

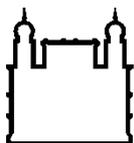
MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

Mestrado em Biologia Computacional e Sistemas

RECONSTRUÇÃO DA REDE METABÓLICA DA *PSEUDOMONAS*
AERUGINOSA CCBH4851

THIAGO GIANNINI RAMOS

Rio de Janeiro
Agosto de 2018



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

THIAGO GIANNINI RAMOS

Reconstrução da Rede Metabólica da *Pseudomonas aeruginosa* CCBH4851

Dissertação apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Mestre em Biologia Computacional e Sistemas

Orientador (es): Prof. Dr. Fabricio Alves Barbosa da Silva
Prof. Dr. Márcio Argollo Ferreira de Menezes

RIO DE JANEIRO

agosto de 2018

Ramos, Thiago Giannini.

RECONSTRUÇÃO DA REDE METABÓLICA DA *PSEUDOMONAS AERUGINOSA* CCBH4851 / Thiago Giannini Ramos. - Rio de Janeiro, 2018.
123 f.; il.

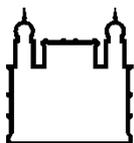
Dissertação (Mestrado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2018.

Orientador: Fabrício Alves Barbosa da Silva.

Co-orientador: Márcio Argollo Ferreira de Menezes.

Bibliografia: f. 105-112

1. Modelagem Computacional. 2. Redes Metabólicas. 3. Sistemas de Informação. 4. Ferramentas Computacionais para Biologia de Sistemas. 5. Biologia de Sistemas. I. Título.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

AUTOR: THIAGO GIANNINI RAMOS

**RECONSTRUÇÃO DA REDE METABÓLICA DA *PSEUDOMONAS AERUGINOSA*
CCBH4851**

**ORIENTADOR (ES): Prof. Dr. Fabrício Alves Barbosa da Silva
Prof. Dr. Márcio Argollo Ferreira de Menezes**

Aprovada em: 27 / 08 / 2018

EXAMINADORES:

Prof. Dr^a. Ana Paula D'Alincourt C. Assef (LAPIH/IOC/FIOCRUZ) – Presidente

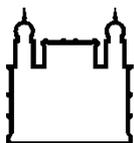
Prof. Dr. Rodolpho Mattos Albano (DB/UERJ) – Revisor.

Prof. Dr. Marcelo Ferreira da Costa Gomes (PROCC/FIOCRUZ)

Prof. Dr^a. Marisa Fabiana Nicolás (LNCC/MCT) – Suplente.

Prof. Dr. Marcos Paulo Catanho de Souza (BCS/IOC/Fiocruz) – Suplente.

Rio de Janeiro, 27 de agosto de 2018.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

Anexar a cópia da Ata que será entregue pela SEAC já assinada.

Dedico ao meu filho e minha esposa,
pois minha família é a minha fonte de
felicidade, energia e sabedoria.

AGRADECIMENTOS

À minha família, a Júlia e o Jorginho, pelo apoio, carinho e compreensão incondicional.

Aos meus pais, Edilson e Regina (Maria), os pais são os verdadeiros e eternos guardiões.

Aos meus sogros, Sérgio e Betânia, JG e Carol, são verdadeiros pais.

Ao meu afilhado, João Pedro (patuléia) e novamente ao Jorginho (cuíca), afinal, sem os momentos de descontrações, não há mestrado possível.

Ao meu amigo Fábio. Ele merece uma linha exclusiva de agradecimento, foi crucial naqueles momentos em que pensei em desistir.

Aos amigos, pela amizade, amigos são para o que der e vier.

Aos meus orientadores, Fabrício Silva e Márcio Argollo, pela oportunidade, disponibilidade, dedicação e atenção ao longo desses anos.

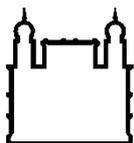
Aos integrantes do grupo de pesquisa do qual faço parte, pessoas que fizeram parte da minha formação.

A todos na coordenação da BCS, especialmente o Ernesto e a Rose, pela atenção e cuidado para com os alunos do programa, são sempre atenciosos.

À Fundação Oswaldo Cruz, pela oportunidade de crescimento profissional, com o suporte financeiro.

“Na vida, nada é fácil, nada vem de graça, nada é como imaginamos, mas se não lutarmos, pelo que realmente queremos, o que não é fácil, se torna impossível”

Jean Pierre Ginhoux



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

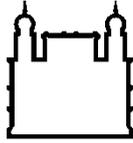
RECONSTRUÇÃO DA REDE METABÓLICA DA *PSEUDOMONAS AERUGINOSA* CCBH4851

RESUMO

DISSERTAÇÃO DE MESTRADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Thiago Giannini Ramos

As altas taxas de mortalidade e morbidade associadas a infecções nosocomiais (adquiridas em ambientes hospitalares) causadas por bactérias multi-resistentes a antibióticos constituem um problema de saúde pública no Brasil. Nesta dissertação estudou-se um desses patógenos, *Pseudomonas aeruginosa*, cuja cepa altamente virulenta CCBH4851 está presente em vários hospitais públicos brasileiros. A partir de dados genômicos desta cepa, gerados no Laboratório de Pesquisa em Infecção Hospitalar do Instituto Oswaldo Cruz (LAPIH/IOC/FIOCRUZ), e de modelos computacionais que integram a visão sistêmica de processos celulares a informação no nível molecular, a pesquisa realizada objetivou reconstruir a rede metabólica da *P. aeruginosa* CCBH4851. Para auxiliar o processo de reconstrução foram criados dois sistemas WEB: o primeiro obtém um mapa metabólico inicial da CCBH4851 a partir da coleção de seus genes anotados e armazenados em um arquivo de extensão GBK e o mapa de referência da cepa PAO1. O segundo sistema auxilia na curadoria manual do mapa reconstruído a partir da criação de um espaço virtual para discussão com especialistas. Entre os resultados obtidos com a aplicação desta metodologia, destacam-se: 1) primeira reconstrução do mapa metabólico da *P. aeruginosa* CCBH4851 capaz de gerar biomassa; 2) adaptação do protocolo T&P para possibilitar tal reconstrução; 3) os sistemas web que possibilitaram tais resultados; 4) uma base de conhecimento para armazenamento de informações dos modelos reconstruídos. Com estes resultados, espera-se ter contribuído para a geração de modelos futuros da *P. aeruginosa* CCBH4851 que auxiliem na descoberta de novas opções terapêuticas.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

RECONSTRUCTION OF THE METABOLIC NETWORK OF *PSEUDOMONAS AERUGINOSA* CCBH4851

ABSTRACT

MASTER DISSERTATION IN COMPUTATIONAL BIOLOGY AND SYSTEMS

Thiago Giannini Ramos

The high mortality and morbidity rates associated with nosocomial infections (acquired in hospital environments) caused by multidrug-resistant antibiotic bacteria are a serious public health problem in Brazil. In this dissertation, we studied one of these pathogens, *Pseudomonas aeruginosa*, whose highly virulent strain CCBH4851 is present in several Brazilian public hospitals. Based on genomic data from this strain, generated by the Hospital Infection Research Laboratory of the Oswaldo Cruz Institute (LAPIH / IOC / FIOCRUZ), and from computational models that integrate the systemic vision of cellular processes and information at the molecular level, the main objective of this research is to reconstruct the metabolic network of *P. aeruginosa* CCBH4851. To support the reconstruction process two WEB systems were created: the first obtains an initial metabolic map of the CCBH4851 strain from the collection of its annotated genes, stored in a GBK extension file, and q reference metabolic network of the PAO1 strain. The second system supports the manual curation of the reconstructed network by creating a virtual space for discussion with specialists. Among the results obtained with the application of this methodology, the following stand out: 1) first reconstruction of the metabolic network of *P. aeruginosa* CCBH4851 capable of generating biomass; 2) adaptation of the T & P protocol to enable such reconstruction; 3) the web systems that enabled such results; 4) a knowledge base for storing information from the reconstructed models. With these results, we intend to contribute to the generation of future models of *P. aeruginosa* CCBH4851 that will be useful for the discovery of new therapeutic options.

ÍNDICE

RESUMO	IX
ABSTRACT	X
1 INTRODUÇÃO	18
1.1 Pseudomonas aeruginosa CCBH4851	19
1.2 Modelagem de Redes Metabólicas em Escala Genômica	20
1.2.1 Reações químicas e Enzimas.....	21
1.2.2 Vias Metabólicas.....	23
1.2.3 Protocolo de Reconstrução de Redes	23
1.3 CONstraint-Based Reconstruction and Analysis – COBRA	26
1.3.1 Flux Balance Analysis (<i>FBA</i>)	28
1.3.2 Análise e Preenchimento de Lacunas (Gapfilling)	31
1.3.3 COBRApy (COBRA for Python).....	31
1.4 Representação e Armazenamento de Dados Biológicos	32
1.4.1 Systems Biology Markup Language (SBML)	32
1.4.2 Bases de Dados Biológicos	33
1.5 Sistemas WEB e Aspectos Tecnológicos	34
1.5.1 Java EE	35
1.5.2 BioJava.....	36
1.5.3 Jsoup (HTML Parser).....	37
1.5.4 Jsbnl	38
1.6 Justificativa	38
2 OBJETIVOS	40
2.1 Objetivo Geral	40
2.2 Objetivos Específicos	40
3 MATERIAL E MÉTODOS	41
3.1 Etapa 1 - Reconstrução inicial (draft)	42
3.1.1 NETGen – Sistema <i>web</i> de Geração de Redes Metabólicas.....	43
3.2 Etapa 2 - Refinamento da rede metabólica por especialistas	61
3.2.1 CurSystem – Sistema <i>WEB</i> de Curadoria Assistida de Redes Biológicas	63

3.3	Etapa 3 - Avaliação inicial através de métodos matemáticos e computacionais	75
3.4	Base de Conhecimento da <i>Pseudomonas aeruginosa</i> CCBH4851	76
4	RESULTADOS	80
4.1	Reconstrução Inicial	80
4.2	Gaps para geração de Biomassa e Curadoria Assistida	82
4.3	Modelos mPAO1_v2008 x mPAO1_v2017 x mCCBH4851_v2	86
4.4	Reações mapeadas e agrupadas por subsistema.....	87
5	DISCUSSÃO	90
5.1	Aspectos Computacionais	90
5.1.1	Desempenho de processamento e Importância da base de conhecimento	91
5.2	Aspectos do Modelo Biológico da <i>P. aeruginosa</i> CCBH4851	93
5.2.1	Sobre a escolha do Modelo de Referência mPAO1_v2017.....	94
5.2.2	Reações de Biomassa (pseudo-reações)	95
5.2.3	Exemplo de análise com o Modelo mCCBH4851_v2	96
5.2.4	Validação de reações inferidas e dificuldades encontradas	98
6	CONCLUSÕES	102
7	REFERÊNCIAS BIBLIOGRÁFICAS	104
	APÊNDICES E/OU ANEXOS	112
7.1	Apêndice A – Equação de Biomassa.....	113
7.2	Apêndice B – Código fonte dos sistemas	114
7.3	Apêndice C - Manual do CurSystem	115
7.4	Apêndice D - Análise Complementar da Rede Metabólica mCCBH4851_v2.....	116
7.5	Apêndice E – Saída da análise do modelo mCCBH4851_v2 feita no FindTargetsWeb	123

ÍNDICE DE FIGURAS

Figura 1 - Exemplo de reação química retirado de (25).	22
Figura 2 - Protocolo de Reconstrução de Redes Metabólicas retirado de (7).	26
Figura 3 - Conceitos fundamentais dos métodos <i>COBRA</i> retirado de (32).	27
Figura 4 - Rede metabólica artificial de uma célula hipotética retirado de (15).	29
Figura 5 - Representação do modelo hipotético através de equações diferenciais (a) e um matriz estequiométrica (b) retirado de (15).	30
Figura 6 - Passos da metodologia adaptada de Reconstrução de Redes Metabólicas feitas a partir do protocolo <i>T&P</i> (7)	42
Figura 7 – Arquitetura conceitual do Sistema <i>WEB NETGen</i>	44
Figura 8 - Exemplo de Referência Cruzada de IDs de Genes (38).	45
Figura 9 - Exemplo de Referência cruzada (cross-references)	46
Figura 10 - Método que interconecta o <i>Python</i> e executa o <i>FBA</i> do <i>COBRAPy</i> internamente no <i>NETGen</i>	50
Figura 11 - <i>Script</i> em <i>Python</i> para a simulação do cálculo de Biomassa do modelo recém gerado	51
Figura 12 - <i>Screenshot</i> do <i>Workspace</i> do <i>NETGen</i> mostrando as <i>APIs</i> referenciadas	53
Figura 13 - Unidade de apresentação para <i>login</i> do <i>NETGen</i>	54
Figura 14 – Unidade de apresentação principal do <i>NETGen</i>	54
Figura 15 - Ambiente de execução de Reconstruções Metabólicas	55
Figura 16 - Janela de Diálogo com o Assistente de Reconstrução com as três etapas de preparação para a execução de uma nova reconstrução	56
Figura 17 - Aba de inserção de informações e upload dos arquivos necessários para a reconstrução	57
Figura 18 - Aba para inferência manual de reações na reconstrução do novo modelo	58
Figura 19 - Janela de resumo e confirmação de configuração da reconstrução que será executada	59
Figura 20 - Unidade de execução em background da reconstrução	60
Figura 21 - E-mail de resultados da nova reconstrução do modelo alvo	61
Figura 22 - Fluxo de trabalho de Curadoria Assistida da Rede Metabólica	63
Figura 23 - Espaço de Trabalho com Classes (modelo, controle e visão) e <i>Libraries</i> utilizadas no projeto do <i>CurSystem</i>	64

Figura 24 – (a) Modelo de Interação das Páginas de Login e (b) Cadastro de Desafios	65
Figura 25 - Modelo de interação da Página <i>Challenges Discussions</i>	66
Figura 26 - Modelo de Interação da Página <i>Close Challenges</i>	67
Figura 27 - Unidade de apresentação para <i>login</i> do CurSystem	68
Figura 28 – Unidade de apresentação “ <i>Home</i> ” (página principal)	68
Figura 29 - Página de Desafios (unidade de apresentação “ <i>List of challenges</i> ”)	69
Figura 30 - Janela de diálogo para cadastro de desafios (unidade de apresentação “ <i>New challenge</i> ”)	70
Figura 31 – Página de respostas para soluções possíveis de cada desafio (unidade de apresentação “ <i>Challenge discussions</i> ”)	71
Figura 32 - Janela de diálogo para cadastro da resposta (unidade de apresentação “ <i>Answer to</i> ”)	72
Figura 33 - Janela de diálogo para cadastro de comentários de respostas (unidade de apresentação “ <i>Discussion list</i> ”)	73
Figura 34 – Página de validação de respostas (unidade de apresentação “ <i>Close challenges</i> ”)	74
Figura 35 - Janela de diálogo para justificar e validar resposta	74
Figura 36 - <i>Script</i> em <i>Python</i> utilizando o <i>COBRAPy</i> para análise de modelos metabólicos representados por arquivos em formato <i>SBML</i>	76
Figura 37 – Arquitetura da interconexão entre os sistemas NETGen, CurSystem e a Base de Conhecimento da <i>P. aeruginosa</i> CCBH4851	77
Figura 38 - Modelo de Entidade e Relacionamento (ER) da Base de Conhecimento da <i>Pseudomonas aeruginosa</i> CCBH4851	79
Figura 39 - Resultados da análise da primeira rede gerada (mCCBH4851_v1) feita no pacote <i>COBRAPy</i>	82
Figura 40 - E-mail com o resultado e análises gerados pelo NETGen	82
Figura 41 - Relatório da segunda versão gerado do modelo.	84
Figura 42 - Resultado do <i>COBRAPy</i> utilizando o arquivo <i>SBML</i> da segunda versão do modelo (mCCBH4851_v2)	84
Figura 43 - Reações apontadas na análise de <i>gap</i> cadastradas no CurSystem	85
Figura 44 - Desafios cadastrados para a Modelagem da Rede de Regulação Gênica da <i>P. aeruginosa</i> CCBH8451	91

Figura 45 - Resultado de execução de uma reconstrução com pouca ou nenhuma informação na Base de Conhecimento	92
Figura 46 - Resultado de execução de uma reconstrução com a maioria das informações contidas na Base de Conhecimento	93
Figura 47 - Dados comparativos das diferentes reconstruções da <i>P. aeruginosa</i> PAO1 (iMO1056/2008, iMO1086/2015 e mPAO1/2017) retirado de (6).	95
Figura 48 - Análise executada no <i>FindTargetsWeb</i> a partir do modelo reconstruído mCCBH4851_v2 (9)	97
Figura 49 - E-mail de resultado do processamento da análise do <i>FindTargetsWeb</i> através do modelo de rede metabólica mCCBH4851_v2	98
Figura 50 - Dados completos do gene com ID NP_250713.1 da última versão do genoma da <i>P. aeruginosa</i> CCBH4851	100
Figura 51 - Resultado da análise de Ortologia via BLASTp para o gene NP_250713.1 do genoma da <i>P. aeruginosa</i> CCBH4851	100
Figura 52 - Resultados obtidos no COBRApy através de análise de meio mínimo do modelo mCCBH4851_v2	116
Figura 53 - Resultado da nova Simulação do modelo utilizando somente reações de uptake indicadas pelo COBRApy	118
Figura 54 - Resultado da análise e cálculo do Meio Mínimo do modelo mCCBH4851_v2 para gerar o valor de biomassa 1.04 mmol/g DCW/h	119
Figura 55 - Simulação de geração de biomassa no meio mínimo estrito utilizando somente as reações indicadas pelo COBRApy.	120
Figura 56 - Resultado da biomassa gerada pela análise <i>loopless</i> da Rede da <i>P. aeruginosa</i> CCBH4851 - mCCBH4851_v2	121
Figura 57 - Gráfico de Dispersão entre os dois fluxos de biomassa do modelo da <i>P. aeruginosa</i> CCBH4851	122
Figura 58 - Lista de inibidores indicados pela análise do modelo da rede metabólica da <i>P. aeruginosa</i> CCBH4851 feita pelo <i>FindTargetsWeb</i> (9).	123

LISTA DE TABELAS

Tabela 1 - Classificação de enzimas e suas respectivas funções (24,26).....	23
Tabela 2 – Dados dos diferentes modelos reconstruídos da <i>P. aeruginosa</i> PAO1 (6,12).	81
Tabela 3 – Dados mapeados da primeira rede metabólica reconstruída da <i>P. aeruginosa</i> CCBH4851 no NETGen.	81
Tabela 4 - Reações apontadas pela análise de lacunas no Modelo mCCBH4851_v1	83
Tabela 5 - Comparação entre os dois modelos gerados no NETGen	85
Tabela 6 - Dados comparativos entre os modelos mCCBH4851_v2, mPAO1_v2008 e mPAO1_v2017 (6,12).	86
Tabela 7 - Relação de reações mapeadas por subsistema do modelo mCCBH4851_v2.....	88
Tabela 8 - Relação de reações com subsistema desconhecido	89
Tabela 9 - Diferentes modelos que contém a reação de demanda energética "R_ATPM"	99
Tabela 10 - Reações de <i>uptake</i> e fluxos mínimos para geração de Biomassa no valor de 12.470970 mmol/g DCW/h indicadas na análise de meio mínimo	117
Tabela 11 – Lista de reações indicadas pelo COBRApy e utilizadas na simulação de geração de biomassa no valor de 1.04 mmol/g DCW/h para meio mínimo com restrições	119

LISTA DE SIGLAS E ABREVIATURAS

CCBH	Coleção de Culturas de Bactérias de Origem Hospitalar
SBML	Systems Biology Markup Language
EC	Enzyme Commission
KEGG	Kyoto Encyclopedia of Genes and Genomes
BRENDA	Braunschweig Enzyme Database
Java EE	Java Enterprise Edition
API	Application Programming Interface
JDBC	Java Data Base Connection
WEB	World Wide Web
XML	eXtensible Markup Language
HTML	Hypertext Markup Language
SGML	Standard Generalized Markup Language
NCBI	National Center for Biotechnology Information
IDE	Integrated Development Environment
GPL	General Public License
COBRA	Constraint-Based Reconstruction and Analysis
ABF	Análise de Balanço de Fluxo
FBA	Flux Balance Analysis
DNS	Domain Name Server
LGPL	Library General Public License
JSBML	Java Systems Biology Markup Language
COBRApy	Constraint-Based Reconstruction and Analysis for Python
DOM	Document Object Model
CSS	Cascading Style Sheets
XSS	Cross-site Scripting
URL	Uniform Resource Locator
SGBD	Sistema Gerenciador de Banco de Dados

1 INTRODUÇÃO

No Brasil, o número de infecções contraídas na rede hospitalar por bactérias da família *Pseudomonas aeruginosa* é significativo e constitui um problema de saúde pública (1). A bactéria *Pseudomonas aeruginosa* é um organismo versátil e oportunista. Seu alto grau de virulência em pacientes imunocomprometidos, somado ao seu amplo sistema de defesa contra antibióticos, são responsáveis por uma mortalidade e morbidade elevada em nosso sistema de saúde (2). O boletim da OMS (3) publicado em 2017 aponta 12 bactérias que apresentam resistência a antibióticos e preocupam os pesquisadores da área e a *Pseudomonas aeruginosa* consta no topo da lista com um nível alto de notificações. Tais características fazem dela um organismo de interesse da comunidade científica para pesquisas que possam identificar novos alvos terapêuticos (4).

