibido pelo dksA, logo,  $\text{fabF} \leftarrow \text{fadR NOT dksA}$ .
- 5) fabG é ativado pelo fadR e inibido pelo dksA, logo,  $\text{fabG} \leftarrow \text{fadR NOT dksA}$ .

6) *fabZ* é ativado pelo CPRX e GadE, logo,  $fabZ \leftarrow CPRX \text{ AND } GadE$ .

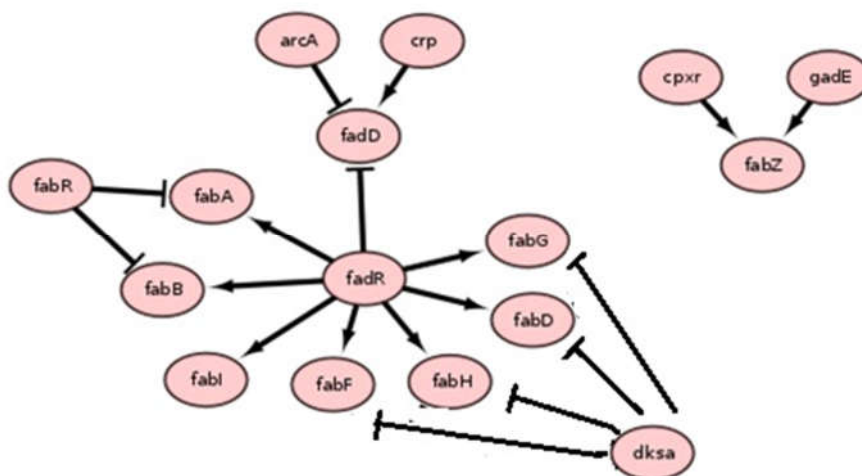
7) *fabA* é ativado pelo *fadR* e inibido pelo *fabR*, logo,  $fabA \leftarrow fadR \text{ NOT } fabR$ .

8) *fabI* é ativado pelo *fadR*, logo  $fabI \leftarrow fadR$

9) *fadD* é ativado pelo CRP e inibido pelo *fadR* e *ArcA*, logo,  $fadD \leftarrow CRP \text{ NOT } ArcA \text{ NOT } fadR$ .

Desse modo, foi criada uma rede booleana baseada no modelo de Kauffman mostrando a relação entre esses genes de acordo com a relação de ativação e inibição entre eles. A figura a seguir mostra essa relação.

FIGURA 13: REDE BOOLEANA



FONTE: Elaborada pelo autor.

#### 4.2 DEFININDO AS REGRAS BOOLEANAS

Com base na imagem acima é possível identificar o gene *fadR* como um nó importante no funcionamento da rede de um modo geral, pois é o principal ativador da relação com oito arestas saindo dele. A simulação feita pelo software online BooleSim

atribui valores *true* e *false* para cada gene e gera uma simulação baseada na regra booleana adotada. Portanto, de acordo com a regra:

$fabH = fadR \ \&\& \ (! \ dkSA)$

$fabD = fadR \ \&\& \ (! \ dkSA)$

$fabB = fadR \ \&\& \ (! \ fabR)$

$fabF = fadR \ \&\& \ (! \ dkSA)$

$fabG = fadR \ \&\& \ (! \ dkSA)$

$fabZ = cpxr \ \&\& \ gadE$

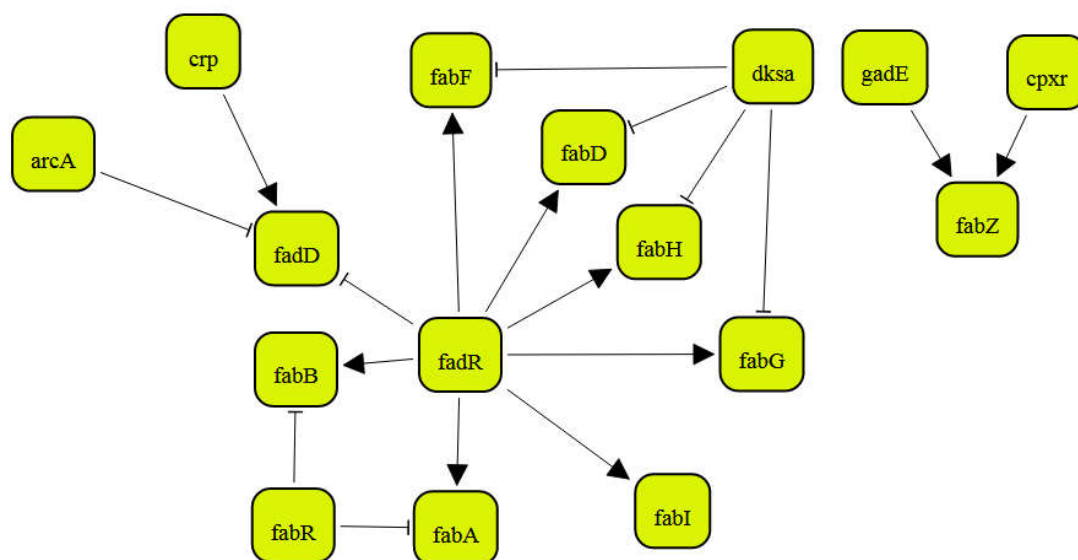
$fabA = fadR \ \&\& \ (! \ fabR)$

$fabI = fadR$

$fadD = crp \ \&\& \ (! \ fadR) \ \&\& \ (! \ arcA)$

temos:

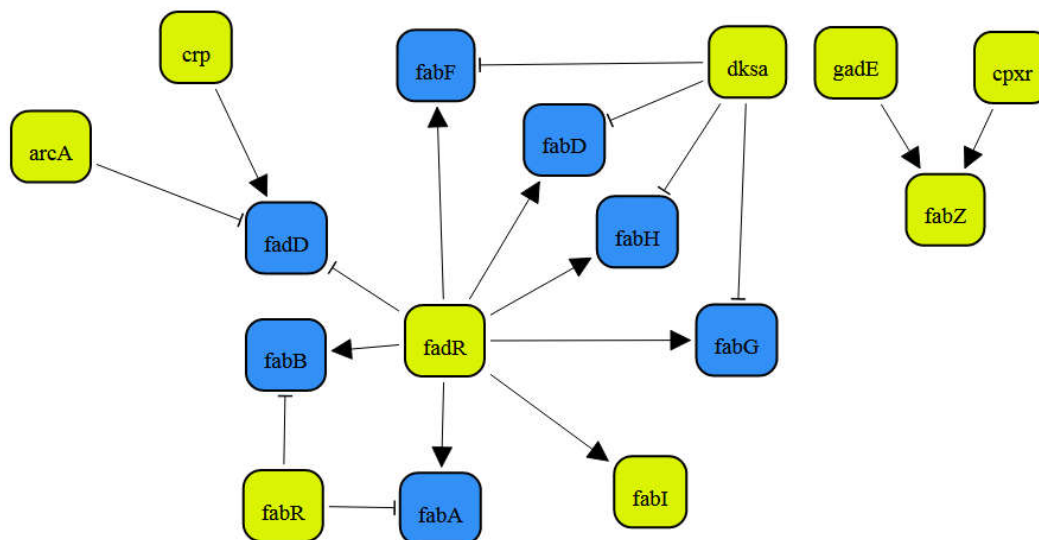
FIGURA 14: MAPA DA REDE NO SOFTWARE BOOLESIM



FONTE: Elaborada pelo autor.

O software BooleSim trabalha com duas cores para visualização: amarela (ativo, recebe valor 1) e azul (inativo, recebe valor 0). Ao executar a simulação de acordo com o esquema acima temos:

FIGURA 15: EXECUÇÃO NO BOOLESIM COM TODOS OS GENES RECEBENDO VALOR 1



FONTE: Elaborada pelo autor.