Uma das abordagens de pesquisa utilizadas na Biologia para construção de conhecimento, denominada como empírica, é executada através de experimentos observáveis e sem inferência da dinamicidade em torno do sistema biológico (5). Em contrapartida, existem abordagens teóricas que visam construir modelos matemáticos / computacionais, que embora sejam precisos, muitas vezes não representam de forma fidedigna o sistema biológico real (5). Quando estas duas abordagens são tratadas com uma inter-relação estreita, os resultados se mostram mais promissores (5). Na perspectiva da abordagem empírica, é importante destacar que é um grande desafio para a biologia, estudar e entender como moléculas individuais e suas interações formam o fenótipo complexo de um organismo. Essa dificuldade está sendo superada com a evolução da plataforma computacional (computadores de alto desempenho x softwares) voltada para biologia de sistemas e pela adoção crescente de modelos matemáticos/computacionais nos projetos de pesquisa. Seguindo essa abordagem, destaca-se a modelagem computacional de redes metabólicas para a realização de análises satisfatórias relacionadas às propriedades metabólicas em escala genômica, tais como crescimento, virulência e mecanismos de resistência a antibióticos (6).

Este trabalho tem como objetivo principal a reconstrução do primeiro modelo computacional da rede metabólica da *Pseudomonas aeruginosa* CCBH4851 que simule a geração de biomassa, além de desenvolver Ferramentas de Suporte (Sistemas WEB) para apoiar a reconstrução através da adaptação do método

proposto no protocolo *T&P* (7). A cepa escolhida para a pesquisa foi encontrada e isolada pelo Laboratório de Pesquisa em Infecção Hospitalar do IOC/Fiocruz (LAPIH), a última versão do genoma fechado foi anotada pelo grupo de pesquisa em modelagem computacional de bactérias multirresistentes da Fiocruz (1,8). Para obter a primeira versão, foram implementados dois sistemas que irão auxiliar os pesquisadores nos passos do método adaptado do protocolo de reconstrução computacional *T&P*, descrito por *Thiele e Palsson* em 2010 (7). O modelo computacional nessa primeira versão possibilita testar hipóteses relacionadas com o crescimento da bactéria, incluindo experimentos *in silico*, nos quais se pode efetuar nocautes de genes em algumas vias metabólicas essenciais. Esses genes seriam candidatos em uma busca de alvos para criação de novos fármacos (9).

1.1 *Pseudomonas aeruginosa* CCBH4851

A *Pseudomonas aeruginosa* é uma bactéria que vive nos mais variados ambientes: já foi encontrada na água, no solo, em vegetais, na maioria dos alimentos, nos animais, em ambientes hospitalares e até em ambientes mais hostis à sua sobrevivência (10). Dentre suas características, podem ser citadas seu formato baciliforme, sua propriedade Gram-negativa, respiração aeróbica e indícios de atividades de fermentação (10). Seu crescimento geralmente acontece aos 37° graus Celsius, mas pode apresentar crescimento com a temperatura variando de 5°C a 42°C. Seu tamanho é aproximadamente de 1,5 a 3,0 micrometros de comprimento e 0,5 micrometros de largura e sua mobilidade é possível através da apresentação de um único flagelo (11).

A bactéria possui um genoma que é classificado como um dos maiores do mundo procariótico e que codifica um número alto de proteínas com funções regulatórias, de transporte e grau de virulência (11). Ademais, seu poder de codificação lhe confere sua versatilidade e capacidade de adaptação ao meio (12). Seu sucesso em sobreviver como patógeno vem da sua versatilidade metabólica (10). A *P. aeruginosa* possui todas as grandes faixas de classificação de fatores de virulência bacterianos, como proteínas ligadoras de ferro, endotoxinas, pili, flagelo, formação de biofilme, entre outras (11).

Porém, muito embora a *Pseudomonas aeruginosa* seja um microrganismo importante em infecções em indivíduos com comprometimento no sistema imunológico, em pessoas imunologicamente saudáveis isto raramente acontece

(1,12). Há indícios de que a hospitalização esteja relacionada com o aumento da colonização de bactérias em pacientes, principalmente em pacientes submetidos a ventilação mecânica e com queimaduras graves (11).

A cepa de *Pseudomonas aeruginosa* estudada neste trabalho foi colocada em destaque por suas características de resistência a antibióticos. A cepa CCBH4851 (pertencente ao clone denominado ST277) foi isolada de um cateter de um paciente internado em um hospital em Goiás (1). A mesma foi depositada na Coleção de Culturas de Bactérias Adquiridas em Hospitais, situada no Laboratório de Pesquisa em Infecção Hospitalar do Instituto Oswaldo Cruz. A cepa se mostrou resistente para a maioria dos antibióticos testados, com exceção da Polimixina B.

O genoma possui 6.719.789 pb, e foi sequenciado utilizando a plataforma de sequenciamento Illumina MiSeq (1). Foi construída uma biblioteca genômica através da marcação por transposon com o kit Nextera XT DNA Sample Prep e sequenciada duas vezes com os kits 300 e 500 ciclos MiSeq Reagent v2, produzindo, respectivamente, 1.445.561 e 1.232.439 leituras pareadas (*paired-end reads*). Ao todo, foram gerados cento e cinquenta contigs e sua anotação foi feita através da plataforma *Rapid Annotation* utilizando *System Technology* (RAST) v2. No total, foram anotados 6.347 genes codificadores de proteínas e 60 genes codificadores de RNA, 55 tRNAs e 5 rRNA. Em torno de 568 subsistemas foram identificados (1).

1.2 Modelagem de Redes Metabólicas em Escala Genômica

Nesta seção, são apresentados conceitos envolvendo a modelagem de Redes Metabólicas, ponto central desse trabalho e subárea da Biologia de Sistemas. Entretanto, não se pode deixar de descrever assuntos da Biologia Molecular como enzimas, reações químicas e vias metabólicas, temas básicos intimamente relacionados à área de modelagem de redes metabólicas.

Os estudos feitos a partir da biologia molecular e os avanços das técnicas de sequenciamento de genomas (algoritmos de bioinformática), tornaram viável a reconstrução metabólica a partir de uma sequência genômica (7,13,14,15). Com tais informações cria-se a possibilidade de construir mapas topológicos que representam os caminhos metabólicos, onde a geração de um conjunto selecionado de produtos, dado um conjunto particular de substratos disponíveis, é escrita em termos de reações.

Um modelo metabólico que consiga reproduzir matematicamente esses caminhos metabólicos podem ser usados para responder, por exemplo, a capacidade de crescimento do organismo que está sendo representado, uma das hipóteses centrais por trás da ideia de vida como conjuntos auto replicantes (16,17).

Um micro-organismo precisa de uma fonte básica de energia e um conjunto de metabólitos acumulados em quantidade definidas pela composição celular, e posteriormente produzirem sua estrutura física para criação de uma nova célula (18). Este processo pode ser representado também no modelo metabólico, é feito através da criação de uma pseudo-reação, que usualmente, é denominada reação de biomassa (19). A reação de biomassa é formada pelos metabólitos (substratos com índices estequiométricos definidos) relativos da composição celular.

Um processo de reconstrução metabólica suficientemente preciso, é um modelo que represente o desenho da via dos fluxos metabólicos de reação observados experimentalmente em organismos in vivo, que permita a simulação de geração de biomassa (19,20). Em organismos procarióticos simples a maximização de seus processos se traduz quase que integralmente na taxa de crescimento (21). Esse tipo de comportamento pode ser simulado através da modelagem matemática / computacional (22), que disponibiliza ferramentas como a COntstraint-Based Reconstruction and Analysis (COBRA) (13), análise baseada em restrições (descrita na seção 1.3).

1.2.1 Reações químicas e Enzimas

O mapa metabólico pode ser visualizado, de forma genérica, como o conjunto de todas as transformações químicas (reações) que um organismo pode realizar (23). Por isso, se faz necessário entender alguns fatores básicos relacionados a essas reações químicas.

Uma reação química pode ser descrita de maneira simplificada como um processo de transformação de um conjunto de metabólitos (substratos - compostos químicos) em outros metabólitos (produtos – compostos químicos) (24). A representação gráfica de uma reação química é mostrada por uma equação separada por uma seta, onde os elementos anteriores a ela são os substratos, e os elementos posteriores são os produtos (24). Por exemplo, na etapa final da glicólise, uma enzima chamada piruvato quinase transfere um grupo fosfato de um substrato (fosfoenolpiruvato) para outro substrato (ADP), gerando assim piruvato e ATP como

produtos. A equação da reação descrita é mostrada abaixo, junto com sua representação gráfica da Figura 1.

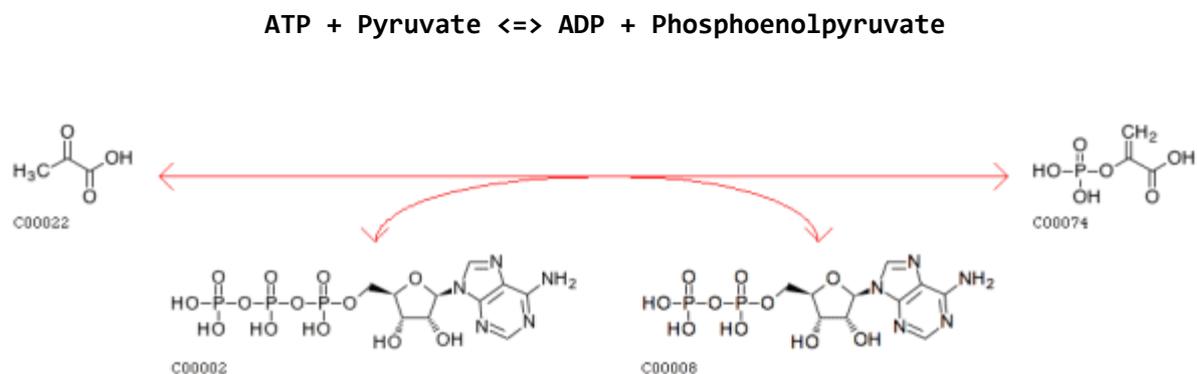


Figura 1 - Exemplo de reação química retirado de (25).

Um produto de uma reação pode ser utilizado como substrato em outra, esse comportamento produz sistemas encadeados de reações, que são classificados conforme suas cadeias e finalidades. Esses sistemas são chamados de vias metabólicas. (5,24).

A maioria das reações de um organismo são executadas lentamente, e essas reações geralmente necessitam de um catalisador para acelerar o processo. Um catalisador é um composto químico capaz de acelerar uma reação sem que ela o consuma no processo. Esses catalisadores de reações dos organismos são proteínas denominadas enzimas (24).

As enzimas são moléculas específicas e normalmente pode catalisar uma reação, ou até mais de uma em alguns casos. O metabolismo de um organismo é composto por uma grande quantidade de enzimas, incluindo algumas essenciais para manter a sobrevivência da célula (23,24). Para uma bactéria por exemplo, pode ser fatal a inativação de certas enzimas, assim prejudicando vias básicas do metabolismo. Por causa da grande quantidade de enzimas, adotou-se um padrão de nomenclatura, esse padrão chama-se *Enzyme Commission Number (EC number ou EC)* (23,26). O EC number é um código formado por quatro campos numéricos separados por ponto, e cada campo está atrelado às informações funcionais que a enzima exerce nos grupos de reações contidos no metabolismo (23,26). Por exemplo, a enzima tripeptide aminopeptidase tem o código EC 3.4.11.4, onde o número 3 representa uma hidrolase (enzima que usa a água para catalisar algumas moléculas), o número 4, indica que atua sobre as ligações peptídicas, o 11 indica que são atuantes sobre o amino-terminal de aminoácidos de um polipeptídeo, e o

último número, o 4 indica ação sobre o amino-terminal final de um tripeptídeo (27). A Tabela 1 lista os grupos funcionais de enzimas catalogadas (25).

CLASSE	FUNÇÃO
EC 1	Oxirredutases
EC 2	Transferases
EC 3	Hidrolases
EC 4	Liases
EC 5	Isomerases
EC 6	Ligases

Tabela 1 - Classificação de enzimas e suas respectivas funções (24,26).

A regulamentação do padrão de nomenclatura é de responsabilidade do *International Union of Biochemistry and Molecular Biology (UIBMB)* (27).

1.2.2 Vias Metabólicas

O conjunto de reações químicas ligadas que exercem funções importantes em um organismo é denominado Via Metabólica. Por sua vez, toda via metabólica é determinada por enzimas estabelecidas pela formação genética de sua célula (23,24). As diversas reações existentes em uma célula produzem uma gama de produtos necessários para o seu funcionamento e estes têm diferentes destinos como reagentes. Portanto, existe uma interligação complexa entre as reações de um organismo. Para a viabilidade de análises e estudos das reações que fazem parte de um determinado processo, as reações são agrupadas em diversas vias metabólicas. Essas vias metabólicas (processos) podem ser de dois tipos: vias de degradação ou de síntese de compostos químicos (24). Uma via metabólica serve como ferramenta de análise e estudo, pois fornece um nível variado de detalhes sobre processos essenciais da célula: ela pode ter mais ou menos reações, abortar as reações em um único organismo ou em um grupo, mostrar ligações com outras reações / vias, entre muitos outros aspectos (24,28).

1.2.3 Protocolo de Reconstrução de Redes

A reconstrução de uma rede metabólica pode ser executada a partir de diferentes metodologias. Até 2010, nenhuma proposta de formalização havia sido

reconhecida na comunidade científica, quando Thiele e Pálsson (7) propuseram um protocolo. O método desenvolvido por esses autores propõe passos detalhados a serem seguidos para que se possa efetuar uma reconstrução de escala genômica em alta qualidade. O documento apresenta um processo generalista para reconstrução de redes metabólicas que tem como alvo a modelagem computacional e inclui reconstrução e análises baseadas em restrições (7,13). Os passos propostos pelo protocolo são quatro e não necessariamente devem ser seguidos de acordo com uma ordenação fixa. A alternância de alguns não compromete a qualidade geral da reconstrução, entretanto, apresenta algumas limitações relativamente à finalização de todas as etapas. A indicação do documento é que as duas primeiras etapas do processo podem ser mescladas para reconstruções bacterianas com foco no refinamento da reconstrução (7,12).

Os quatro passos são:

1- Reconstrução Rascunho (*Draft*)

A primeira etapa, descrita como rascunho, consiste na obtenção de informações com ferramentas automatizadas, computacionais. O ponto de partida é a anotação do genoma do organismo alvo, nesse caso a *Pseudomonas aeruginosa* CCBH4851 (1). A partir da anotação genômica é possível realizar análises computacionais e inferir funções metabólicas. Tais análises, por sua vez, utilizam informações contidas em bases de dados globais, como *KEGG*, *BRENDA*, *Pseudomonas Genome DB*, dentre outras (6,9,13,25). Nesse processo, algumas informações podem ser incluídas de maneira errônea e outras podem estar ausentes pela falta de informações na anotação ou devido a divergências nas bases de dados. Por essa razão, cada grupo de pesquisa cria suas próprias ferramentas computacionais para lidar com esse problema, o que é negativo em se tratando da padronização da informação e formatação dos dados. Por outro lado, permite a expansão de um novo campo de pesquisa (7,12,19).

2- Curadoria Manual

Nesta etapa o processo se faz de forma oposta à primeira. O trabalho é desenvolvido de maneira completamente manual por especialistas na área (7). Posteriormente à elaboração de um modelo gerado de forma automática, os especialistas procedem a uma reavaliação e aperfeiçoamento do modelo, revisando-

o gene por gene (19). Nesse processo, os genes têm sua posição na rede metabólica confirmados e também se levantam possíveis lacunas, *gaps* de genes que deveriam estar presentes e não estão. Em particular, todas funções metabólicas e suas respectivas reações são confrontadas individualmente relativamente à literatura disponível para o organismo alvo (20). Há necessidade de participação de especialistas experientes da área, pois nem toda anotação possui um score de confiabilidade alto e a maioria dos bancos de dados bioquímicos são inconsistentes e incompatíveis (7,23).

3- Conversão Computacional / Matemática do Modelo

Depois de obter um resultado satisfatório nas etapas anteriores, ou seja, construir um modelo que atende os objetivos de pesquisa relacionados ao organismo (13,20), faz-se necessário convertê-lo para formatos matemáticos e computacionais (13). É nessa fase que são escolhidos alguns parâmetros importantes do modelo para a viabilidades de estudos pretendidos, tais como limite e abrangência de escopo fenotípico (12). É importante observar que o modelo inicial nem sempre é idêntico ao modelo final, pois com as inúmeras curadorias iterativas podem ser gerados modelos que diferem em escopo e limites no sistema (7).

4- Execução e Avaliação da Rede Metabólica

Esse estágio do processo de reconstrução consiste na verificação, avaliação e validação da rede metabólica (7). O conjunto metabólico (modelo) criado na etapa de conversão é posto em testes e um dos principais consiste em verificar se o modelo tem capacidade de sintetizar precursores de biomassa (aminoácidos, nucleotídeos trifosfatos e lipídios), componentes básicos para duplicação da célula (12,13,14). A partir destes testes são identificadas funções metabólicas inexistentes na reconstrução. Como o protocolo propõe um processo iterativo, esses *gaps* podem ser solucionados gradativamente (7). A decisão acerca da parada desse processo de iteratividade é levada em consideração o foco de estudo para qual a reconstrução se destina. Assim, o modelo se daria como concluído quando estivesse pronto para uso em análises.

A Figura 2 exhibe os passos propostos pelo protocolo T&P (7).

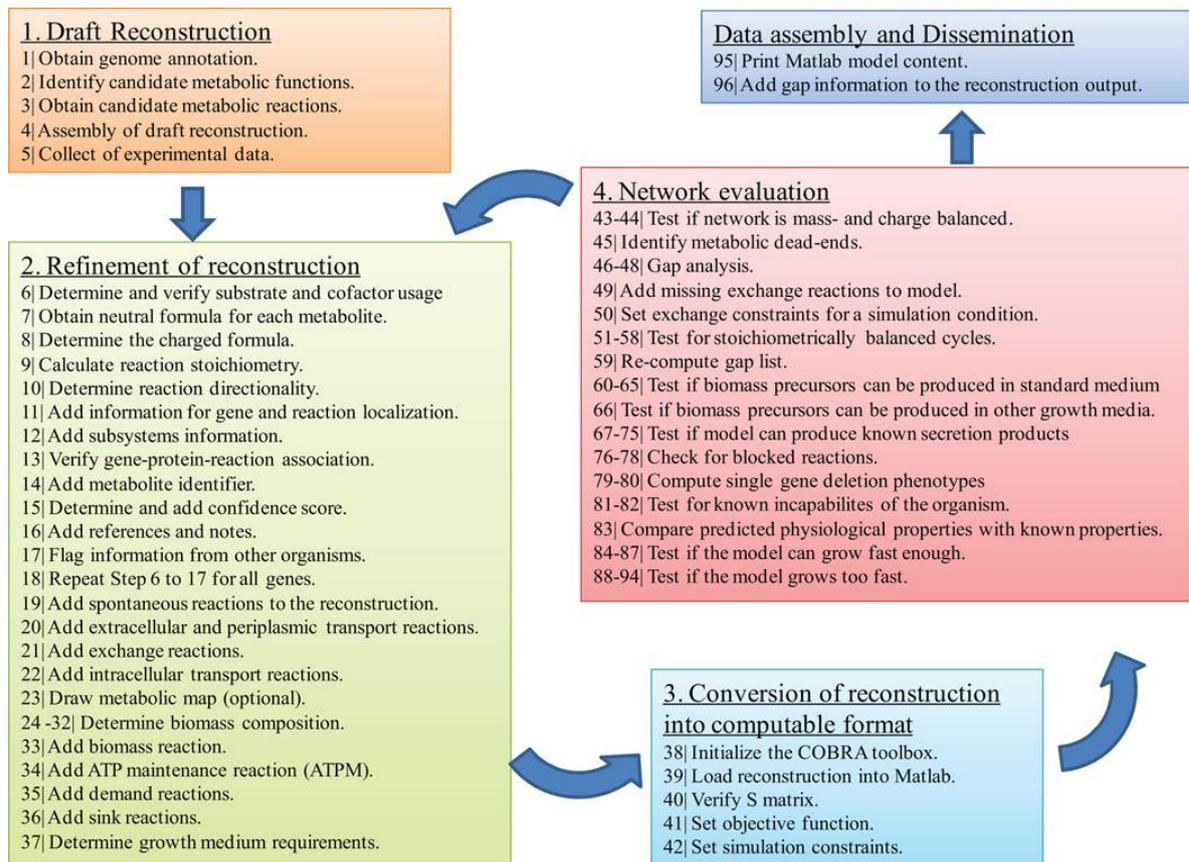


Figura 2 - Protocolo de Reconstrução de Redes Metabólicas retirado de (7).

1.3 COnstraint-Based Reconstruction and Analysis – COBRA

Após as conclusões dos primeiros projetos de sequenciamento de genomas bacterianos, o entendimento da relação genótipo-fenótipo começou a mudar (29). As sequências completas de genes formam um conjunto de informações abrangentes, relacionados aos elementos genéticos que definem um organismo. Para a compreensão desses processos celulares complexos, como o metabolismo, por exemplo, bases de conhecimento foram surgindo com variadas representações matemáticas/computacionais (5). As representações matemáticas possibilitam cálculos referentes aos estados fenotípicos com base em parâmetros genéticos e ambientais. Eminentemente, representações mecanicistas da relação genótipo-fenótipo metabólico microbiano são criadas (29,30).

Os modelos de redes metabólicas em escala genômica baseados em restrições armazenam a relação genótipo-fenótipo, considerando simultaneamente restrições fenotípicas impostas pelas leis físico-químicas e pela genética. Essa área de pesquisa surgiu a partir da percepção de que as relações quantitativas genótipo-fenótipo poderiam ser criadas utilizando genomas. A enorme produção de dados

gerados a partir dos estudos de genomas acelerou o desenvolvimento de métodos de reconstrução baseados em restrições e ferramentas de análise de redes metabólicas. Todo esse arsenal foi reunido em um poderoso *framework* analítico, o *COBRA* (*CO*nstraint-*B*ased *R*econstruction and *A*nalysis) (30,31).