A relação adotada para os testes mostra que na presença de um inibidor o resultado será a inibição de determinados genes como mostra a tabela-verdade a seguir:

TABELA 2: RELAÇÃO ENTRE O REPRESSOR E UM GENE

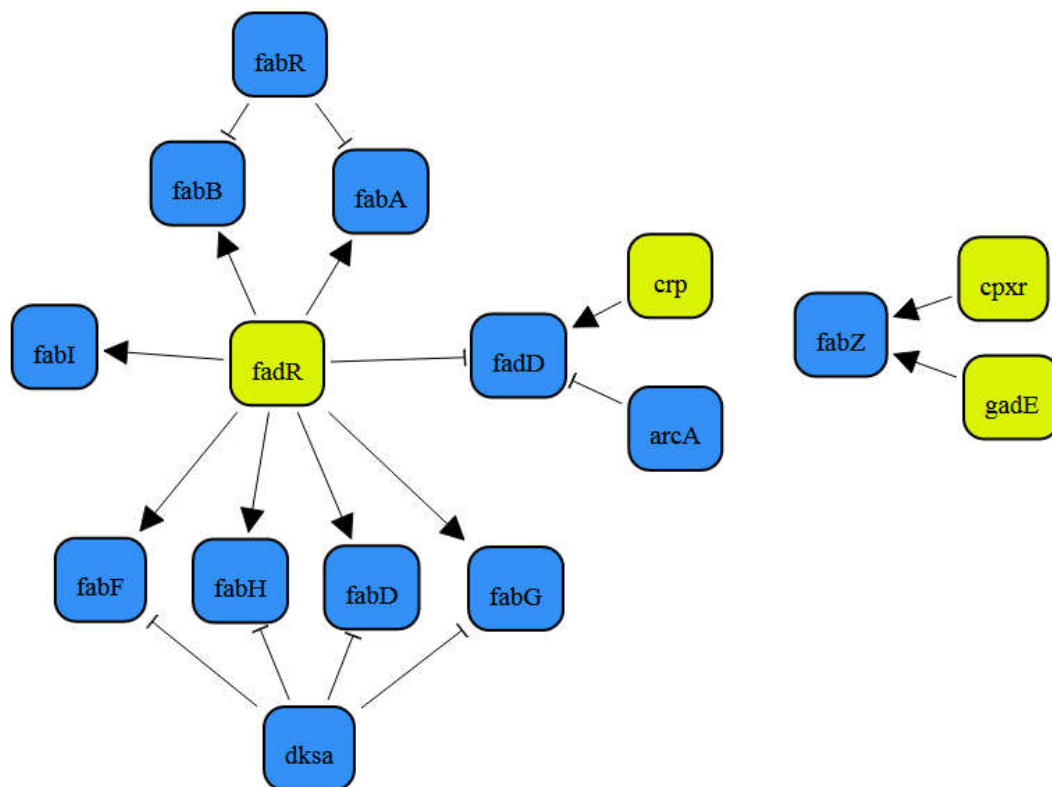
Repressor	Gene	saída
0	0	0
0	1	1
1	0	0
1	1	0

FONTE: Elaborada pelo autor

Devido à esta situação, quando a simulação foi executada (Figura 17) com os genes inibidores ativados (fabR, arcA, dksA), todos os genes relacionados a ele foram desativados (cor azul), pois, neste caso, eles atuam como inibidor. A maioria dos genes inibidos nessa simulação possui atividade catalítica e de suplementação para outros substratos como o fabF. Logo, o processo da síntese dos ácidos graxos sofre uma diminuição de produção na sua via metabólica.

Outra situação ocorre quando desativamos a maioria dos genes inibidores da rede como mostra a figura a seguir:

FIGURA 16: MAIORIA DOS GENES INIBIDORES ATIVADOS

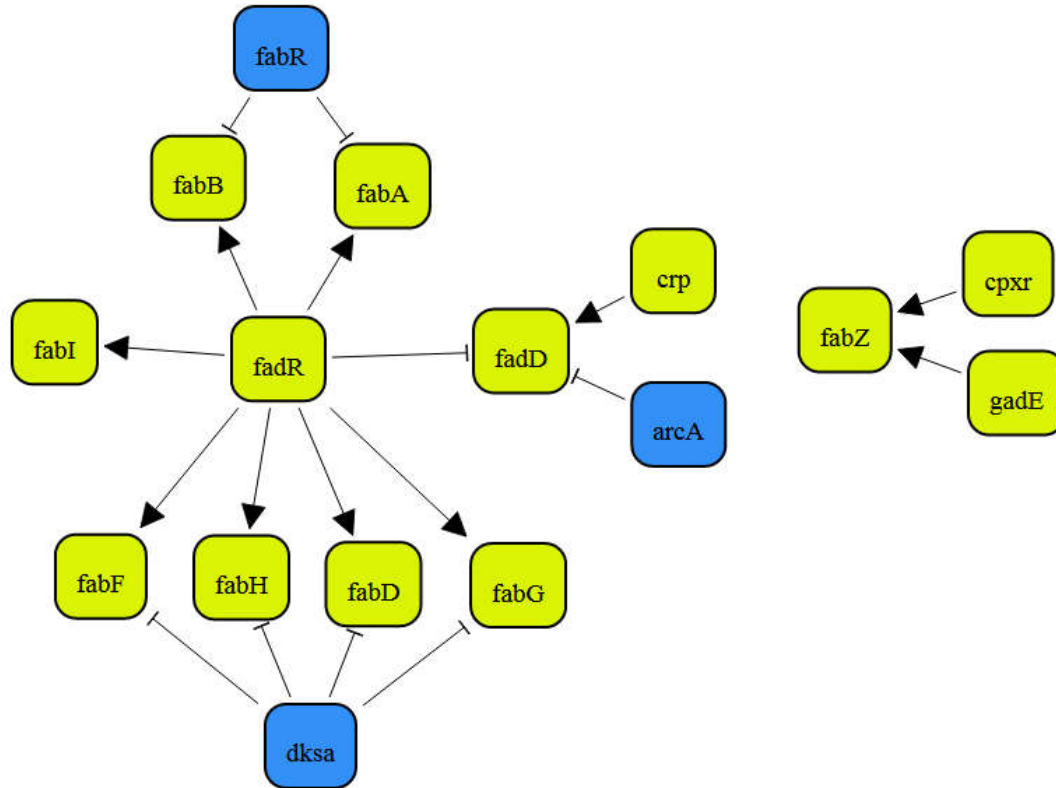


FONTE: Elaborada pelo autor.

O gene fadR não foi desativado porque ele possui propriedades de inibição e ativação. Como nesse caso ele está ligado diretamente com a ativação de sete genes e apenas inibindo um deles, ele não foi desativado.

O resultado dessa simulação é apresentado na figura a seguir:

FIGURA 17: SIMULAÇÃO COM OS INIBIDORES DESATIVADOS

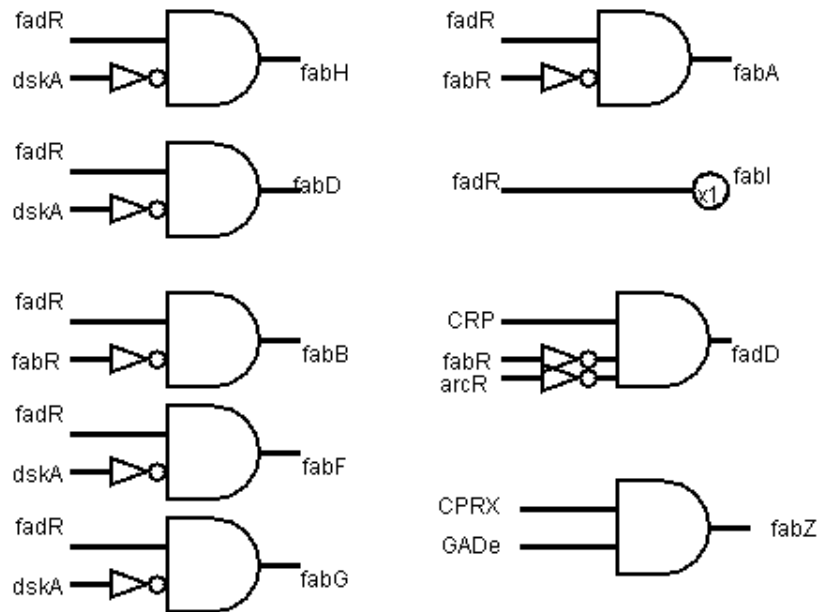


FONTE: Elaborada pelo autor.

Nesta simulação, temos o nível máximo de ativação dos genes para estes genes estudados, pois os inibidores foram desativados como mostra a imagem figura 17. Essas relações permitem organizar e perceber as interações entre os genes de maneira mais simplificada. Foi identificado que o gene FadR é o *attractor* da via, pois é por ele que ocorre a maior quantidade de fluxo, fazendo assim, a conectividade entre diferentes partes da rede. Isso condiz com a sua função de ativador e repressor e seu relacionamento com o transporte de enzimas.

Diante dos resultados da figura 17 percebemos uma inconsistência no gene central fadR, pois ele deveria inibir o gene fadD. Neste caso, precisaremos analisar a plataforma e as regras atribuídas para essa relação para entender o motivo dessa inconsistência.

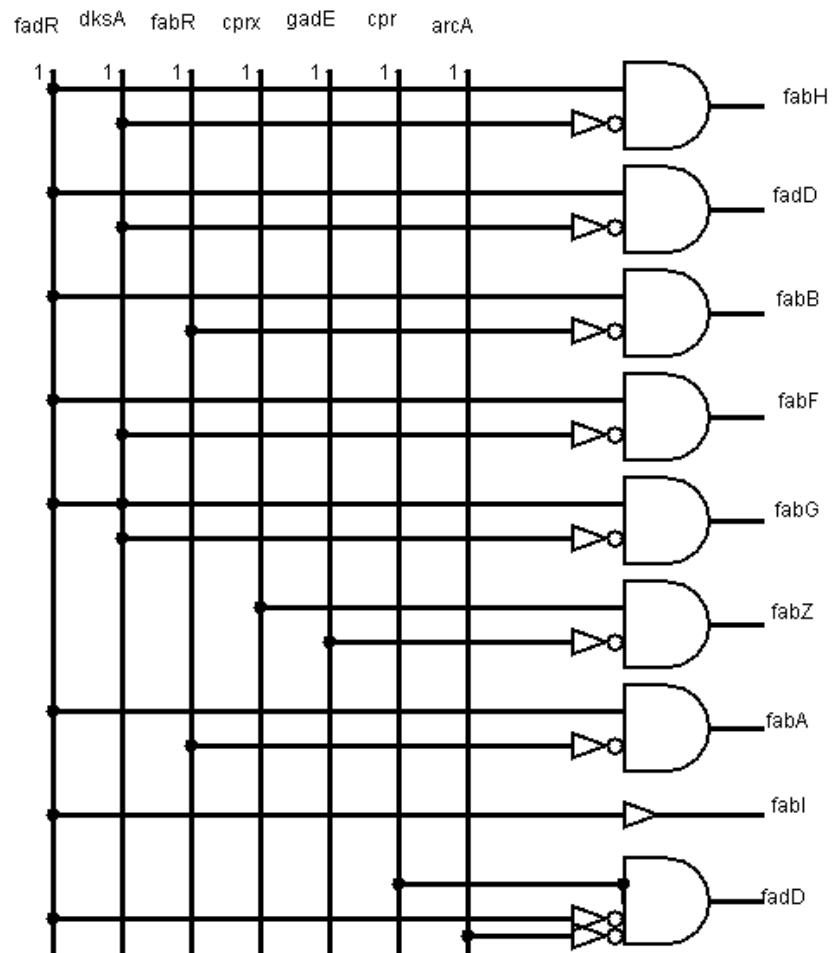
FIGURA 18: PORTAS PARA CADA REGRA DOS GENES



FONTE: Elaborada pelo autor

O objetivo desses circuitos é mostrar que é possível aumentar o nível da representação da modelagem booleana a partir das tabelas-verdade que são geradas com as regras. Através desses circuitos que estão englobados na área de biologia sintética, é possível simular o comportamento dos componentes da via rapidamente inserindo fatores externos como luz e temperatura. Para este trabalho não foi feita essa simulação, apenas as possíveis ligações entre os genes. Na figura seguinte, é apresentado uma nova forma de representação para os dados.