O *COBRA* possui alguns conceitos importantes para o entendimento da biologia de sistemas inerente aos modelos. Tais conceitos incluem: 1) imposição de restrições físico-químicas que limitam o fenótipo computável (Figura 3.a-d); 2) a identificação e descrição matemática/computacional de pressões seletivas evolutivas, como condições ambientais (Figura 3.e) e 3) um entendimento de escala genômica do metabolismo celular caracterizando todos produtos metabólitos de uma célula (Figura 3.f). Tais conceitos podem ser satisfatoriamente representados como na Figura 2, abaixo:

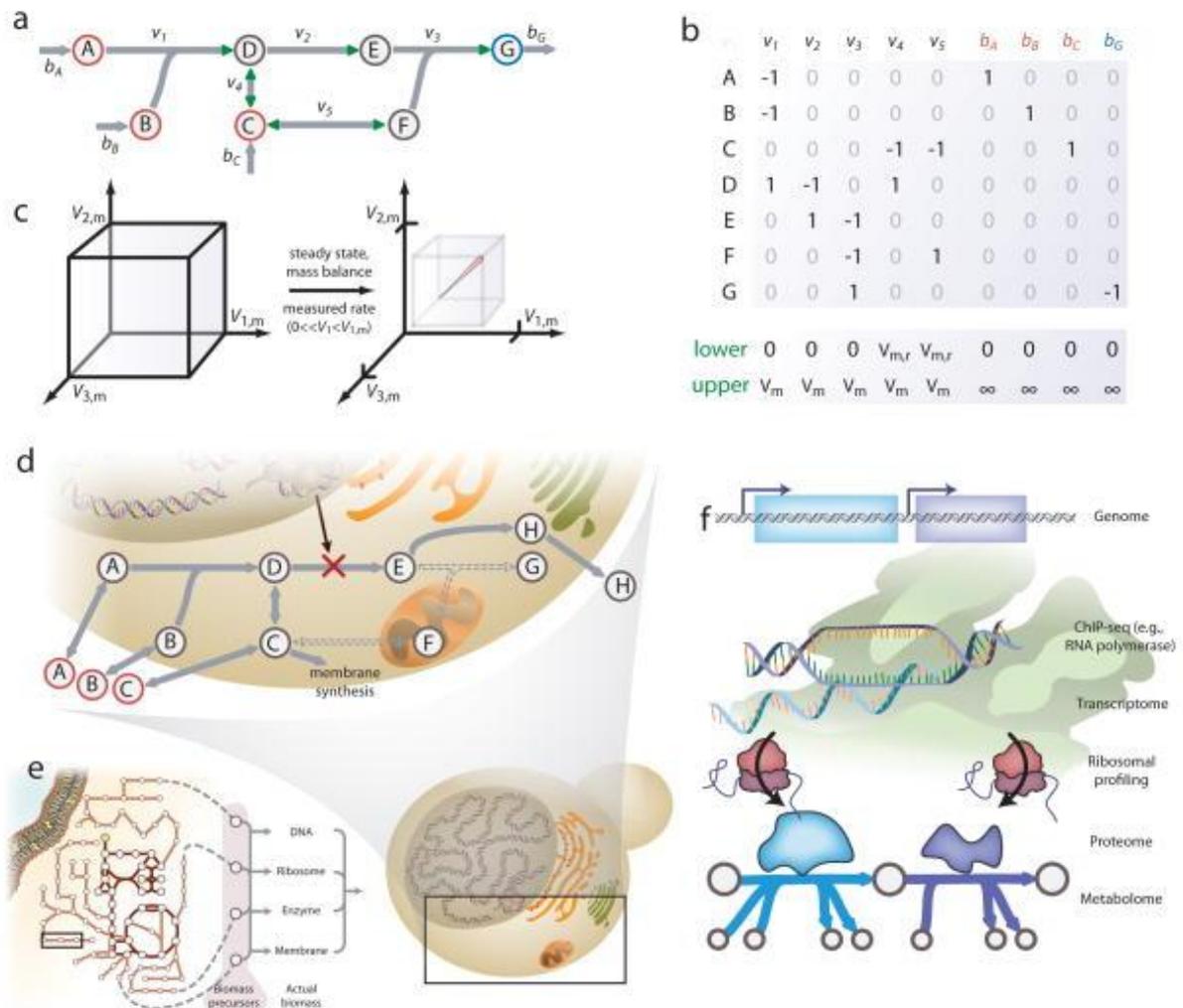


Figura 3 - Conceitos fundamentais dos métodos *COBRA* retirado de (32).

Na Figura 3.a (14,32) uma mistura complexa de moléculas (vermelho) pode interagir e reagir para produzir produtos finais. O fluxo de reações do metabolismo é representado matematicamente por uma matriz estequiométrica (Figura 3.b). As colunas representam a estequiometrias das reações e valores positivos e negativos representam produtos e reagentes respectivamente. O fluxo de cada reação é limitado pela termodinâmica e pelas capacidades catalíticas do modelo ($V_m = V_{max}$), e essas são descritas pelos limites inferiores e superiores de cada reação (Figura 3.b - verde). O “espaço de soluções” é formado por todas as distribuições possíveis e são resultados das restrições de reações existentes do modelo (Figura 3.c). Restrições adicionais podem reduzir ainda mais o “espaço de soluções”. Como exemplo desse tipo de restrição, tem-se balanço de massa, suposição de estado estacionário e taxas de consumo de metabólitos.

Uma das funções cruciais do *COBRA*, a função objetivo (reação de biomassa, nesse caso), descreve uma pressão evolutiva que possibilita o crescimento microbiano. Além disso, também detalha as demandas metabólicas que dão suporte básico à construção de produtos mais elaborados de componentes essenciais celulares, como membranas, macromoléculas e ATP (Figura 3.e).

Por fim, na Figura 3.f, a associação do genoma com o modelo metabólico é feita através da representação matemática do genoma transcrito, proteínas e reações. Esse tipo de esquema, comumente conhecido como associação Gene-Proteína-Reação (GPR), é bastante utilizado para descrever a associação de genes nos modelos e fornece uma interface rica para integração de informações (29,30,33).

1.3.1 Flux Balance Analysis (FBA)

Um dos principais métodos do *framework COBRA* é o *Flux Balance Analysis (FBA)*, também conhecido como Análise de Balanço de Fluxo (ABF) (34). Em meados da década de 90 essa metodologia começou a ser amplamente utilizada para auxiliar pesquisadores na reconstrução de modelos metabólicos. O *FBA* foi desenvolvido por Bernhard Palsson, em um laboratório na Universidade da Califórnia, em San Diego (29). O *FBA* utiliza a otimização linear (área da matemática) para calcular o fluxo de reações do modelo no estado estacionário, sem levar em consideração os parâmetros cinéticos (30). Uma das tarefas do *FBA* é a tentativa de prever a taxa de crescimento do organismo representado no modelo reconstruído. O *FBA* também simula a taxa de produção de um ou mais metabólitos

da rede. O estado da rede é representado pelo conjunto de reações, que é interpretado como sendo um único fluxo subdividido em módulos (reações). Cada reação tem a sua entrada (substratos) e sua saída (produtos) e a soma dos fluxos da entrada de metabólitos sempre será igual a soma do fluxo de saída, formando uma corrente contínua de processamento de metabólitos (34).

No *FBA*, os fluxos metabólicos são apresentados matematicamente como um problema de programação linear. Através de uma função objetivo que deverá ser otimizada, os fluxos de todas as reações são estratificados e maximizados como a solução do problema. Em modelos metabólicos, geralmente, a reação de biomassa é escolhida como função objetivo para a simulação da taxa de crescimento. Na prática, a reação de biomassa subtrai os metabólitos precursores no meio (simulado) para produzir biomassa e o organismo se duplicar. Os valores de fluxo para esse tipo de simulação são normalizados em peso celular seco por hora por milimoles por grama (mmol/g DCW/h) (34).

Apresenta-se uma implementação prática em um modelo metabólico artificial de célula hipotética para demonstrar os conceitos por trás do *FBA*. O modelo é representado por $M = 7$ metabólitos e $N = 12$ reações, das quais: a) cinco ocorrem dentro da célula, b) seis transportam metabólitos para o interior da célula e c) uma é a pseudo-reação que representa o acúmulo de metabólitos precursores de biomassa (15,35). A Figura 4 apresenta a representação do mapa metabólico relacionado aos pseudogenes.

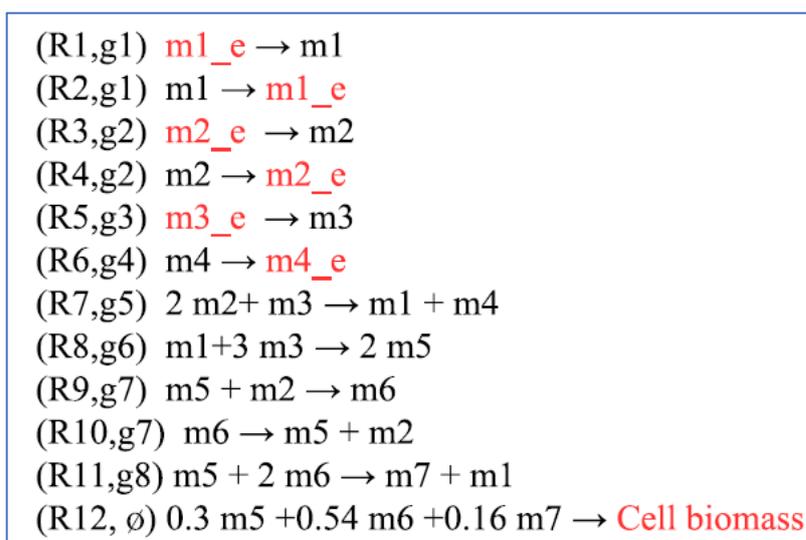


Figura 4 - Rede metabólica artificial de uma célula hipotética retirado de (15).

As reações ocorrem simultaneamente dentro da célula e as concentrações de metabólitos são alteradas no tempo como resultado da diferença entre suas taxas de consumo e síntese de todas as reações que participam do processo de geração de biomassa (15). Adotando a taxa de fluxo da reação j como f_j e estequiometria do metabólito i na reação j como S_{ij} , onde o metabólito com valor zero está ausente, com valor negativo é um substrato e valor positivo é um produto. Pode-se afirmar que a concentração de metabólito i avança ao longo do tempo. Então, a formulação matemática da descrição pode ser dada pela equação abaixo.

$$\frac{d[m_i]}{dt} = \sum_{j=1}^N S_{ij} f_j.$$

Para rede hipotética exposta acima descreve-se a evolução temporal das concentrações de metabólitos do modelo hipotético como um conjunto de equações diferenciais (Figura 5.a) e através de uma matriz estequiométrica (Figura 5.b).

a)

$$\begin{aligned} \frac{d[m1]}{dt} &= (1)f_1 + (-1)f_2 + (1)f_7 + (-1)f_8 + (1)f_{11} \\ \frac{d[m2]}{dt} &= (1)f_3 + (-1)f_4 + (-2)f_7 + (-1)f_9 + (1)f_{10} \\ \frac{d[m3]}{dt} &= (1)f_5 + (-1)f_7 + (-3)f_8 \\ \frac{d[m4]}{dt} &= (-1)f_6 + (1)f_7 \\ \frac{d[m5]}{dt} &= (2)f_8 + (-1)f_9 + (-1)f_{11} + (-0.3)f_{12} \\ \frac{d[m6]}{dt} &= (1)f_9 + (-1)f_{10} + (-2)f_{11} + (-0.54)f_{12} \\ \frac{d[m7]}{dt} &= (1)f_{11} + (-0.16)f_{12} \end{aligned}$$

b)

$$\frac{d[mi]}{dt} = \sum_{j=1}^N S_{ij} f_j$$

$$S = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & -2 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 & -3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & -1 & 1 & -1 & -0.3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -2 & -0.54 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -0.16 \end{pmatrix}$$

Figura 5 - Representação do modelo hipotético através de equações diferenciais (a) e um matriz estequiométrica (b) retirado de (15).

Uma reconstrução metabólica de um genoma, geralmente, tem milhares de reações e metabólitos, o que torna inviável a solução através de conjuntos de equações diferenciais (15). Entretanto, para organismos que crescem em estado estacionário, o tempo não é um fator que influencia as concentrações de metabólitos no meio (6,12,35,36). Portanto, o conjunto de equações podem ser compactados para a expressão:

$$\sum_{j=1}^N S_{ij} f_j = 0 \quad \forall i$$

A equação, por sua vez, pode ser escrita em uma notação reduzida como

$$\mathbf{S} \vec{f} = \vec{0}$$

Essa equação representa um conjunto linear de equações integradas em fluxo, simplificando a resolução do problema. Por causa do maior número de reações do que metabólitos, o problema possui múltiplas soluções, que reflete as diversas estratégias de geração de biomassa dos organismos (12,34,35).

1.3.2 Análise e Preenchimento de Lacunas (Gapfilling)

A análise de preenchimento de lacunas é uma etapa muito importante na reconstrução de modelos de redes metabólicas. Consiste em utilizar ferramentas computacionais / matemáticas para verificação de informações referentes ao modelo. Uma das possibilidades de análise, é executar algoritmos para comparar um modelo reconstruído (*source*) com o outro em reconstrução (*target*), e buscar reações faltantes para simulação de geração de biomassa com valor acima de zero. Este tipo de processamento existe em ferramentas como o *COBRAPy* (pacote COBRA para a linguagem Python) (14). O método faz o preenchimento do *gap* do modelo e descobre quais reações precisam ser adicionadas a um modelo para torná-lo viável (taxa de biomassa > 0). O *COBRAPy* tem uma implementação de preenchimento de lacunas muito semelhante à de Reed et al (30), onde usa-se o recurso de programação linear inteira mista para descobrir o menor número de reações que precisam ser adicionadas para uma coleção de reações definida pelo usuário, ou seja, um modelo universal (30).

1.3.3 COBRAPy (COBRA for Python)

O *COBRA toolbox* é uma das implementações do *framework COBRA*, e foi desenvolvido inicialmente para integração com o MATLAB, propondo disponibilizar computacionalmente os métodos através de uma interface de alto nível de abstração (14). O avanço e sucesso desta ferramenta fizeram do atual *COBRA toolbox* uma estrutura padrão para modelagem baseadas em restrições do modelo. Para

alavancar ainda mais a pesquisa baseada em métodos *COBRA*, houve expansão para outras linguagens de programação (13,14,33). Uma das implementações alternativas ao *COBRA toolbox* é o *COBRAPy* (14), que é um pacote de métodos *COBRA* escritos na linguagem de programação *Python*. Sua principal diferença é sua implementação, que tem como base o paradigma orientado à objetos. Portanto sua arquitetura disponibiliza os métodos *COBRA* através de classes, atributos e métodos bem definidos. Para os pesquisadores, abre novas possibilidades, pois não ficarão atrelados à uma ferramenta proprietária como o *MATLAB*. O *COBRAPy* é totalmente livre e regido pela lei GNU, qualquer pessoa pode pegar seu código, alterar e utilizar sem qualquer preocupação com direitos autorais (14).

1.4 Representação e Armazenamento de Dados Biológicos

Depois do mapeamento das vias metabólicas e suas respectivas reações, um estudo requer uma maneira de se representar e armazenar o modelo da rede metabólica. Diferentes estudos (37) utilizam um ou dois formatos de representação. Quando se trata de representação visual, pesquisadores preferem utilizar Grafos. Como material para entrada de dados em ferramentas computacionais de análise de redes, a escolha é a Linguagem de Marcação para Sistemas Biológicos (*SBML*) (37). O *SBML* também foi a escolha nesta investigação.

Outra opção de armazenamento para as informações geradas, foi a construção de uma base de conhecimento, a qual será estruturada com base em conceitos em outras bases de conhecimento, como o *KEGG* e o *Pseudomonas Genoma DB* (26,38). As tecnologias citadas são descritas nas sessões 1.4.1 e 1.4.2.

1.4.1 Systems Biology Markup Language (*SBML*)

eXtensible Markup Language (XML) é uma linguagem criada a partir da demanda de navegação e troca de informações pelos sistemas *WEB* (39). Sua base foi, assim como o *Hipertexto Markup Language (HTML)*, a *Standard Generalized Language (SGML)*. Sua sintaxe, diferentemente das outras, fornece significado ao texto marcado para detalhamento do mesmo. Duas características mais importantes definem seu grande sucesso e adoção por sistemas computacionais: primeiro, é definida como uma “linguagem de marcação extensível”, entretanto, é uma linguagem de meta-marcação, isto é, sua principal funcionalidade é definir outras linguagens de marcação; segundo, em outras linguagens de marcação, limita-se a

descrever a estrutura lógica de um documento e como essa estrutura será apresentada. Diferentemente, o *XML* pode definir a semântica de dados de forma generalista, deixando o projetista escolher tais informações (39).

A *SBML* é uma dessas linguagens baseadas em *XML* (37), com grande número de ferramentas computacionais sendo desenvolvidas nos últimos anos. Com isto, surgiu a necessidade de um padrão de troca e representação de modelos redes de reações biológicas. O padrão *SBML* atualmente encontra-se no Level 3 versão 2. Nessa versão há informações detalhadas de componentes do modelo, tais como: compartimento, substratos e produtos, reações, parâmetros globais do modelo, definição de unidades de medida, associações Gene-Proteína-Reação (GPR), e extensões de informações para sua leitura e processamento por pacotes computacionais de análise (40).

1.4.2 Bases de Dados Biológicos

As bases de dados biológicas são produtos de longas pesquisas em bioinformática e representam avanços significativos na área de modelagem de redes metabólicas (5). Todas reconstruções de redes atualmente utilizam em larga escala estes recursos (7,12,15). Nesta seção serão citadas as bases que contribuíram de forma decisiva para a conclusão deste trabalho.

1.4.2.1 Pseudomonas Genome DB e PseudoCAP

O *Pseudomonas Genome DB* é um banco de dados mantido pela comunidade acadêmica para o estudo da bactéria *Pseudomonas aeruginosa* (38). O projeto iniciou-se pelo financiamento das empresas Cystic Fibrosis Foundation e a PathoGenesis Corporation (41). O objetivo proposto era reunir uma força tarefa para o sequenciamento e a anotação do genoma da bactéria *Pseudomonas aeruginosa* PAO1. A anotação envolveu diversos pesquisadores da PathoGenesis e dos cientistas do Projeto *Pseudomonas aeruginosa* Community Annotation (*PseudoCAP*), entre outras coisas, ligações de outros dados externos a pesquisa (38,41).

A base de dados é, atualmente, repositório para diversas informações sobre *Pseudomas aeruginosa*, incluindo armazenamento de modelos de redes metabólicas curados e não curados (*draft*), anotações de genomas, links e cruzamentos de dados para outras plataformas de dados biológicos, tabelas de conversões de IDs de genes, entres outras coisas (41).

1.4.2.2 KEGG

O *Kyoto Encyclopedia of Genes and Genomes*, mais conhecido como KEGG, é outra base de dados biológicos bastante conhecida e utilizada pela comunidade científica (26). É uma base de conhecimento que fornece uma forma sistêmica de análise das funções genéticas, fornecendo um apanhado de referências cruzadas para vários outros tipos de análises. Como exemplos, podem ser mencionadas as informações genômicas que são armazenadas no banco de dados de genes, os quais por sua vez compõem inúmeras coleções de genomas parcial ou completamente sequenciados (26). O *KEGG* teve sua origem no Japão, iniciado no âmbito do Programa *Human Genome Program* do ministério da educação. Todos os dados do projeto são disponibilizados através de plataformas sistêmicas sobre serviços *web* (*APIs* de acesso) chamado de *GenomeNet service*. Os dois módulos mais importantes são o KEGG GENE e o KEGG PATHWAY.

O *KEGG GENE* são coleções de genomas e seus genes, reunindo informações adicionais de proteínas, reações e componentes (metabólitos das reações). O *KEGG PATHWAY* fornece informações para representação de funções de ordem superior em termos da rede de moléculas interagentes. Qualquer usuário pode acessar o sistema *KEGG* através de uma via funcional (modo *top-down*), ou iniciar a busca a partir de informações genômicas (modo *bottom-up*) (25).

1.5 Sistemas WEB e Aspectos Tecnológicos

Sistemas *WEB* são aplicações que utilizam tecnologias criadas para internet como interface para o usuário ter acesso. O desenvolvimento de sistemas *WEB* envolve linguagens de programação, acesso a banco de dados, entres outras tecnologias. Diferentemente, os *websites* consistem em documentos indexados e endereçados através de um servidor de nomes (DNS). No período anterior ao surgimento da internet, os tipos de sistemas existentes consistiam em aplicações para *desktop*, instaláveis em cada computador (42).

A evolução da tecnologia e dos protocolos de segurança de acesso à internet popularizaram o desenvolvimento de Sistemas *WEB*. Algumas das vantagens de se ter um sistema na plataforma web são:

- Facilidade de instalação e manutenção;
- Facilidade para acesso dos usuários;

- Descentralização dos recursos computacionais;
- Designer responsivo (adaptativo) para acesso de múltiplos dispositivos;

Para o desenvolvimento de Sistemas *WEB* existem inúmeras opções disponíveis, tais como Java (Oracle), .NET Framework (Microsoft), Python, Apache Foundation, entre outras. Java e .NET Framework são mais antigas e consolidadas atualmente, mas outras como o Python vem ganhando seu espaço (43).

1.5.1 Java EE

O *Java EE*, conhecido anteriormente como *J2EE*, é uma plataforma lançada para suprir as necessidades de implementação de um servidor de aplicações robusto. É considerado informalmente um padrão, embora, na realidade, seja definido por uma especificação. Pode-se afirmar ainda a plataforma *Java EE* como o ambiente mais relevante para desenvolvimento e execução de aplicações comerciais, pois provê escalabilidade, interoperabilidade, segurança e confiabilidade, o que facilita aos desenvolvedores de software a construção de modelos de aplicações comerciais complexas e distribuídas (42,43,44).