FIGURA 19: Representação das entradas lógicas



FONTE: Elaborada pelo autor

## 5 CONCLUSÃO

Com a elaboração das relações entre os genes podemos concluir que de fato a modelagem com redes booleanas oferece simplicidade entre as relações e entre os genes. A escolha do organismo modelo *E. coli* k-12 facilitou a busca pelos genes estudados, principalmente com a utilização do banco de dados RegulonDB, exclusivo para a *E. coli* k-12.

O processo de extração dos dados do KEGG e simplificação através do Cytoscape é demorado devido à ausência de uma ferramenta específica para este processo. No entanto, realizar o processo manualmente ajuda na adaptação e a entender melhor o fluxo das informações, porém, para vias maiores o processo manual pode se tornar muito longo ou inviável.

O estudo do organismo é a parte mais importante para o começo da modelagem booleana. Não se pode aplicar as regras booleanas sem o entendimento das relações gênicas. Após a análise do banco de dados da *E. coli* k-12 e identificação dos ativadores e inibidores, foi possível realizar a relação entre eles e partir para a definição das regras.

Com esses dados podemos concluir dois dos principais objetivos da modelagem booleana: simplicidade e abstração. A simplicidade se deu pelo fato da relação acontecer de forma bem direta com as reações implícitas. Este mesmo fato permite identificar a abstração do modelo. Por ser um modelo simples não é possível identificar reações e componentes mais internos

Portanto, o esquema realizado mostrou uma maneira de simplificar as relações entre os genes do banco de dados KEGG para a via de regulação dos ácidos graxos e como podemos realizar as simulações entre os componentes da via utilizando uma modelagem booleana. Este processo pode ser repetido para o estudo de outras vias.

Diante do exposto, é interessante continuar a exploração dessa área em busca de ferramentas, predições e novidades.

### 5.1 TRABALHOS FUTUROS

A modelagem booleana é somente uma das formas de modelagem disponíveis e a mais simples. No entanto, isto não significa que seja uma abordagem ineficiente. Para trabalhos futuros seria interessante uma otimização em determinadas partes do processo.

Uma dessas otimizações possíveis seria na parte de extração de dados do KEGG para a simplificação da rede gênica. Para a via deste trabalho foi possível realizar o processo manualmente, porém para vias maiores poderá ser inviável. Com

isso, seria interessante criar um *pipeline* pelo menos semiautomático para contribuir com modelagens de outras vias gênicas.

Um trabalho recente liderado pelo Instituto de Tecnologia de Massachusetts – MIT (do inglês, Massachusetts Institute of Technology) (disponível em: [www.cellocad.org](http://www.cellocad.org)) recebe sequências de nucleotídeos e converte essas sequências em representação binária, e, a partir disso, realiza combinações entre elas para determinar as portas lógicas e o melhor circuito que representa aquela entrada.

Sendo assim, um trabalho futuro para ser realizado a partir da modelagem booleana seria utilizar a plataforma desenvolvida pelo MIT ou outra similar para concluir a parte final deste trabalho e gerar um circuito de acordo com as vias do organismo estudado.