O padrão *Java EE* consiste em:

- Regras de construção para desenvolver aplicações comerciais.
- Uma implementação de referência para disponibilizar um modelo operacional do *Java EE*.
- Um conjunto de testes de compatibilidade para que terceiros possam verificar a qualidade do funcionamento de seus produtos com o padrão.
- Diversas *APIs (Application Programming Interface)* para disponibilizar o acesso genérico para recursos comerciais e de infraestrutura.
- Tecnologias para simplificar o desenvolvimento em *Java* no âmbito comercial, distribuído, tolerante a falhas e multicamadas, amplamente baseado em componentização.

Ainda que se pudessem executar remotamente aplicativos desenvolvidos em Java nos servidores na ocasião do lançamento da linguagem *Java*, não havia inicialmente um padrão especificado e desenvolvido pela Sun disponível para execução de tais aplicações remotamente em servidores (42,44). Isto porque o Java – originalmente denominado como *Oak* – foi criado para ser uma linguagem de

software para o desenvolvimento apenas de aplicações domésticas e outros dispositivos afins.

Com a progressão da Internet, a linguagem também evoluiu adquirindo novas ferramentas para o desenvolvimento no lado do cliente, como applets – mini aplicativos para executar no navegador e *JavaBeans* – classes com codificação padronizada. Posteriormente, diversas especificações de *APIs* foram incluídas na plataforma, assim como o *Java Database Connectivity (JDBC)* – driver *API* para conexão de banco de dados, gerenciamento e operações com banco de dados. Todas estas inovações tinham o propósito de suprir necessidades do mercado para acesso aos recursos exigidos pelas aplicações comerciais (44).

Contudo, ficou bastante explícito após a introdução do *Java* que seu uso no lado do cliente em um ambiente de sistemas baseados em navegador (sistemas *WEB*) não era viável. Assim, a Sun enfrentou vários desafios, como a latência no carregamento remoto do servidor de bibliotecas *Java* para que uma aplicação pudesse ser executada no lado do cliente (42,43).

A simplicidade do *Java*, a arquitetura independente de plataforma e o conjunto de *APIs* disponíveis, assim como sua conformação com a *WEB* bem definida, eram pontos positivos para o seu sucesso no desenvolvimento de softwares comerciais. Foram estas características juntamente com a facilidade de adicionar as tecnologias do *Java* às páginas *HTML*, padrões em *WEB Sites*, que levaram a uma ampla utilização do *Java* para desenvolvimento centrado para *WEB*. Compreendendo tal fato, juntamente com a ascensão do uso de *Java* para aplicações baseadas na *WEB*, a Sun, para o aproveitamento deste potencial, verificou que se fazia necessária a adaptação da linguagem *Java* à *WEB* para que aplicativos no lado cliente pudessem ser executados remotamente em um servidor (42,43,44).

1.5.2 BioJava

O *BioJava* é um projeto de código aberto (dirigido por comunidade) que implementa uma biblioteca (*API*) para processamento de dados biológicos através da linguagem *Java*. A *API* é dividida em inúmeros módulos (pacotes) independentes. Os módulos são divididos por funcionalidades e ou áreas de estudos da biologia. Os módulos e suas respectivas funcionalidade são (45):

- *Core* (núcleo da *API*) – fornece classes que representam sequencias de nucleotídeos e aminoácidos, e seus relacionamentos;

- Estrutura de proteínas – fornece ferramentas (métodos) para representar e manipular estruturas biomoleculares 3D e comparação da estrutura de proteínas;
- Genoma e sequenciamento – fornece classes que representam sequências gênicas e diversos formatos de arquivos com diferentes padrões de escrita;
- Alinhamento – fornece algoritmos para processamento de alinhamentos de sequências;
- *ModFinder* – fornece métodos para identificar e classificar modificações de proteínas em estruturas de proteína 3D;
- Propriedades de aminoácidos – fornece uma gama análise de propriedades físico-químicas para as proteínas;
- Desordem de proteínas – fornece algoritmos para previsão de regiões desordenadas de proteínas;
- Serviço WEB – fornece chamadas para métodos remotos de serviços de bioinformática através da utilização do protocolo *REST*;

A biblioteca *BioJava* encontra-se na versão três e a sua utilização possibilita um nível alto de suporte para o desenvolvimento de sistemas que envolvem dados biológicos em *Java*, pois fornece uma plataforma variada de processamento de informações biológicas em alto desempenho. A *API* pode ser acessada e está disponível para download em <http://www.open-bio.org>, assim como as outras implementações feitas pela *Open Bioinformatics Foundation (OBF)* (45).

1.5.3 Jsoup (HTML Parser)

O *parser* é um analisador sintático que processa uma entrada de dados. Geralmente um texto reconhecível por humanos que possuem certas regras específicas, e monta uma estrutura de como é sua composição (44,46). Um *HTML Parser* é um tipo de *parser*, entretanto, suas regras são configuradas para manipulação de arquivos do tipo *HTML* (43,44). O *Jsoup* é uma *API Java* que implementa classes e métodos em linguagem de programação *Java* para extrair e manipular dados de arquivos *HTML*, por tanto é considerado um *HTML Parser* para a linguagem *Java* (46). O *Jsoup* traz as seguintes funções de manipulação de arquivos *HTML*:

- Leitura e análise de arquivos *HTML* de um URL, arquivo ou cadeias de caracteres;
- Localizar e extrair dados, usando seletores de passagem *DOM* ou *CSS*;
- Manipular os elementos *HTML*, atributos e texto;
- Limpar o conteúdo enviado pelo usuário contra uma lista segura de *white lists*, para evitar ataques *XSS*;
- Organização de arquivos de saída do tipo *HTML*;

1.5.4 Jsbnl

O *SBML* (*Systems Biology Markup Language*) é a linguagem de marcação padrão, baseada em *XML* (*eXtensible Markup Language*), para armazenamento e troca de modelos computacionais quantitativos em biologia de sistemas (37). O *SBML* é suportado por mais de 210 pacotes de softwares, esse índice alto de referências é caracterizado pelas regras bem definidas e simplificadas, incluindo a disponibilidade de sua sólida *API libSBML* (47). A *API libSBML*, em até certo momento, tinha sua implementação nativa somente em linguagem C/C++, com interfaces (*JNI - Java Native Interface*) que davam suporte ao uso em outras linguagens de programação. Esse tipo de abordagem trazia um certo nível de dificuldade para os desenvolvedores Java, pois sua compatibilidade não é simples para ser incorporada em qualquer aplicação (37,47,48).

Alguns grupos iniciaram suas próprias implementações nativas em *Java*, afim de ter total suporte aos algoritmos e métodos de processamento de arquivos do tipo *SBML*. Um tempo depois, vários desses grupos se juntaram para desenvolver em conjunto uma biblioteca única, evitando duplicações, para o processamento de arquivos *SBML*. Desse esforço surgiu o *JSBML* (*Java Systems Biology Markup Language*), uma *API* totalmente implementada em *Java* (nativa) e *Open Source* sob os termos *LGPL 2.1* (*Library General Public License*) (47).

1.6 Justificativa

O aumento de casos de infecções hospitalares no Brasil causado por bactérias resistente a antibióticos requer novas pesquisas que sirvam à elaboração de alternativas de terapia (3). Uma das abordagens de pesquisa utilizadas na

biologia para construção de conhecimento é a empírica, executada através de experimentos observáveis e sem inferência da dinamicidade em torno do sistema biológico. Em contrapartida, existem abordagens teóricas que visam construir modelos matemáticos / computacionais que, embora sejam precisos, muitas vezes não representam de forma fidedigna o sistema biológico real (5,6). Quando estas duas abordagens são tratadas com uma inter-relação estreita, os resultados se mostram mais promissores. Este trabalho realizou a primeira reconstrução do modelo metabólico da bactéria (1).

Essa primeira reconstrução irá tornar viável uma série de estudos e aplicações de métodos matemáticos/computacionais a fim de criar versões de modelos computacionais. Tais modelos darão suporte ao avanço da compreensão de numerosos processos biológicos do organismo que está sendo representado. Os modelos de referência possuem uma série de vantagens para as pesquisas experimentais, dentre elas a simulação do desenvolvimento rápido do ciclo de vida e os testes de restrição de expressão de genes (nocautes), entre outras análises que se feitas em laboratório demandariam tempo muito mais elevado (6,7,12). Uma possibilidade concreta de utilização atual das versões de modelos gerados seria a identificação de novos alvos terapêuticos através do uso de sistemas como o FindTargetsWeb (9).

Uma outra abordagem computacional é o modelo computacional integrado de célula inteira (28). Trata-se de abordagem mais complexa, composta por diversos modelos que compõem cada processo da célula. Tais modelos se integram para formar uma única representação computacional da célula inteira. Esse tipo de abordagem está sendo desenvolvida por outros projetos. O modelo de rede metabólica desenvolvido pode vir a ser um dos futuros módulos integrados. Isso irá viabilizar estudos relacionados aos mecanismos de resistência a antibióticos apresentados pela bactéria (12,28).

2 OBJETIVOS

2.1 Objetivo Geral

Executar uma reconstrução da rede metabólica da *Pseudomonas aeruginosa* CCBH4851 que simule a geração de biomassa acima de zero.

2.2 Objetivos Específicos

- Empreender releitura do método de *Thiele e Palsson* e desenvolver uma metodologia simplificada para a reconstrução da rede (7).
- Desenvolver uma ferramenta (sistema *web*) para execução da reconstrução automatizada do modelo de rede metabólica da *P. aeruginosa* CCBH4851.
- Desenvolver uma segunda ferramenta computacional (sistema *web*) para a curadoria da rede obtida através da reconstrução automatizada.
- Aplicar ferramentas computacionais existentes para uma avaliação preliminar de geração de biomassa da rede reconstruída (14,34).

3 MATERIAL E MÉTODOS

Neste capítulo é discutido o método criado pela adaptação da releitura do protocolo *T&P* para a reconstrução da rede metabólica (objetivo principal), assim como as ferramentas computacionais desenvolvidas para possibilitar a aplicação do método. Inicialmente, caracteriza-se de modo geral o método. Na seção 3.1, descreve-se a reconstrução automatizada e o sistema NETGen. A seção 3.2 aborda a curadoria assistida com o sistema CurSystem. A seção 3.3 detalha a avaliação preliminar empreendida através de ferramentas computacionais. Por fim, na seção 3.4, expõe-se onde as informações geradas foram armazenadas, ou seja, a arquitetura da base de conhecimento que integra todas ferramentas do método.

O trabalho teve como ponto de partida o resultado do último sequenciamento e anotação do genoma da bactéria *Pseudomonas aeruginosa* CCBH4851, disponível no portal Pseudomonas (<http://pseudomonas.procc.fiocruz.br/index.php/data/p-aeruginosa-ccbh-4851>). O tipo de abordagem utilizada para a reconstrução adotou o protocolo indicado para o desenvolvimento de modelos metabólicos em escala genômica. O processo completo descrito pelo protocolo é dividido em quatro fases principais: a primeira fase consiste na criação de um modelo draft; a segunda fase, na cura manual; a terceira, na conversão para um formato matemático/computacional; e a última, na avaliação da rede (7). Porém, no contexto de execução deste trabalho, não houve tempo hábil para sua aplicação completa. Por esse motivo, estabeleceu-se uma adaptação do protocolo que objetivou sua simplificação. Isto implicou em uma redução aos seus dois primeiros passos, com uma avaliação preliminar de geração de biomassa que consiste na aplicação de métodos matemáticos e computacionais para verificar se o modelo gera cálculos de biomassa coerentes com os sistemas biológicos, e essencialmente valores acima de zero (28).

A Figura 6 exhibe todos os passos da metodologia adaptada pelo presente trabalho.



Figura 6 - Passos da metodologia adaptada de Reconstrução de Redes Metabólicas feitas a partir do protocolo *T&P* (7)

O protocolo *T&P* sugere uma metodologia genérica (7), indicando detalhadamente as fases a serem percorridas para uma reconstrução de qualidade. Entretanto, a execução de todas elas ultrapassa o escopo desta pesquisa. Assim, todas as fases foram adaptadas para a especificidade do organismo alvo da reconstrução, gerando uma metodologia própria. Entre as adaptações, houve a necessidade de criação das próprias ferramentas computacionais para a execução das etapas um e dois, obtendo-se como produto dois sistemas *web* que futuramente poderão ser utilizados ou adaptados para outros organismos alvos e não somente a bactéria *Pseudomonas aeruginosa* CCBH4851 (1).

A seguir, a uma descrição detalhada das etapas e suas respectivas adaptações, incluindo a ferramentas geradas (sistemas).

3.1 Etapa 1 - Reconstrução inicial (draft)

A reconstrução inicial é crucial e é através da anotação do genoma da *Pseudomonas aeruginosa* CCBH4851 (1) que o processo se instaura. Esta é a principal entrada do sistema de geração de rede, sistema *web* de reconstrução NETGen. A reconstrução é feita por meio de dois algoritmos importantes: o primeiro efetua análise e comparação da anotação do genoma com modelos matemáticos/computacionais gerados em outras reconstruções de redes metabólicas de outros microrganismos, como é o caso da *Pseudomonas aeruginosa* PAO1 (12) e *Mycoplasma genitalium* (28). A recomendação do protocolo *T&P* é que

sempre se inicie por esse passo (7). Quando há um modelo de referência curado disponível, que é o caso deste trabalho com o modelo da *Pseudomonas aeruginosa* PAO1 (6,12,36), obtêm-se muitas informações valiosas por causa das semelhanças morfológicas, fisiológicas e da maioria de suas vias metabólicas (7). O segundo algoritmo, também utiliza os dados da anotação genômica, e efetua uma busca automática exaustiva por reações da rede metabólica em bases de dados especializadas em informações de vias metabólicas, como o *KEGG*, *BRENDA* e o *Pseudomonas DB* (25,38,49). Essas bases de dados irão auxiliar no fornecimento de informações acerca das reações da *Pseudomonas aeruginosa* CCBH4851 que não foram encontradas no primeiro algoritmo. Ao encontrar qualquer tipo de informação associada, os algoritmos a armazenam junto às informações colhidas no primeiro algoritmo, constituindo uma Base de Conhecimento. Esta base é um dos produtos desenvolvidos para servir de repositório da *P. aeruginosa* CCBH4851.

3.1.1 NETGen – Sistema web de Geração de Redes Metabólicas

O NETGen surgiu por conta de uma necessidade específica para geração de rede metabólica da *Pseudomonas aeruginosa* CCBH4851. Existem alguns sistemas de reconstrução de redes disponíveis, como o BiGG, o BioModelsDB, o ModelSEED e outros ainda em desenvolvimento. Porém, cada qual com suas características (50), estão ainda distantes da entrega de um suporte abrangente de compatibilidade para modelagens de diferentes organismos. Este fato motivou o desenvolvimento do NETGen.

A arquitetura do sistema NETGen foi pensada e estruturada para possibilitar a compatibilidade entre diferentes APIs disponíveis de interfaces que processam dados biológicos. A linguagem utilizada como *default* (não é a única) foi Java com a extensão do padrão *Java EE*, a qual disponibiliza recursos para modularização e desenvolvimento de software *web* corporativo (43,51). Como mostrado na Figura 7, o NETGen tem passos bem definidos que vão desde a leitura de arquivos com os formatos *GBK* e *SBML* como entrada, até a criação, análise, exportação e envio do novo modelo para o usuário responsável pela reconstrução (37,52). Permeando todo o processo de execução, pode-se citar a referência e a inclusão de APIs de pesquisa bem sólidas, como por exemplo a *BioJava* (processamento de informações em arquivos de dados biológicos) e a *JSBML* (API para processamento de arquivos *SBML* escrita em *Java*) (47). Outra inclusão utilizada para o processo é integração com a linguagem *Python*, para execução do pacote *COBRApy* e suporte da análise

prévia do modelo gerado e enviado para usuário. Os dois algoritmos desenvolvidos são descritos nos próximos parágrafos (14).

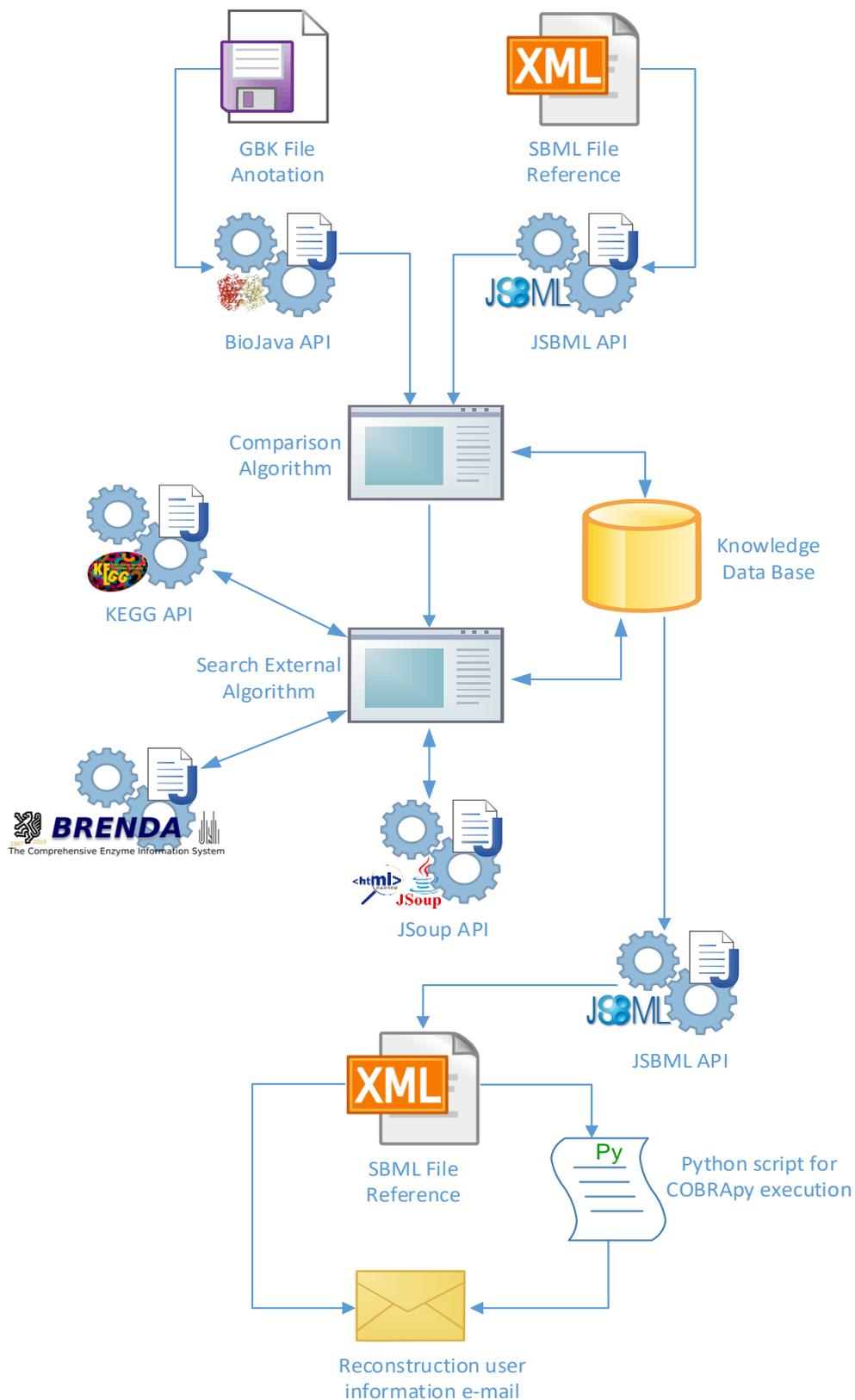


Figura 7 – Arquitetura conceitual do Sistema WEB NETGen

O primeiro algoritmo (Algoritmo 1) do sistema faz uma comparação entre os dois organismos (informações digitais biológicas): o organismo de referência reconstruído e modelado em formato *SBML* e o organismo alvo da reconstrução com anotação do genoma em arquivo formato *GBK* (40,52). Para promover tal comparação é necessária uma tabela de equivalência de identificadores, que está disponível em diferentes formatos (TXT e CSV) e pode ser acessada pelo site do *Pseudomonas Genome DB*. Esta tabela contém informações chamadas de *cross-references* (de/para) de identificadores, ou seja, IDs de genes (e consequentemente suas proteínas) registradas em diferentes bases de dados (Figura 8) (38).

Gene Feature Overview

Strain	<i>Pseudomonas aeruginosa PAO1 (Reference)</i>
Locus Tag	PA0704
Name	
Replicon	chromosome
Genomic location	776787 - 778181 (- strand)
Transposon Mutants	2 transposon mutants in PAO1
Transposon Mutants in orthologs	3 transposon mutants in orthologs

Cross-References

RefSeq	NP_249395.1
GI	15595901
Affymetrix	PA0704_at
Entrez	880754
GenBank	AAG04093.1
INSDC	AAG04093.1
NCBI Locus Tag	PA0704
protein_id(GenBank)	gb AAG04093.1 AE004506_3 gnl PseudoCAP PA0704
TIGR	NTL03PA00705
UniParc	UPI00000C510F
UniProtKB Acc	Q9I5M2
UniProtKB ID	Q9I5M2_PSEAE
UniRef100	UniRef100_Q9I5M2
UniRef50	UniRef50_F9Y5M0
UniRef90	UniRef90_V8EFN8

Figura 8 - Exemplo de Referência Cruzada de IDs de Genes (38).

O algoritmo interpreta cada *RefSeqGeneID* (identificador oficial de proteínas do NCBI) disponível no arquivo *GBK* da anotação do organismo alvo e guarda sua equivalência em *LocusTag* ID da tabela (38,45,52). Esse último é o identificador padrão utilizado em arquivos do tipo *SBML* para modelos de *P. aeruginosa* e por isso a conversão é necessária (6). Com *LocusTagID* executa-se uma varredura no arquivo do modelo de referência (arquivo *SBML*) e verifica-se a existência de determinado gene, com suas respectivas reações e metabolitos. Caso exista a referência, o algoritmo faz a importação para o novo modelo. Ao fim da execução desse primeiro algoritmo há um novo modelo candidato com todas possíveis reações encontradas no cruzamento entre os organismos (dados digitais) (38,52).

A Figura 9 demonstra um exemplo de referência cruzada de IDs:

Strain	Pseudomonas aeruginosa PAO1 (Reference)
Locus Tag	PA0704
RefSeq	NP_249395.1

Figura 9 - Exemplo de Referência cruzada (cross-references)

Segue abaixo o Algoritmo 1 e sua implementação em pseudocódigo, que seleciona cada reação que será incluída na primeira formação do novo modelo em *SBML* (37):

Algoritmo 1 - Comparação entres os dois Organismos (dados digitais)

Entrada: Arquivo GBK (anotação do genoma do organismo alvo), Arquivo SBML (modelo do organismo de referência)

Saída: Informações de Reações armazenadas no Banco de Dados

```
1: procedimento reconstrucaoPrimeiraParte()
2:   NovaReconstrucao : Objeto–Representa uma Execução de Reconstrução
3:   ler arquivoGBK
4:   ler arquivoSBML
5:   ler tabelaCrossRef
6:   para cada sequencia em arquivoGBK faça
7:     se existe sequencia em tabelaCrossRef faça
8:       locusTagID <- pegarLocusTagID(sequencia, tabelaCrossRef)
9:       se existe locusTagID em arquivoSBML faça
10:        adiciona infoReações em NovaReconstrucao
11:      fim-se
12:    fim-se
13:  fim-para
14:  armazena NovaReconstrucao em BD
15: fim-procedimento
```

O segundo algoritmo desenvolvido (Algoritmo 2) inclui as *APIs* de busca do *KEGG* para a tentativa de extrair informações de reações que não foram encontradas no modelo de referência na etapa anterior (25). A base de dados possui *APIs* com implementação em *Java*, o que facilita as chamadas remotas para as buscas de informações por outros programas desenvolvidos também em *Java* (48). Essas dão suporte para busca de informações através de variados campos, como por exemplo, sequências de aminoácidos e EC numbers, que são informações geralmente encontradas na anotação genômica em formato GBK. Ainda há uma segunda opção implementada pelo segundo algoritmo, a mineração de conteúdo de páginas *WEB* (38,52). Para a execução desta tarefa, foi utilizada a *API Jsoup HTML*

Parser, uma interface também desenvolvida em *Java* e que executa processos que facilitam a extração e navegação pelas marcações de um documento *HTML* (46).

A função de incrementar o modelo (*SBML*) com reações que foram encontradas pela busca externa em bases dados é descrita abaixo pelo Algoritmo 2. Note-se que o algoritmo está dividido em dois módulos, um com o procedimento principal e o segundo com uma função auxiliar.

Algoritmo 2 parte 1 – Busca externa de Informações de Reações em Base de Dados

Entrada: ListaDeGeneNaoEncontrado - Lista de Genes NÃO Encontrados
NovaReconstrucao – Objeto criado para representar uma execução

Saída: Informações Adicionais de Reações armazenadas no Banco de Dados

```
1: procedimento reconstrucaoSegundaParte()
2:   para cada geneNaoEncontrado em ListaDeGeneNaoEncontrado faça
3:     se geneNaoEncontrado contem EcNumber faça
4:       funçãoArmazenarNoBD(geneNaoEncontrado.EcNumber)
5:     senão se
6:       EcNumber <- buscarLocusTagID(geneNaoEncontrado.LocusTagID)
7:       funçãoArmazenarNoBD(EcNumber)
8:     senão se
9:       EcNumber <- buscarSeqAA(geneNaoEncontrado.SeqAA)
11:      funçãoArmazenarNoBD(EcNumber)
12:   fim-se
13: fim-para
14:fim-procedimento
```

Algoritmo 2 parte 2 – Busca externa de Informações de Reações em Base de Dados

```
15: função ArmazenarNoBD(EcNumber : Caracter)
16:         resultadoKEGG <- buscarKEGG(geneNaoEncontrado.EcNumber)
17:         se resultadoKEGG != Nulo faça
18:             infoReações <- extraiInfoReações(resultadoKEGG)
19:             adiciona infoReações em NovaReconstrucao
20:         fim-se
21: fim-função
```

Toda informação processada do modelo pelos algoritmos é armazenada na base de conhecimento. A base de conhecimento também foi desenvolvida como produto desse trabalho e sua principal função é guardar as informações das versões geradas de modelos da *P. aeruginosa* CCBH4851. O funcionamento desta base será descrito mais adiante, na seção 3.4 deste capítulo.

Ao final da execução dos dois algoritmos, o sistema executa um método (rotina programática) auxiliar para efetuar a leitura dos dados armazenados da base de conhecimento e, a partir das informações lidas, cria o arquivo no formato *SBML* da versão do modelo correspondente (37,44). O fluxo de execução segue e o arquivo *SBML* é utilizado como entrada em dois outros métodos do sistema.

O primeiro método implementa a integração com a linguagem *Python* e executa a análise de balanço de fluxo (*FBA*) através do uso do pacote *COBRAPy* (14,34). Ainda não existe nenhuma *API* ou *framework COBRA* para a linguagem *Java*. Contudo, a integração do *Python* e o *COBRAPy* para executar internamente a análise no sistema NETGen ficou facilitada pelo suporte da plataforma *Java EE* para interconexão com outras linguagens de programação (43). O código do método ficou conciso como pode ser observado na Figura 10.

```

private String executarAnaliseFBA(String _nomeArquivoSBML) {
    try {
        String log = "";

        String scriptFBA = "python testeCCBH4851.py " + _nomeArquivoSBML;

        File diretorio = new File(pathAreTrabalho);

        Runtime runtime = Runtime.getRuntime();

        Process processo;

        String[] nargs = { "/bin/bash", "-c", scriptFBA }; // Unix/Linux/MacOSX

        System.out.println("INICIO ANALISE FBA");

        processo = runtime.exec(nargs, null, diretorio);

        BufferedReader bReader = new BufferedReader(new InputStreamReader(processo.getInputStream()));

        String linha;
        while ((linha = bReader.readLine()) != null) {
            log += "\n" + linha;
            System.out.println(linha);
        }

        System.out.println("FIM SCRIPT");
        //FIM DO SCRIPT

        return log;
    } catch (Exception e) { e.printStackTrace(); return ""; }
}

```

Figura 10 - Método que interconecta o *Python* e executa o *FBA* do *COBRapy* internamente no *NETGen*

Na Figura 11, analisamos o código que é invocado pelo método integrador que está contido na Figura 10. O código é escrito em *Python* para análise de geração de Biomassa do modelo recém gerado (14).

```
testeCCBH4851.py x
1  #!/usr/bin/python
2
3  import cobra.test
4  import copy
5  import os
6  from os.path import join
7  from math import log
8  import copy
9  import pandas
10 import sys
11 pandas.options.display.max_rows = 100
12
13 data_dir = os.getcwd()
14
15 sbml_file = str(sys.argv[1])
16 nomeRede = str(sys.argv[2])
17
18 ccbh = cobra.io.read_sbml_model(sbml_file)
19
20 print ccbh.reactions.get_by_id("PA01_Biomass")
21
22 print ('----- CCBH4815 -----')
23
24 taxaC = ccbh.optimize().f
25 print ('ID: %s') % nomeRede
26 print ('biomass generated: %f mmol/g DCW/h') % (taxaC)
27 tx_em_horas = log(2.0)/taxaC
28 tx_em_minutos = tx_em_horas * 60
29 print ('biomass produced in hours: %f') % (tx_em_horas)
30 print ('biomass produced in minutes: %f') % (tx_em_minutos)
31 print ('-----Reconstruction numbers-----')
32 print ('reactions: %i') % (len(ccbh.reactions))
33 print ('metabolites: %i') % len(ccbh.metabolites)
34 print ('genes: %i') % len(ccbh.genes)
```

Figura 11 - Script em Python para a simulação do cálculo de Biomassa do modelo recém gerado

O método faz a transição do fluxo de execução para o *script* de análise em *Python*, este executa os métodos do pacote *COBRAPy* e retorna o fluxo de execução para o *Java* (14,43). O retorno da execução é capturado em uma cadeia de caracteres, entretanto, é preciso percorrê-la a fim de estruturar a resposta para formar o *log* (histórico). Por último, o método retorna o *log* para o método principal da reconstrução e insere o resultado junto ao e-mail que será enviado para o usuário com as informações da reconstrução.

O segundo método implementa a utilização da *API* de *e-mail* da *Oracle* (53). Através desse método, o sistema envia um e-mail ao responsável pela reconstrução informando o resultado e encaminhando o arquivo gerado como anexo do e-mail.

Todos esses algoritmos e métodos descritos são o que se designa núcleo do sistema ou “*core*”. Trata-se de módulos de programas (métodos) que compõem o

núcleo sistemático (44). Para que todos esses recursos possam estar à disposição de qualquer usuário pela *web*, faz-se necessário encapsulá-los em alguma casca, uma interface para facilitar sua utilização (43). Isto foi feito utilizando as tecnologias *Java* no NETGen e CurSystem, já citadas anteriormente.

A estruturação e o desenvolvimento do sistema em relação à tecnologia e arquitetura escolhidas, junto com os seus *Screenshots* de interface (unidades de representações) e fluxo de execução (48) serão discutidos a seguir.

A tecnologia *Java EE* foi empregada por causa de sua estrutura tecnológica que apoia bastante o desenvolvimento de sistemas *web* (42,43,44). Todas as suas especificações de plataforma são projetadas para se encaixar em projetos com arquitetura de projeto em camadas *MVC* (42,44). A implementação de cada camada do NETGen utilizou um dentre os *frameworks* disponíveis. A camada Visão (*view*) adotou o *JSF* (v2.2.8.21) com *framework Primefaces* (v6.0). Para a camada Controle (*control*), foram escolhidos o *Java Core* (v1.8.131) e *Python* (v2.7). A camada Modelo (*model*) utilizou *JPA* e *JDBC* (42,43,44).

Na Figura 12 podem ser observados dois elementos citados acima, as referências das *APIs* e as camadas implementadas seguindo o padrão *MVC*:

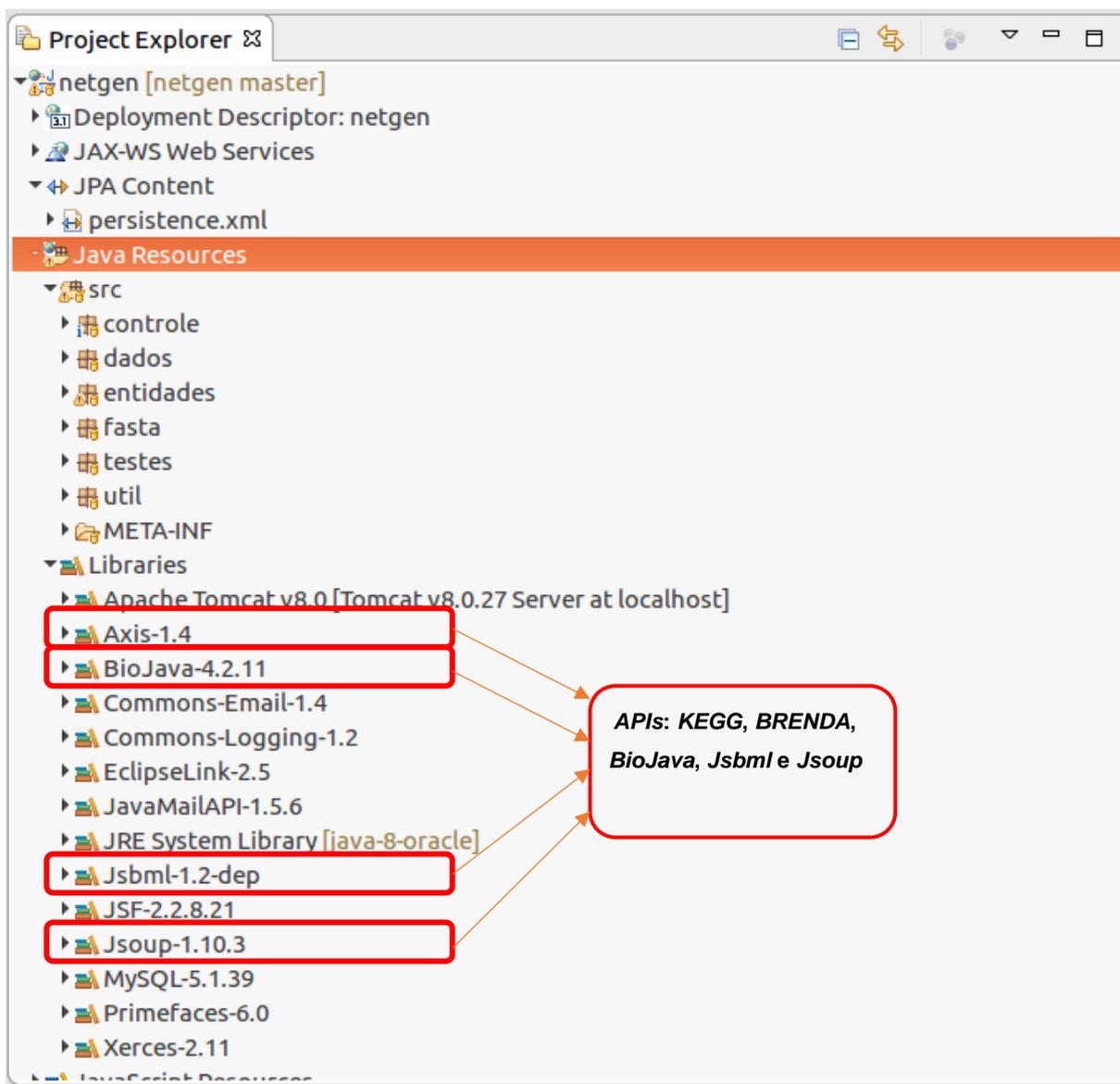


Figura 12 - Screenshot do Workspace do NETGen mostrando as APIs referenciadas

Para execução do NETGen foi implantado o servidor de aplicação *Apache Tomcat* (v8.0.27), o qual fornece todo suporte que um sistema *web* desenvolvido com o padrão *Java EE* demanda, não necessitando outro tipo de serviço adicional (42). O sistema atualmente encontra-se disponível no servidor localizado na Sede do Programa de Computação Científica do Instituto Oswaldo Cruz e pode ser acessado através do link: <http://pseudomonas.procc.fiocruz.br:8185/netgen/>.

A seguir, as principais Unidades de Representação¹ da Interface do NETGen (54) são apresentadas. A Figura 13 exibe a tela de entrada do sistema.

¹ Em uma interface gráfica, uma unidade de apresentação é uma tela ou janela. Já em interfaces *web*, é uma página (54).

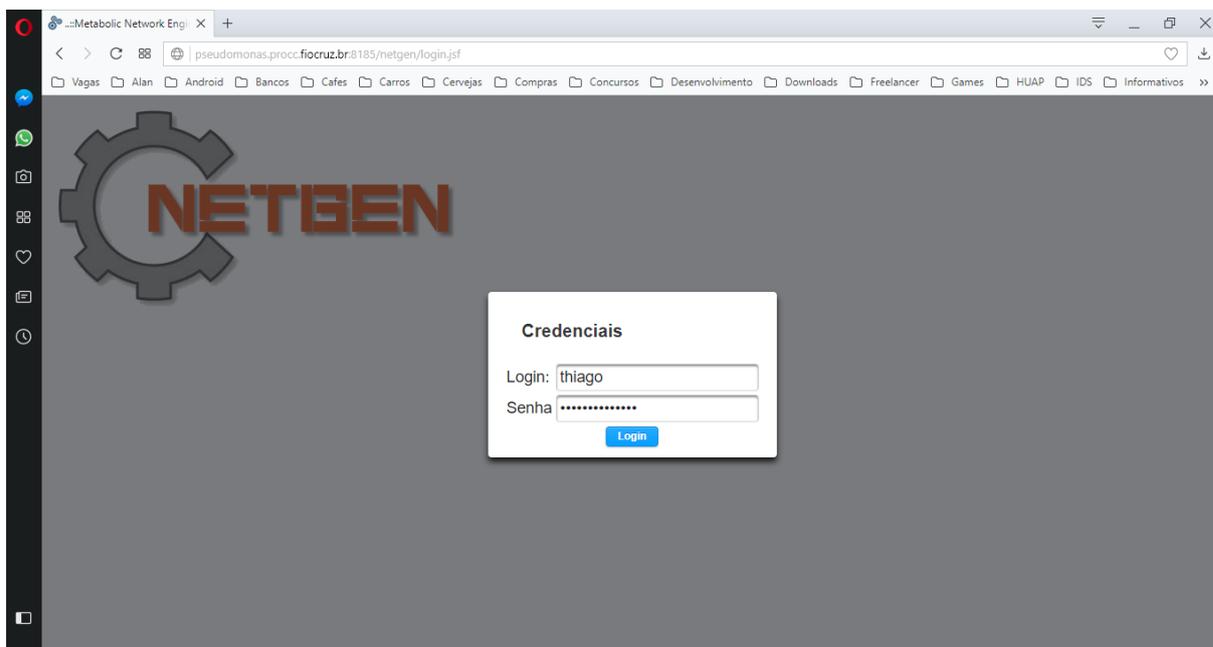


Figura 13 - Unidade de apresentação para *login* do NETGen

O sistema monitora as execuções de reconstrução por usuário, ou seja, toda reconstrução da rede metabólica criada na base de conhecimento é convertida em uma versão atrelada ao usuário da sessão corrente (*login* ativo).

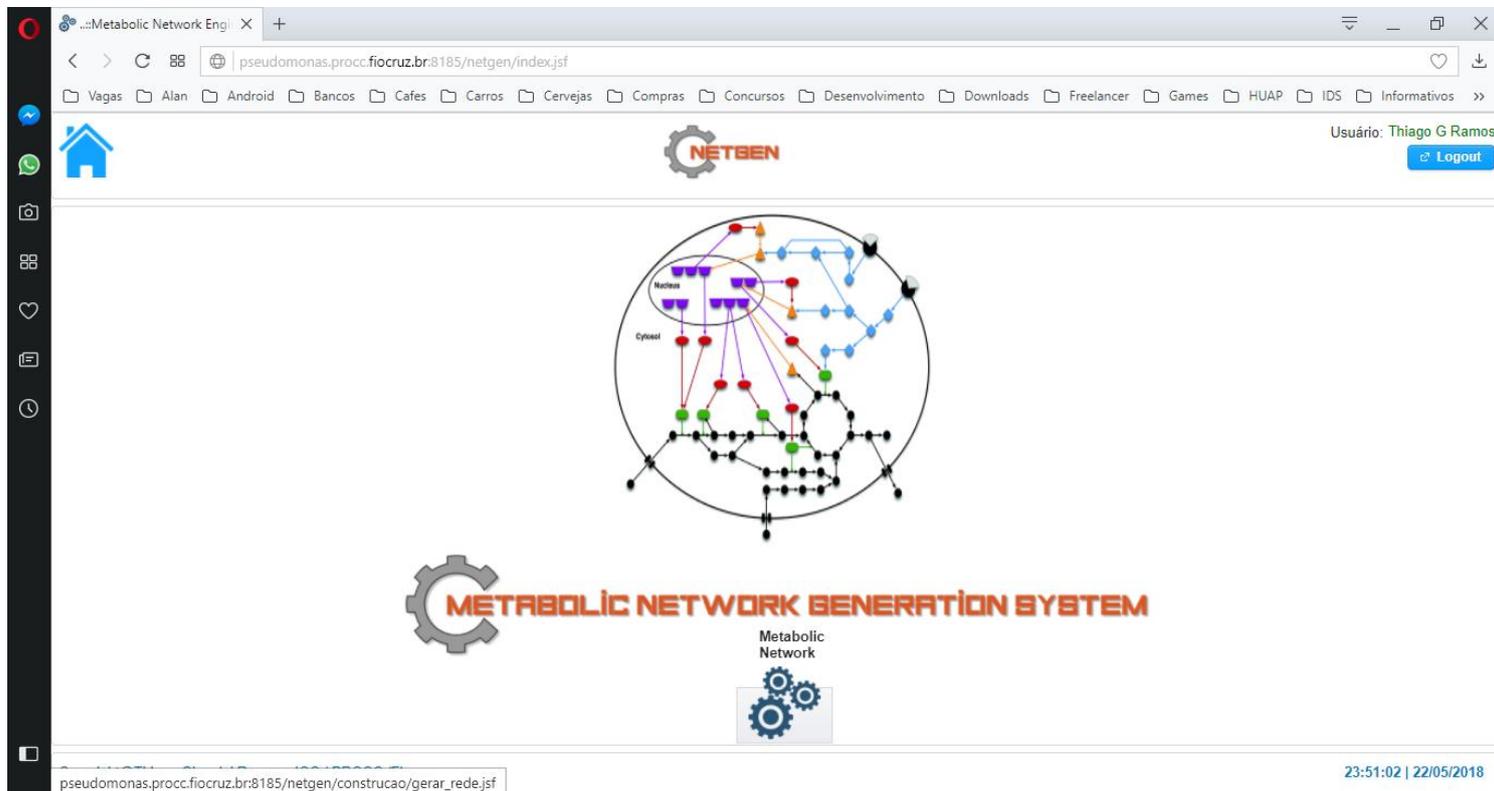


Figura 14 – Unidade de apresentação principal do NETGen

Na página principal (Figura 14), clicando no botão “*Metabolic Network*”, o usuário pode solicitar o redirecionamento para o ambiente de execução de novas reconstruções (Figura 15).