## REFERÊNCIAS

- ALBERT, R., WANG, R. **Discrete Dynamic Modeling of Cellular Signaling Networks**. Elsevier, v 467, Pensilvânia, EUA, 2010.
- ANEMAET, I.G., BEKKER, M., and HELLINGWERF, K.J. Algal photosynthesis as the primary driver for a sustainable development in energy, feed, and food production. *MarBiotechnol* (NY) 12, 619-629. (2010).
- BENNER, SA., SISMOUR, AM. **Synthetic Biology**. Jul, 2005. Disponível em : <https://www.ncbi.nlm.nih.gov/pubmed/15995697>.
- BERG. JM; TYMOCZKO JL, STRYER L. **Biochemistry**. 5th edition. New York: W H Freeman; 2002.
- BIOLOGY, Synthetic. **Applying Engineering to Biology**. Disponível em: <<http://www.synbiosafe.eu/uploads///pdf/EU-highlevel-syntheticbiology.pdf>>. acesso em: 10 dez 2017.
- BIOLOGY, Synthetic. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0959440X08000936#fig1>>. Acesso em: 10 dez 2017.
- BLOMINGDALE, P., **Boolean Network Modeling in Systems Pharmacology**. V 45, n 1, p. 159-180, Fev., 2018.
- BONATO, D. et al., **Genômica Final Web**. Genômica e Biologia de Sistemas. Cap 13. 2015.
- BROADHURST, C.L., et al., **Brain-specific Lipids from Marine, Lecustrine, or Terrestrial Food Resources**. (2002)
- CANNON WB. **Bodily Changes to Pain, Hunger, Fear and Rage**, 2nd ed. New York and London: D. Appleton and Co, 1929.
- CASTRO, Socorro Gama et al. RegulonDB (version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. **Nucleic Acids Research**, Cidade, v. 36, n. 1, p.120-124, jan. 2008.
- CRICK. F. **Central dogma of molecular biology**. *Nature* 227 (5258): 561–3. (August 1970).
- COMMISSION, European. **Opinion on Synthetic Biology I Definition**. Disponível em: <[https://ec.europa.eu/health/sites/health/files/scientific\\_committees/emerging/docs/scenihr\\_o\\_044.pdf](https://ec.europa.eu/health/sites/health/files/scientific_committees/emerging/docs/scenihr_o_044.pdf)>. Acesso em: 12 dez 2017.
- DUBROVA, Kauffman. **Networks: Analysis and Applications**. Ieee Xplore. San Jose, nov. 2005.

FACHINI, A. **Avaliação de Métodos de Inferência de Redes de Regulação Gênica**. São Paulo, 2016. Disponível em: <http://www.teses.usp.br/teses/disponiveis/3/3141/tde-18012017-084639/en.php>. Acessado em jan, 2018.

GENOMICS, Synthetic. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0958166912000298>>. Acesso em: 18 dez 2017.

GIRI, M. et al., **Equivalent Model of P53 Pathway Using Boolean Network**. International Journal of Scientific & Engineering Research, V. 8, n3, Mar, 2017. 8, Issue 3.

HALL, John E. **Tratado de Fisiologia Médica**. 13. Ed. Elsevier Brasil, 2017. Cap. 16, p. 1176.

HERDEWIJN, P., MARLIERE, P. Toward. **Safe genetically modified organisms through the chemical diversification of nucleic acids**. Chemistry & biodiversity 6, 791-808. (2009).

HOBOM, B. **Surgery of genes. At the doorstep of synthetic Biology**. Medizin. Klinik 75, 14–21 (1980).

KACHALO, S. et al., NET-SYNTHESIS: a software for synthesis, inference and simplification of signal transduction networks. **Disponível em** <https://www.ncbi.nlm.nih.gov/pubmed/18033793>. Acesso em jan 2018.

KANEHISA, Minoru; GOTO, Susumu. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, Cidade, v. 28, n. 1, p.27-30, jan. 2000.

KARLEBACK, G. **Modelling and analysis of gene regulatory networks**. Nature, Tel Aviv, Israel, v. 9, p.770-780, out. 2008.

KAUFFMAN, Stuart. "Metabolic stability and epigenesis in randomly constructed nets," **Journal of Theoretical Biology**, vol. 22, p. 437–467, 1969.

KESHAVA, R. et al., **Synthetic Biology: overview and Applications**. 2017. Disponível em: <https://www.sciencedirect.com/science/article/pii/B978012804659300004X>. Acessado em Fevereiro, 2018.

KITANO, Hiroaki. Systems Biology: A Brief Overview. **Science**, Tóquio, v. 295, n. 1, p.1662-1664, 1 mar. 2002.

KOBAYASHI, K., E HIRAISHI, K., **Design of Probabilistic Boolean Networks Based on Network Structure and Steady-State Probabilities**. v. 28, n 8, Ago, 2017.