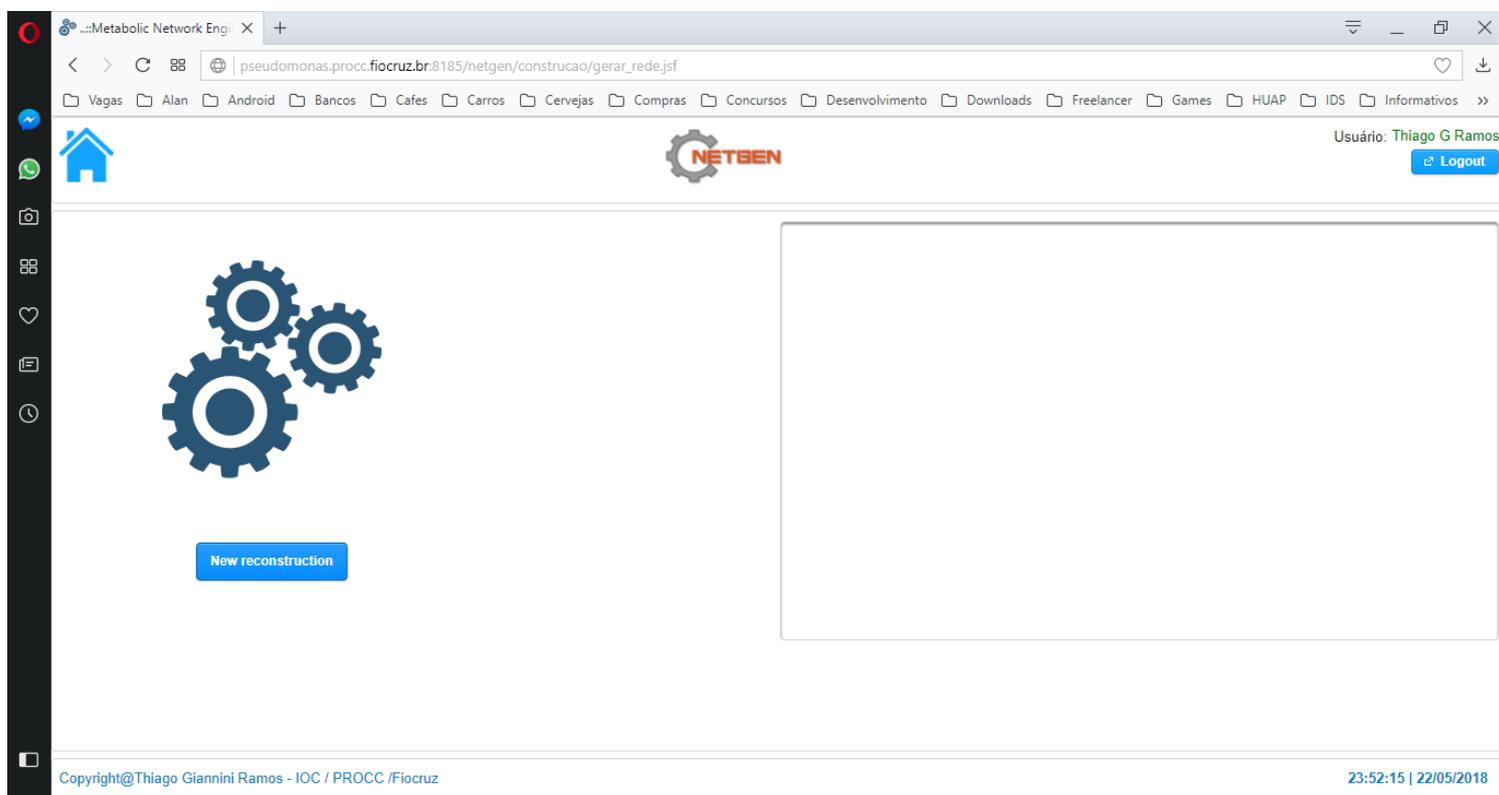


Figura 15 - Ambiente de execução de Reconstruções Metabólicas

Para executar uma nova reconstrução, o usuário precisa acionar o botão “*New reconstruction*”. O sistema mostra uma janela de diálogo (*popup*) com um assistente (*reconstruction wizard*) que auxilia na configuração e definição das informações do modelo de rede metabólica. Na janela (Figura 16) “*New Background Reconstruction*” demonstra a preparação para a execução de uma reconstrução. A preparação é feita em três etapas. Essas etapas estão descritas a seguir.

1 – Aba de informações da reconstrução: o usuário define o título do modelo da rede, número de versionamento (versão na base de conhecimento), nome do arquivo *SBML* de saída, se vai existir a inferência de genes (escolha manual), qual o arquivo de anotação no formato *GBK* será feito *upload* para o sistema e o qual o arquivo do modelo (formato *SBML*) de referência será utilizado para *upload* também no sistema.

2 – Aba de inferência de genes: em seguida, se foi marcada a opção de utilização de inferência, a janela irá apresentar uma lista de reações (*GPR*) carregadas a partir do *SBML* do modelo de referência. O usuário poderá escolher quantas reações estiverem disponíveis na lista e essas serão forçadas como entrada no novo modelo da reconstrução.

3 – Aba de confirmação e execução: nesta última etapa obrigatória, o usuário poderá conferir todas as configurações definidas nas etapas anteriores e, por fim, clicando no botão “*Start*”, a execução da reconstrução será iniciada em *background* (processo executado em modo silencioso) no servidor onde está implantada o sistema *web*.

New Background Execution

Reconstruction Info | Inference | Confirmation

Reconstruction information

Auto ID: 1

Enter your name: Thiago G Ramos

Enter your e-mail: gianninimail@gmail.com

Title of Reconstruction: Reconstrução completa

Version of Reconstruction: 1

Name of output SBML: CCBH4851_v1.xml

Table From(NBCI)/To(PA) for ID's: tabelaNCBI.txt

Inferring genes? YES

Select Upload GBK File:

Select Upload SBML Base File:

→ Next

Start Cancel

Figura 16 - Janela de Diálogo com o Assistente de Reconstrução com as três etapas de preparação para a execução de uma nova reconstrução

New Background Execution

Reconstruction Info
Inference
Confirmation

Reconstruction information

Auto ID: 1

Enter your name: Thiago G Ramos

Enter your e-mail: gianninimail@gmail.com

Title of Reconstruction:

Version of Reconstruction:

Name of output SBML:

Table From(NBCI)/To(PA) for ID's:

Inferring genes? YES

+ Choose
↗ Upload
⊗ Cancel

Select Upload GBK File:

CCBH4851.gb	13.8 MB	<div style="width: 100%; height: 15px; background-color: #007bff; position: relative;"> <div style="width: 100%; height: 100%; background-color: black; position: absolute; top: 0; left: 0;"></div> </div>	✕
-------------	---------	---	---

+ Choose
↗ Upload
⊗ Cancel

Select Upload SBML Base File:

iPAE1146.xml	2.6 MB	<div style="width: 100%; height: 15px; background-color: #007bff; position: relative;"> <div style="width: 100%; height: 100%; background-color: black; position: absolute; top: 0; left: 0;"></div> </div>	✕
--------------	--------	---	---

→ Next

Start
Cancel

Figura 17 - Aba de inserção de informações e upload dos arquivos necessários para a reconstrução

O *layout* da aba “*Reconstruction Info*” exibe ferramentas para a escolha dos arquivos que serão utilizados na reconstrução, os quais são transferidos (via *API JSF*) para uma pasta reservada pelo sistema no servidor (Figura 17). Na existência de arquivos com os mesmos nomes, o sistema os substitui imediatamente pelos novos arquivos escolhidos. Uma restrição imposta na interface da aba refere-se à escolha da extensão do arquivo, pois somente *GBK* para anotações (organismo alvo) e *SBML* (organismo de referência) podem ser submetidos ao servidor.

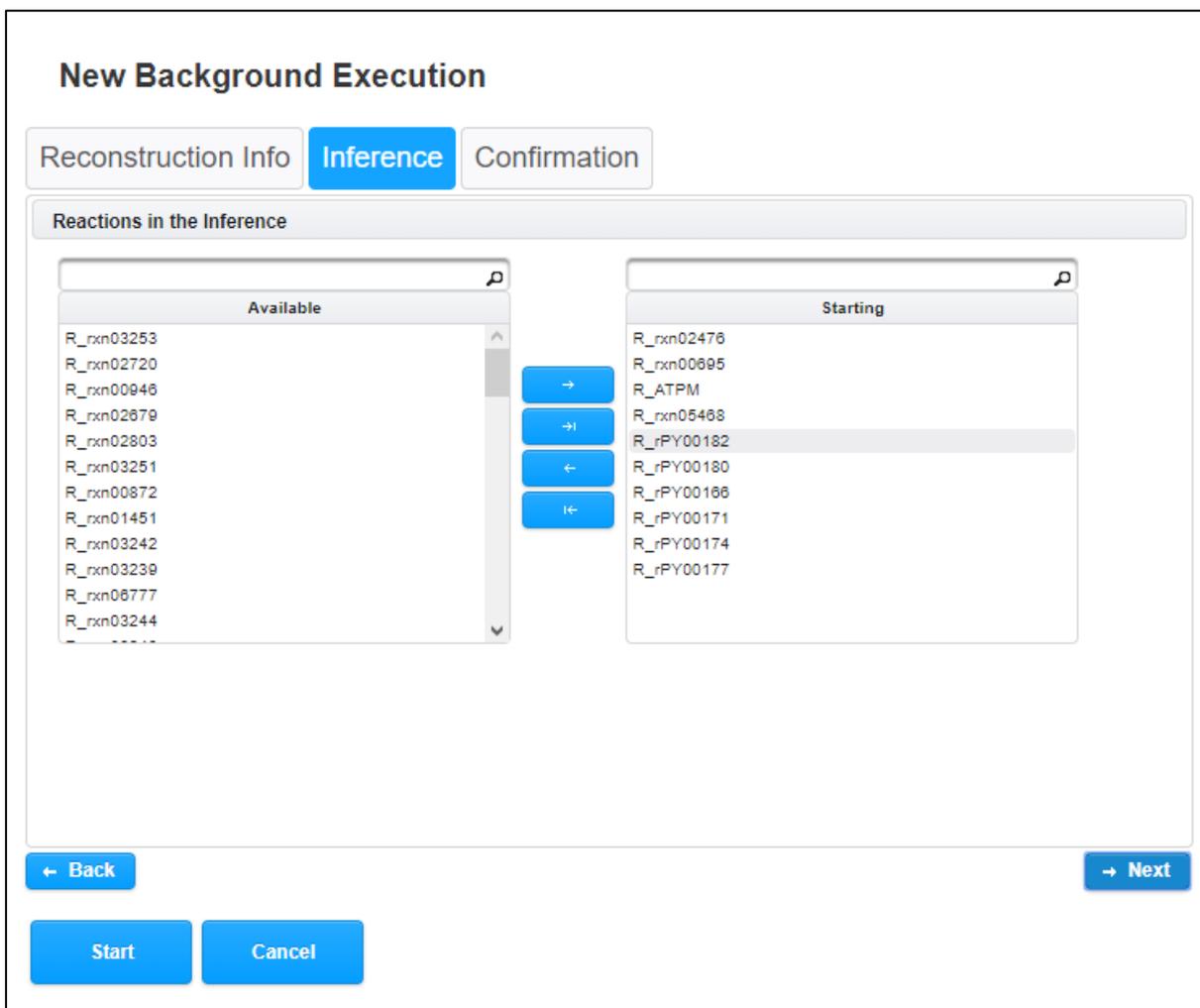


Figura 18 - Aba para inferência manual de reações na reconstrução do novo modelo

A aba de inferências é mostrada na Figura 18 e possibilita a modificação manual com base no modelo de referência. O pesquisador pode selecionar algumas reações para serem forçadas a compor a reconstrução da rede metabólica alvo. Com isso, pode-se testar diferentes combinações e gerar diversas versões do modelo do organismo alvo.

A aba de confirmação exibe um resumo com todas as informações inseridas ou escolhidas pelo pesquisador. A Figura 19 mostra um exemplo de reconstrução de rede metabólicas prestes a ser executada:

New Background Execution

Reconstruction Info Inference **Confirmation**

Confirmation

Enter your e-mail:	gianninimail@gmail.com
Title of Reconstruction:	Reconstrução completa
Version of Reconstruction:	1
Name of output SBML	CCBH4851_v1.xml
Table From(NBCI)To(PA) for ID's	tabelaNCBI.txt
Name of GBK File:	CCBH4851.gb
Name of SBML Base File:	iPAE1146.xml
Inferring genes?	YES

List of Reactions Inferred

1. R_rxn02476
2. R_rxn00695
3. R_ATPM
4. R_rxn05468
5. R_rPY00182
6. R_rPY00180
7. R_rPY00166
8. R_rPY00171
9. R_rPY00174
10. R_rPY00177

← Back

Start **Cancel**

Figura 19 - Janela de resumo e confirmação de configuração da reconstrução que será executada

Após confirmação de todos os dados referentes à reconstrução, o usuário a inicia clicando no botão “Start”. A Figura 20 demonstra uma execução de reconstrução em andamento.

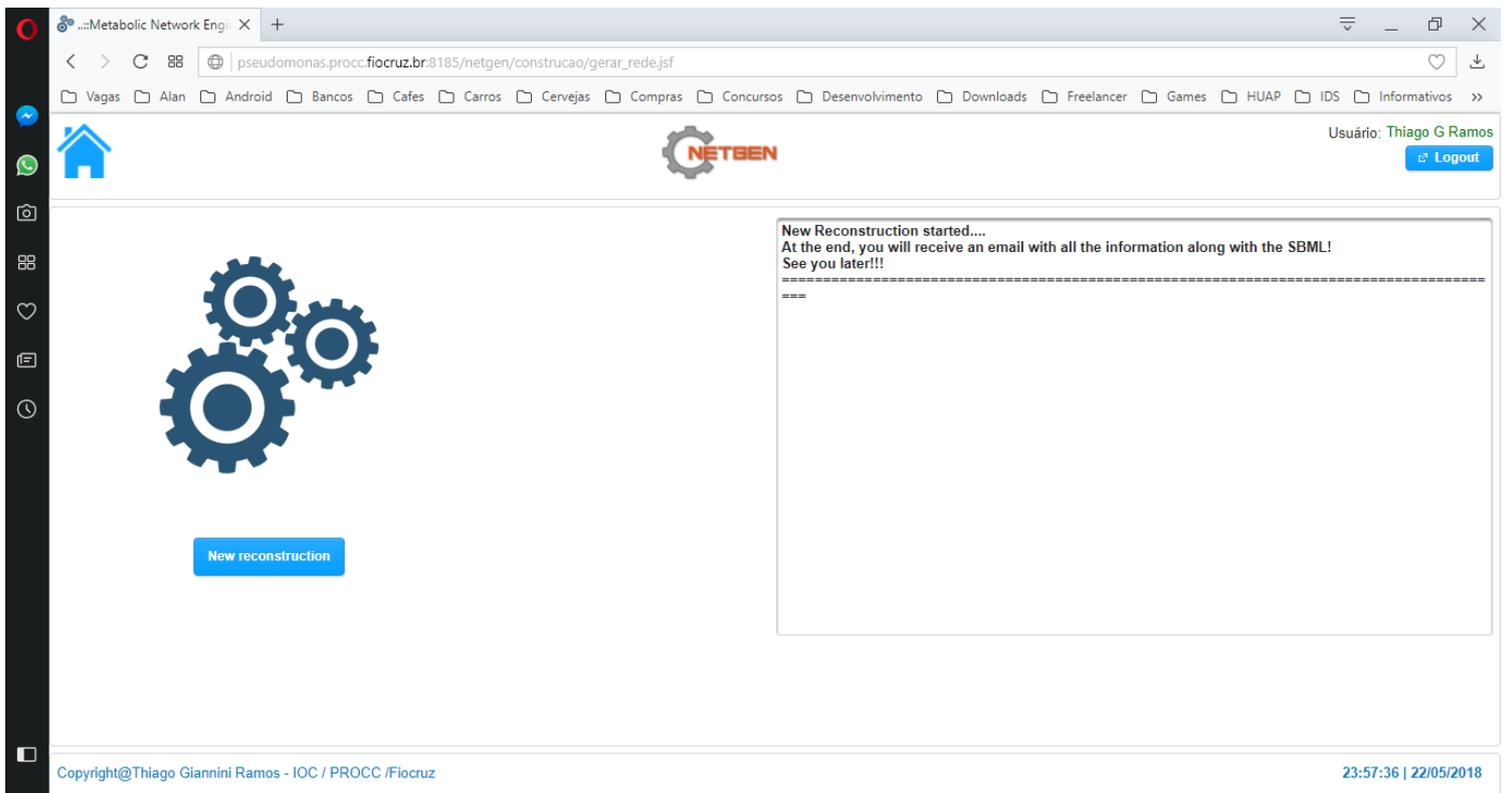


Figura 20 - Unidade de execução em background da reconstrução

Ao iniciar uma execução de reconstrução de rede metabólica, o sistema mostra o *status* e informa ao usuário que em breve será enviado um e-mail contendo todas informações do *pipeline* executado e um anexo com o arquivo do novo modelo gerado em formato *SBML*. Um exemplo de resultado pode ser visto na Figura 21.

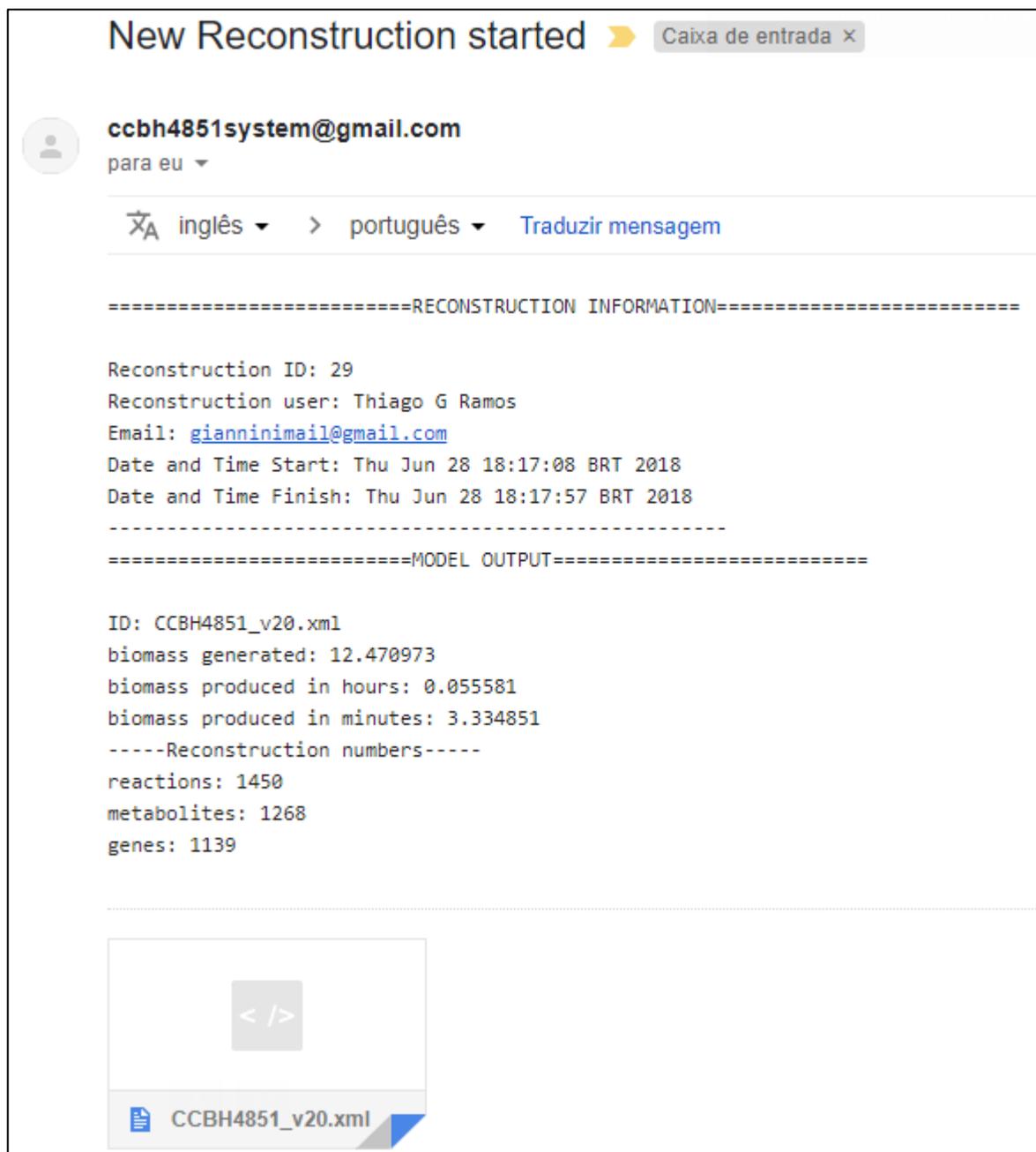


Figura 21 - E-mail de resultados da nova reconstrução do modelo alvo

3.2 Etapa 2 - Refinamento da rede metabólica por especialistas

Na segunda fase, o objetivo foi corrigir erros e adicionar informações que não foram possíveis extrair na primeira etapa de forma automatizada. Essa etapa é muito importante, pois nela se validam os dados extraídos de forma automática na fase anterior e se corrigem eventuais inconsistências, *gaps* da rede (7,12).

Cada reconstrução tem um escopo de estudo e é sempre visando alcançar esse propósito que a curadoria deve se guiar. Porém, na etapa de planejamento do método foi previsto que reunir especialistas para a curadoria seria um grande

desafio. O grupo de pesquisadores e apoiadores são de diferentes instituições do estado do Rio de Janeiro. Além disso, há a dificuldade adicional de compatibilizar as agendas de todos os envolvidos. Por essa razão, fez-se necessário discutir uma forma de interligar e centralizar as informações produzidas de forma independente de local físico e sincronismo de agenda. Concluiu-se, então, que a descentralização e a assincronia só poderia ser tangível através da internet, através de um sistema computacional que funcionasse na internet, um sistema *web*.

A fim de promover o refinamento da rede através da interação remota e assistida entre especialistas, foi proposto um segundo sistema *web*. Trata-se de um Sistema de Desafios (*Challenges*) que traz consigo uma junção de técnicas computacionais, designado *CurSystem*. O *CurSystem* foi desenvolvido para intermediar o trabalho feito pelos especialistas, de maneira que eles possam propor e efetuar possíveis ajustes na rede metabólica da *Pseudomonas aeruginosa* CCBH4851 de forma automatizada.

O objetivo foi desenvolver um software que propusesse desafios para os grupos de especialistas que participarão da curadoria. Esses desafios abarcam questões relacionadas às características que serão discutidas da rede metabólica e cada *gap* analisado tem uma ou mais soluções (respostas) propostas. Para cada resposta, pode haver ou não uma lista de discussão entre os pesquisadores. O responsável pelo desafio define uma solução entre as respostas disponíveis. A resolução aprovada será aplicada na modificação no modelo, gerando uma nova versão da rede metabólica a cada evolução (7). O diferencial ofertado pelo *CurSystem* consiste na maneira como se dá cada evolução do modelo. Ele permite que todas as contribuições dos especialistas sejam feitas de forma assíncrona e distribuída, diminuindo o tempo entre os incrementos no modelo. Toda operação da curadoria é executada *online*, via *internet*. Por isso, para o acompanhamento total da curadoria, qualquer que seja a movimentação no sistema – da discussão à alteração do modelo – é notificada via *e-mail* ao grupo de pesquisadores.

O fluxo da curadoria do sistema realizada pelo grupo de especialistas através da utilização do sistema é exibido abaixo (Figura 22):

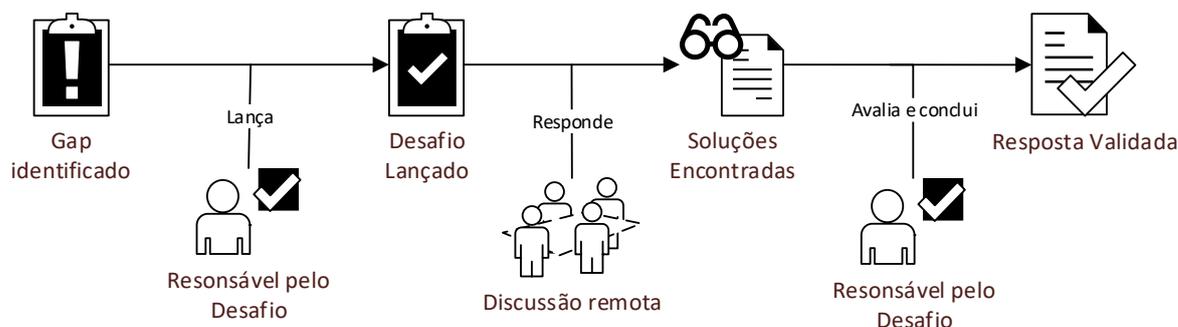


Figura 22 - Fluxo de trabalho de Curadoria Assistida da Rede Metabólica

3.2.1 CurSystem – Sistema *WEB* de Curadoria Assistida de Redes Biológicas

O CurSystem também foi desenvolvido com as tecnologias disponíveis no *Java* com extensão do padrão *EE*. Sua arquitetura não tem o mesmo grau de complexidade que o NETGen, o qual exigiu a implementação de muitas *APIs* integradas. A complexidade do CurSystem está em suas regras de negócio, as quais vinculam os desafios e suas soluções às reações do modelo metabólico depositadas na base de conhecimento. Objetiva-se que cada solução proposta represente uma nova fórmula química ou uma confirmação da existente e que caso a solução seja a escolhida como resposta validada pelo responsável do desafio, ela possa substituir a reação original.

A *Integrated Development Environment* (IDE) utilizada para a implementação foi o Eclipse, uma ferramenta de ambiente integrado que suporta variadas linguagens de programação e inclui outros plugins de integração com servidores e serviços *web* (42). Na Figura 23 expõe-se a organização no *Eclipse* e os pacotes e *Libraries* (bibliotecas de códigos reutilizáveis) do trabalho.

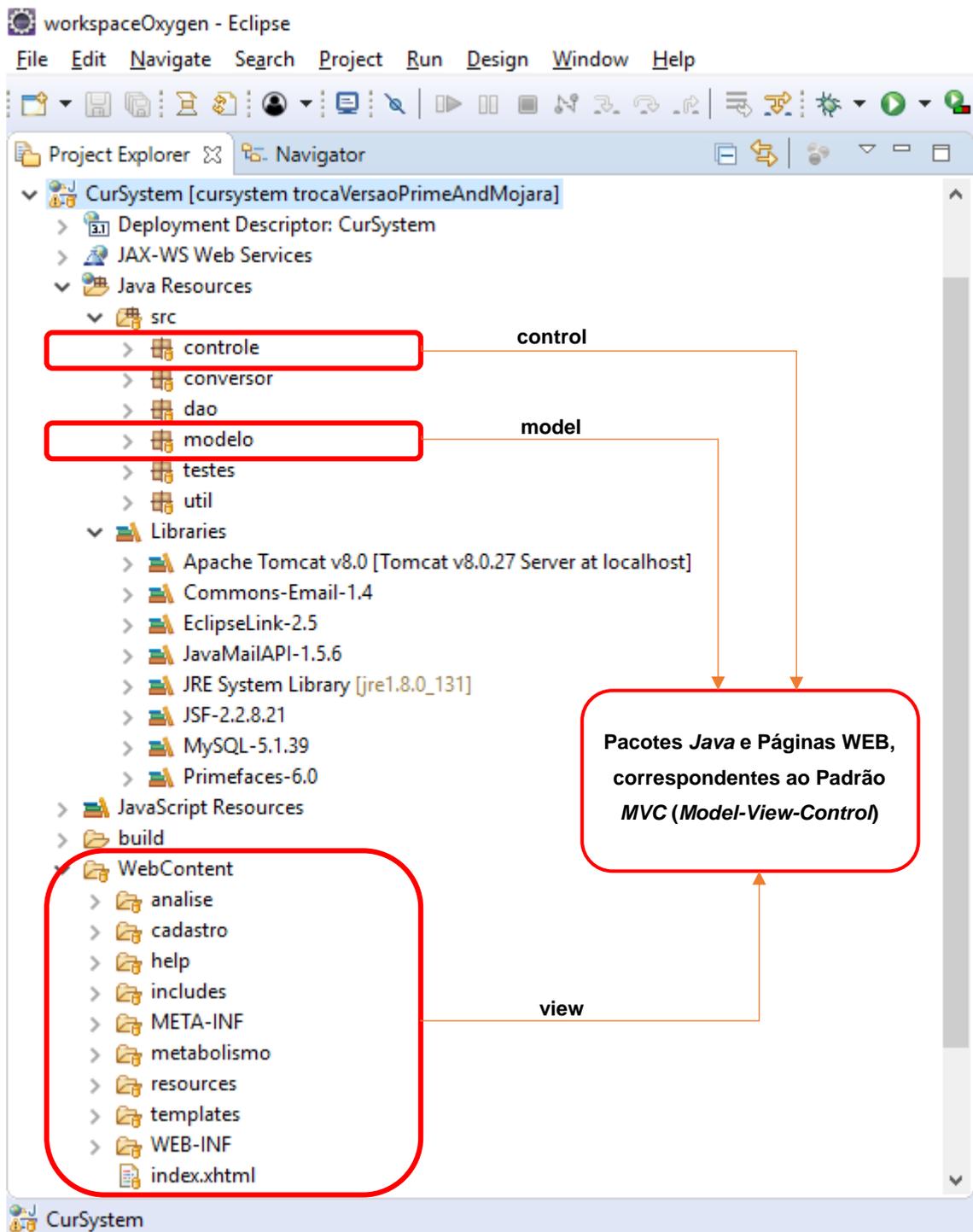


Figura 23 - Espaço de Trabalho com Classes (modelo, controle e visão) e *Libraries* utilizadas no projeto do CurSystem

O CurSystem, assim como o NETGen, também foi dividido em camadas respeitando a arquitetura Modelo-Visão-Controle (MVC). Outro ponto comum em relação ao NETGen é sua preparação para acesso à base de conhecimento, que é projetada no *MySQL*, obrigando os sistemas à utilização de *APIs* de acesso do gerenciador da Oracle (43,55,56). O acesso à base de conhecimento será discutido pormenorizadamente mais adiante.

Abaixo, apresenta-se os principais fluxos do modelo de interação como conversa (MOLIC) e as unidades de apresentação no formato firmado pelas teorias que compõem a multidisciplinaridade em Interação Humano-Computador (54). A Figura 24a representa modelo de interação da Página de Login e a Figura 24b mostra o modelo para cadastro de desafios.

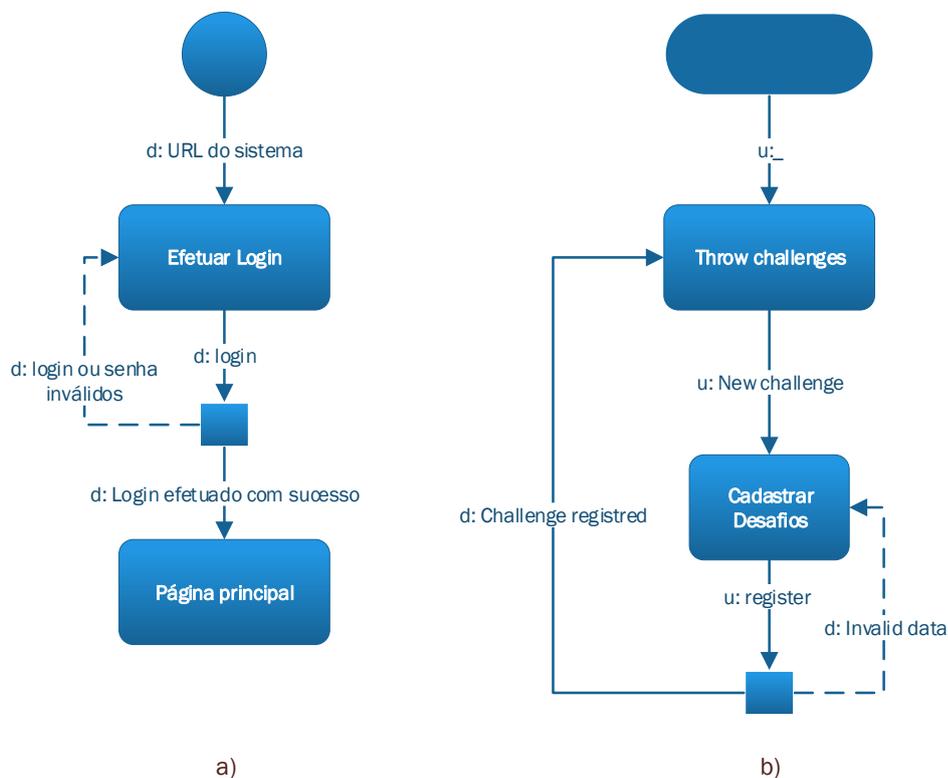


Figura 24 – (a) Modelo de Interação das Páginas de Login e (b) Cadastro de Desafios

O modelo de interação apresentado na Figura 25 mostra como o *status* do ciclo de vida do desafio é alterado. Quando há uma primeira resposta, o desafio passa do *status* “Open” para “Analyzing”. Para isso, o usuário deve acessar a página “Challenge Discussions” e criar uma resposta para o desafio. Nessa página também podem ser feitos comentários sobre outras respostas ou até mesmo as próprias respostas.

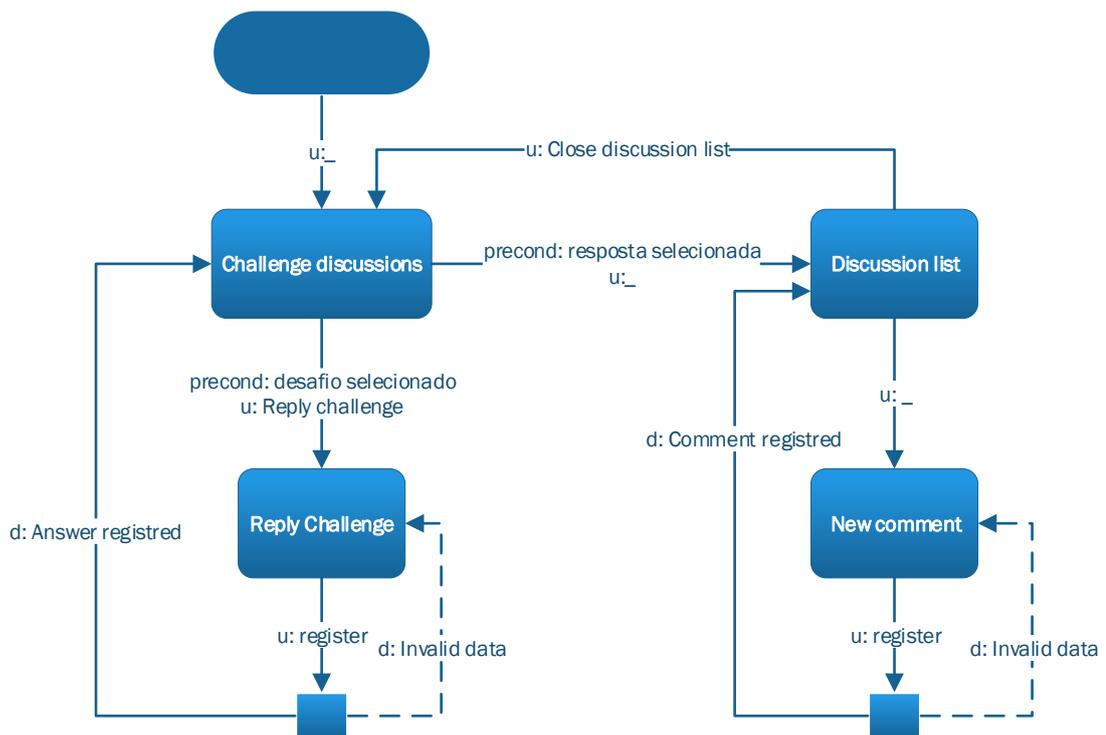


Figura 25 - Modelo de interação da Página *Challenges Discussions*

Na Figura 26 a interação representada pelo modelo mostra como desafio ao ser fechado pelo usuário tem seu *status* alterado para “*CLOSED*”. Porém, somente o responsável pelo desafio pode fechá-lo. O usuário seleciona uma resposta entre outras cadastradas naquele desafio, insere a justificativa e valida a resposta.

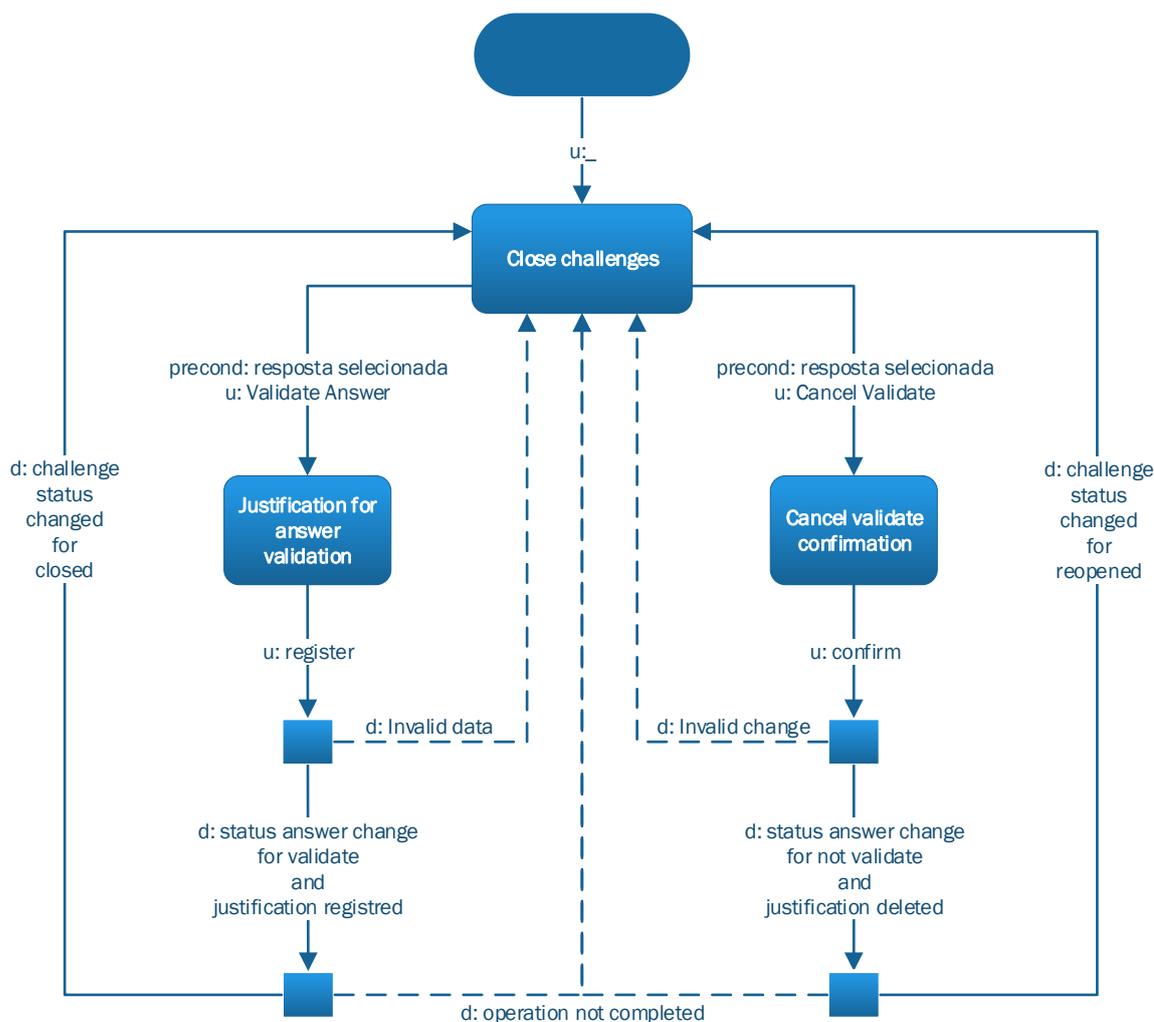


Figura 26 - Modelo de Interação da Página *Close Challenges*

Não foi criado um modelo de interação para todas as interfaces do sistema, pois essa necessidade existiu somente para as unidades mais complexas, onde diferentes ações podem ser executadas em uma única página.

A unidade de login CurSystem é muito similar à do NETGen e pode ser vista na Figura 27. A unidade de apresentação principal centraliza todas os links para as demais e armazena as informações do usuário da sessão corrente (Figura 28).