KOMAROFF, L. et al., **A Bacterial Clone Synthesizing Proinsulin**. National Academy of Sciences. V 35, n 8. Ago, 1978.

LABORATORY, European Molecular Biology et al. UniProt: a hub for protein information. **Nucleic Acids Research**, Newark, v. 43, p. 205-212, out. 2014.

LAL, S. et al., **Web Semantic for Textual and Visual Information Retrieval**. State-of-the-art Information Retrieval Tool for Biological Resources. Ed. IGI Global. Cap 10. fev 2017.

LI, F. et al., **The Yeast Cell-cycle Network is Robustly Designed**. Proceeding of the National Academy of Sciences of the United States of America. San Diego, EUA. Fev, 2004.

LIIVING, Tiina. **Biochemical Fundamentals**. In: KOCK, Ina. **Modeling in System Biology: The Petri Net Approach**. Frankfurt: Springer, 2011. Cap. 2. p. 19-20.

MONOD J; CHANGEUX JP; JACOB F. Allosteric proteins and cellular control systems, **Journal of Molecular Biology**, Paris, v. 6, n. 4, p. 306-329, abril, 1963.

MONTAGUE, M. et al. **Synthetic Genomics: potencial and limitations**. Boudeaux, França. v 23, n 5, p. 659-665. out, 2012.

NATURE, disponível em: <https://www.nature.com/subjects/metabolic-engineering> . Acessado em jan, 2018.

OGATA, Hiroyuki et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. **Nucleic Acids Research**, Cidade, v. 27, n. 1, p.29-34, jan. 1999.

OSBOURN, AE., e FIELD, B., **Operons**. 2009. Disponível em: <https://www.ncbi.nlm.nih.gov/pubmed/19662496>. Acessado em jan, 2018.

PEREIRA, L. Disponível em: <http://slideplayer.com.br/slide/10453419/>. Acessado em Jan, 2018.

REDDY, V. et al., **Petri Net Representations in Metabolic Pathway**. Napperville, EUA. 1993,

REGULON DB. 2017. Disponível em: <<http://regulondb.ccg.unam.mx/index.jsp>>. Acesso em: 29 out. 2017.

RIGOUTSOS, Isidore; STEPHANOPOULOS, Gregory. **Systems Biology and Synthetic Biology: Network, Models and Applications**. 2. ed. Nova York: Oxford, 2007. 335 p.

SAADATPOUR, A., ALBERTS, R. **Boolean Modeling of Biological Regulatory Networks: a methodology tutorial**. Elsevier, Nov, 2013.

SANDOVAL, N.R., et al., **Strategy for directing combinatorial genome engineering in Escherichia coli**. *Proceedings of the National Academy of Sciences of the United States of America*. (2012).

SIMÃO, A. **Aplicação da análise de mutantes no contexto do teste e validação de redes de Petri coloridas**. USP, São Carlos. 2004.

DEMURI, Gouthan; NIELSEN, Jens. **Yeast as a Prototype for Systems Biology**. *Systems Biology and Synthetic Biology*, Lyngby, v. 96, n. 10, p.287-353, jan. 2009.

VICENTE, F. **Integração de dados na Inferência de Redes de Genes: avaliação de informações biológicas e características topológicas**. São Paulo, mai, 2016.

TOMIYAMA, M., **Modelagem e Simulação de Processos Biológicos Usando Redes de Petri Predicado Transição Diferenciais**. Uberlândia. 2007. Disponível em: <https://repositorio.ufu.br/handle/123456789/12600>. Acessado em Janeiro, 2018.

WANG, Ruy Sheng. **Boolean modeling in systems biology: an overview of methodology and applications**. *Iop Science*. Pennsylvania, set. 2012.

WELCH, D.F. **Application of cellular fatty acid analysis**. *Clinical Microbiology Reviews*, v.4, n.4, p. 422-438, 1991.

ZHANG, Y et al., **Stochastic Model of Yest Cell-cycle Network**. *Physica D: non Linear Phenomena*. V219, p. 35-39. Pequim, China. jul, 2006.