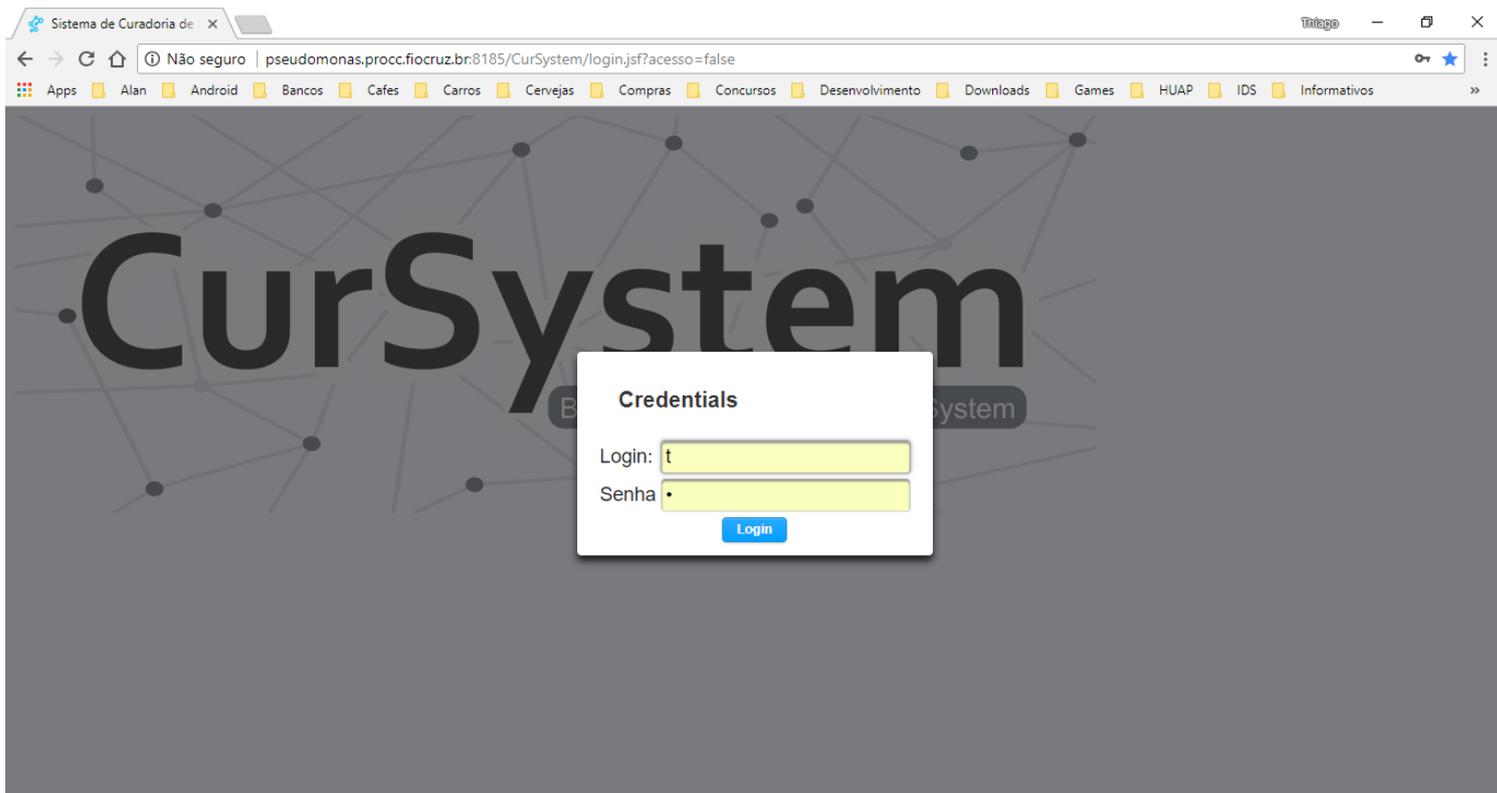


Figura 27 - Unidade de apresentação para *login* do CurSystem

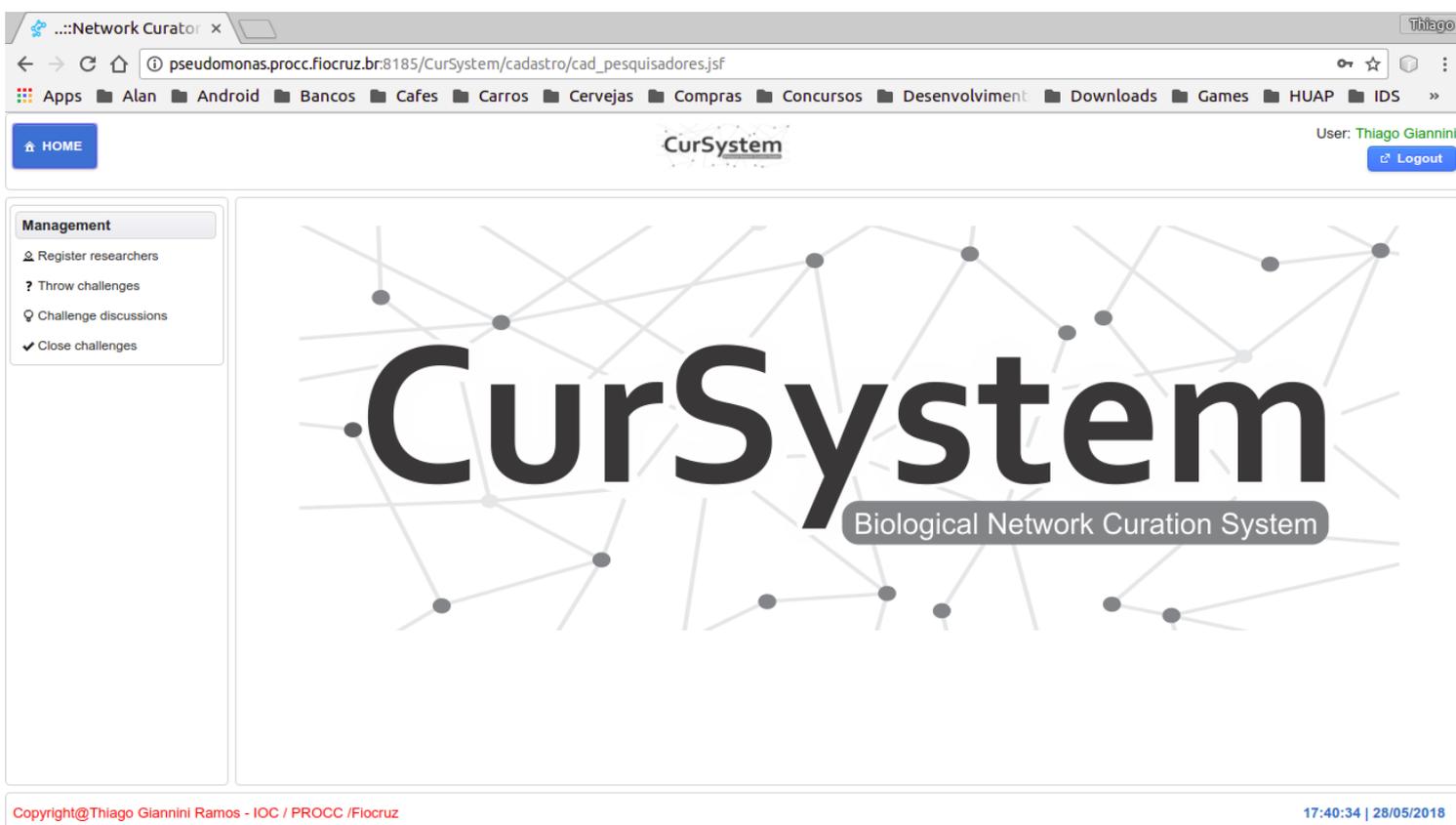


Figura 28 – Unidade de apresentação “Home” (página principal)

O modelo de interação mostrado na Figura 24(a) necessitou ser implementado em duas unidades de apresentação e deu origem à página de desafios e à janela de diálogo de cadastro de desafios. As duas unidades de apresentação podem ser visualizadas nas Figura 29 e Figura 30, respectivamente.

The screenshot shows a web browser window with the URL `pseudomonas.procc.fiocruz.br:8185/CurSystem/cadastro/cad_desafios.jsf`. The page title is "CurSystem" and the user is logged in as "Thiago Giannini". The main content is a table titled "List of challenges".

Id	Title	Status	Project (Network Type)	Owner	Answers	Actions
119	R_ATPM / ATPM - Confirmar de Reação Importada	OPEN	Metabolic Network CCBH4851	Thiago Giannini	1	[Icons]
120	R_rxn02476 / 3-phosphoshikimate 1-carboxyvinyitransferase - Confirmar de Reação Importada	OPEN	Metabolic Network CCBH4851	Thiago Giannini	0	[Icons]
121	R_rxn00695 / glucose-1-phosphate adenylyltransferase - Confirmar de Reação Importada	OPEN	Metabolic Network CCBH4851	Thiago Giannini	0	[Icons]
122	Validação da Reação de Biomassa do Mod. V1 CCBH4851	OPEN	Metabolic Network CCBH4851	Thiago Giannini	0	[Icons]
16	Gene "astA"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
17	Gene "astB"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
18	Gene "oprD"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
20	Proteína "BqsS"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
21	Gene "astR"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	3	[Icons]
22	nirG	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
23	Gene "pscK"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
24	Gene "flpR"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]

At the bottom of the table, there is a pagination control showing page 1 of 9, and a "New" button. The footer of the page contains the copyright information "Copyright@Thiago Giannini Ramos - IOC / PROCC / Fiocruz" and the timestamp "17:39:26 | 28/05/2018".

Figura 29 - Página de Desafios (unidade de apresentação “List of challenges”)

Na página de desafios, o *status* de cada desafio é marcado pelo uso de cores distintas. Os desafios grafados em cor verde são de *status* “COMPLETED” e os de fonte na cor preta são de *status* “OPEN”.

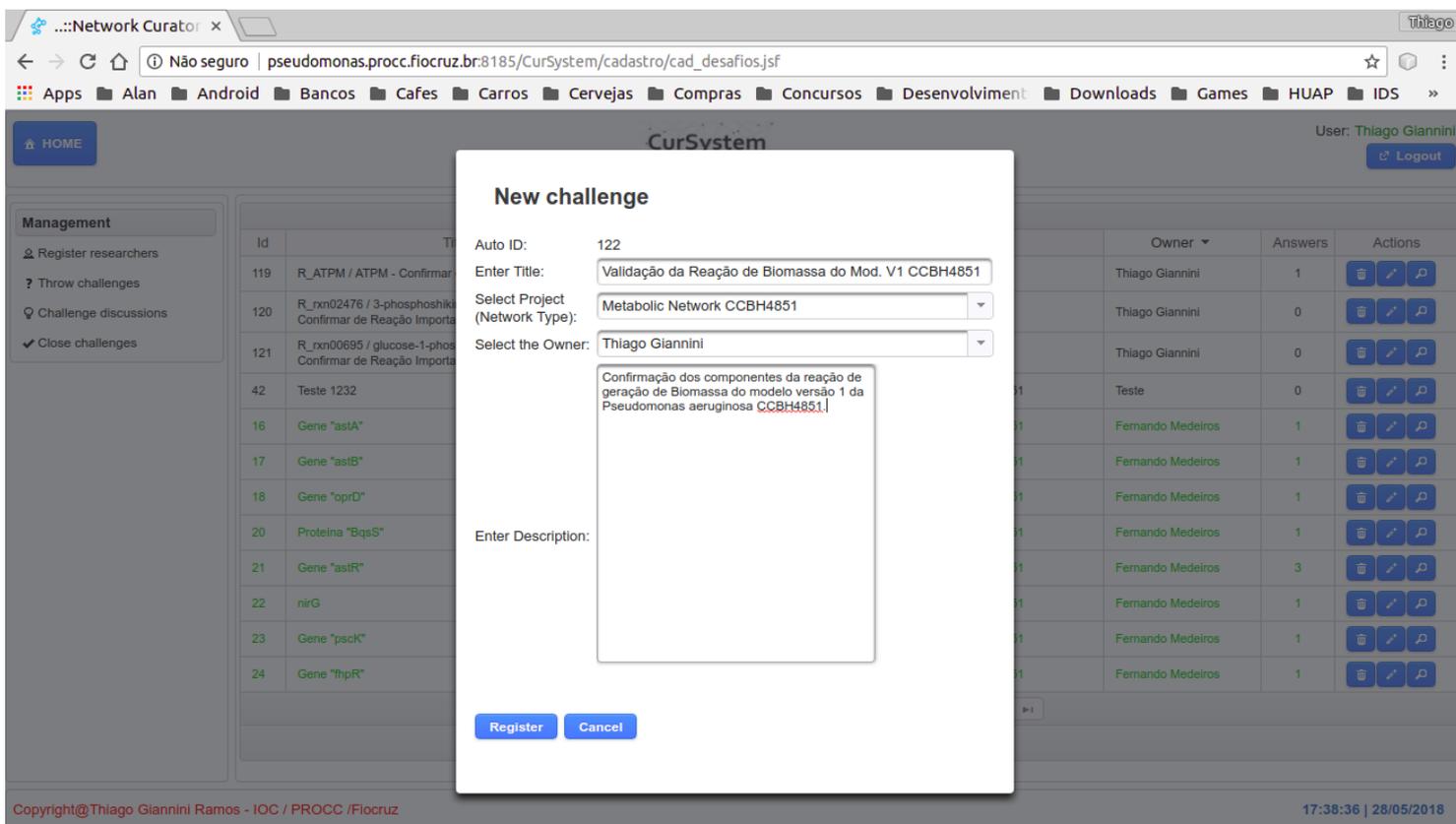


Figura 30 - Janela de diálogo para cadastro de desafios (unidade de apresentação “*New challenge*”)

Para a interface do modelo de interação de respostas e comentários dos desafios também houve necessidade de dividir a página em mais de uma unidade de apresentação: a página “*Challenge discussions*” e as janelas de diálogo “*Answer to*” e “*Discussion list*”. As três unidades de apresentação são exibidas nas Figura 31, Figura 32 e Figura 33, respectivamente.

...:Network Curator x Thiago

pseudomonas.procc.fiocruz.br:8185/CurSystem/analise/challenges_discussions.jsf

Apps Alan Android Bancos Cafes Carros Cervejas Compras Concursos Desenvolvimento Downloads Games HUAP IDS

HOME CurSystem User: Thiago Giannini Logout

Management

- Register researchers
- Throw challenges
- Challenge discussions
- Close challenges

Challenges in your projects

- Gene pqrC
- Gene pqrB
- Gene PA1560
- Gene PA0202
- Gene PA1559
- Gene ack
- Gene mexZ
- Gene mvlA
- Gene mexY
- Gene mexX
- Gene qrh
- Gene oprD
- Gene qsc102
- Gene pstI
- Gene phzC2
- gene "ivy"
- pqrA
- Gene phzC2
- phnC2
- R_ATPM / ATPM - Confirmar de Reação Importada
- R_rxn02476 / 3-phosphohikimate 1-carboxyvinyltransferase - Confirmar de Reação Importada
- R_rxn00695 / glucose-1-phosphate adenyltransferase - Confirmar de Reação Importada

Description of the challenge

R_ATPM / ATPM - Confirmar de Reação Importada

Reação foi importada do Modelo da P. aeruginosa PAO1 como sendo essencial para geração de Biomassa. Essa é uma reação que se encontra em todos os modelos da P. aeruginosa PAO1, desde a versão de 2008. Não tem explicação nos artigos de sua origem.

Reação: $(1.0)H_2O + (1.0)ATP \Rightarrow + (1.0)H + (1.0)Phosphate + (1.0)ADP$

Equação: $(1.0)H_2O + (1.0)C_{10}H_{13}N_5O_{13}P_3 \Rightarrow + (1.0)H + (1.0)HO_4P + (1.0)C_{10}H_{13}N_5O_{10}P_2$

Reply challenge

Answer frame

Selected Challenge Response List

Id	Title	Researcher	*
109	R_ATPM / ATPM - Confirmar de Reação Importada	Márcio Argollo	

Copyright@Thiago Giannini Ramos - IOC / PROCC / Fiocruz 17:30:03 | 28/05/2018

Figura 31 – Página de respostas para soluções possíveis de cada desafio (unidade de apresentação “Challenge discussions”)

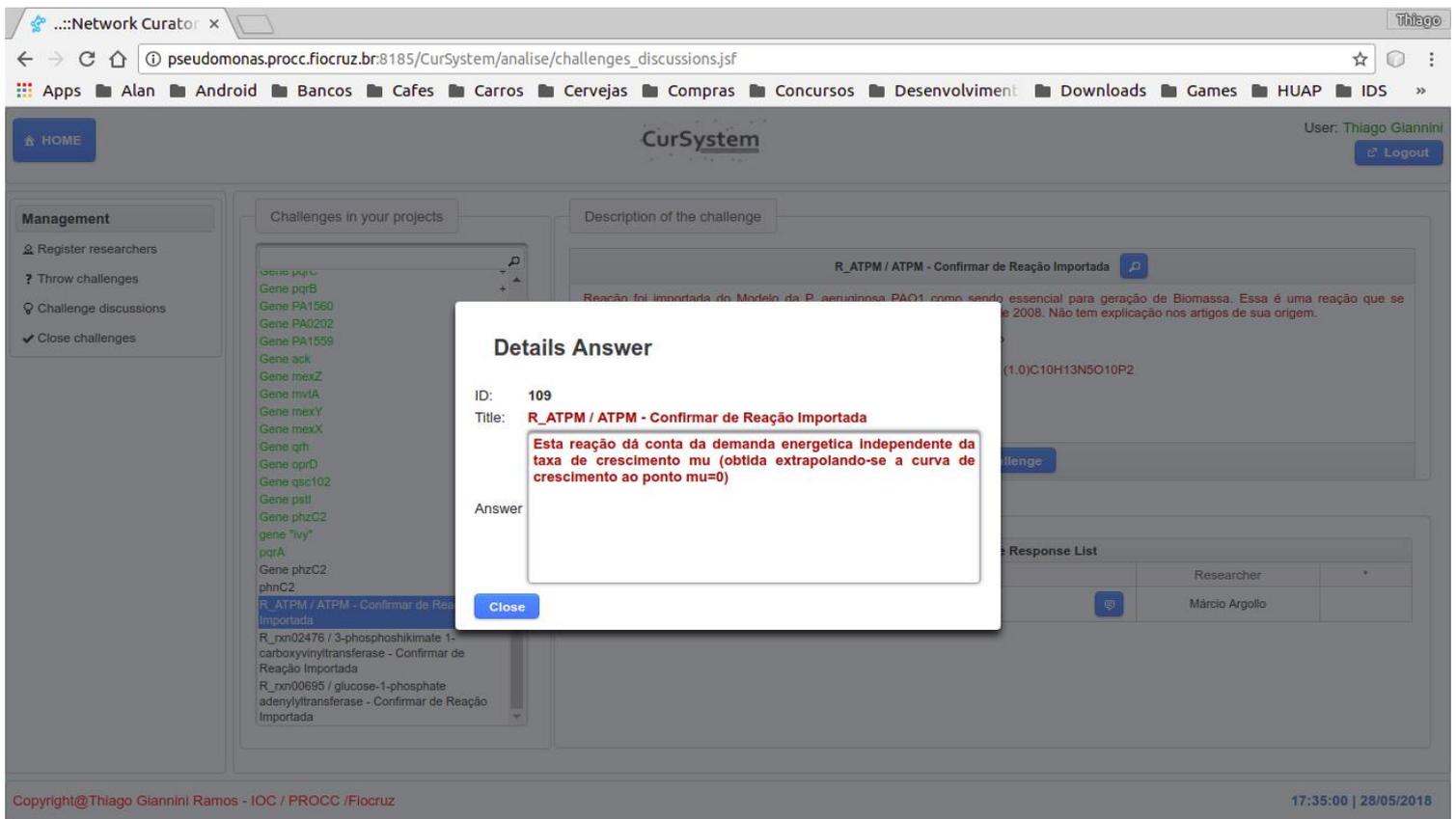


Figura 32 - Janela de diálogo para cadastro da resposta (unidade de apresentação "Answer to")

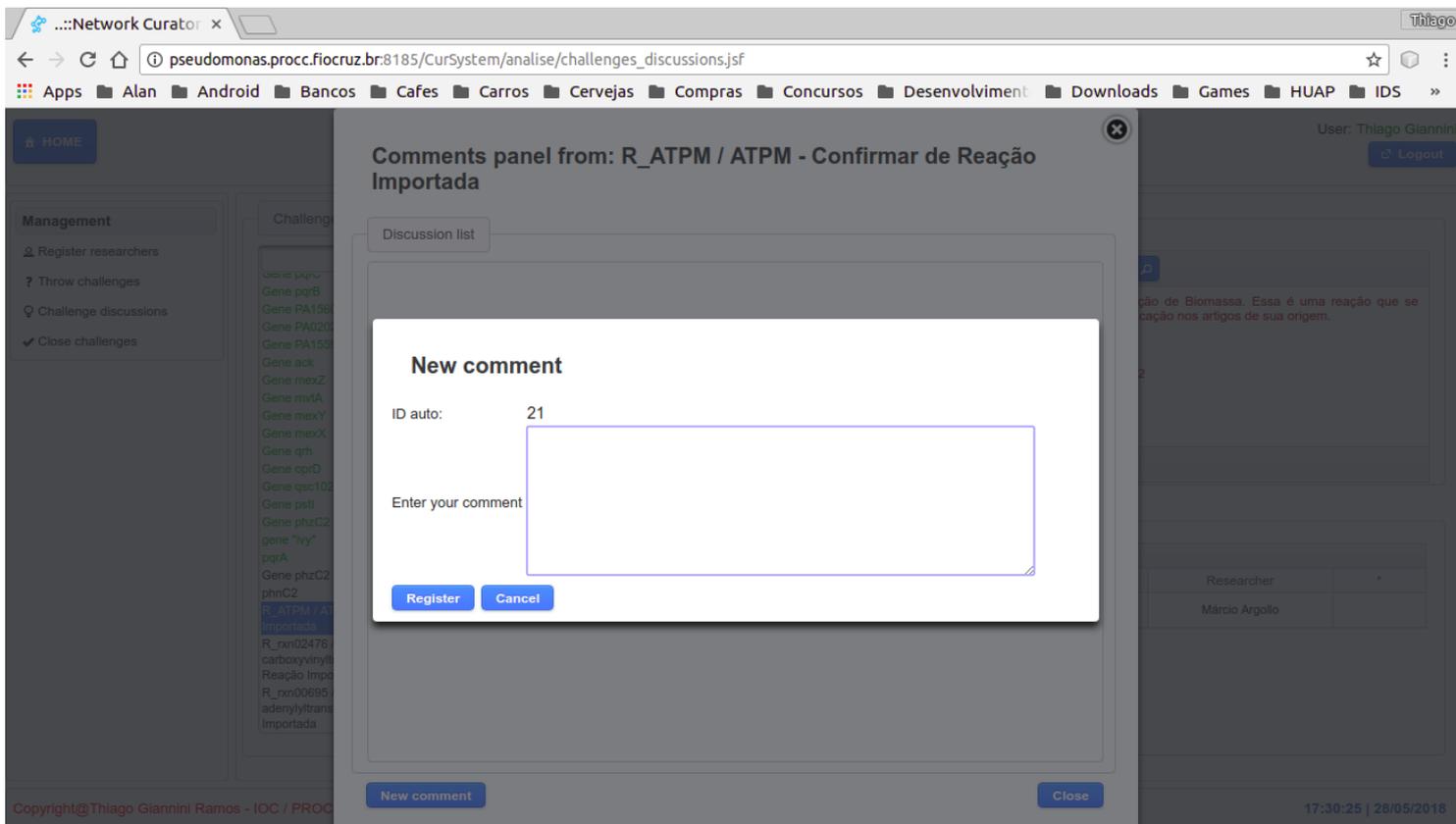


Figura 33 - Janela de diálogo para cadastro de comentários de respostas (unidade de apresentação "Discussion list")

Na interface de resolução de desafios o usuário analisa as respostas e seus comentários, escolhe uma resposta e a valida com uma justificativa. Para esse fluxo de operações há o modelo de interação da Figura 26. As unidades de representação são: a página "Close Challenges" (Figura 34) e a janela de diálogo "Justification for answer validation" (Figura 35).

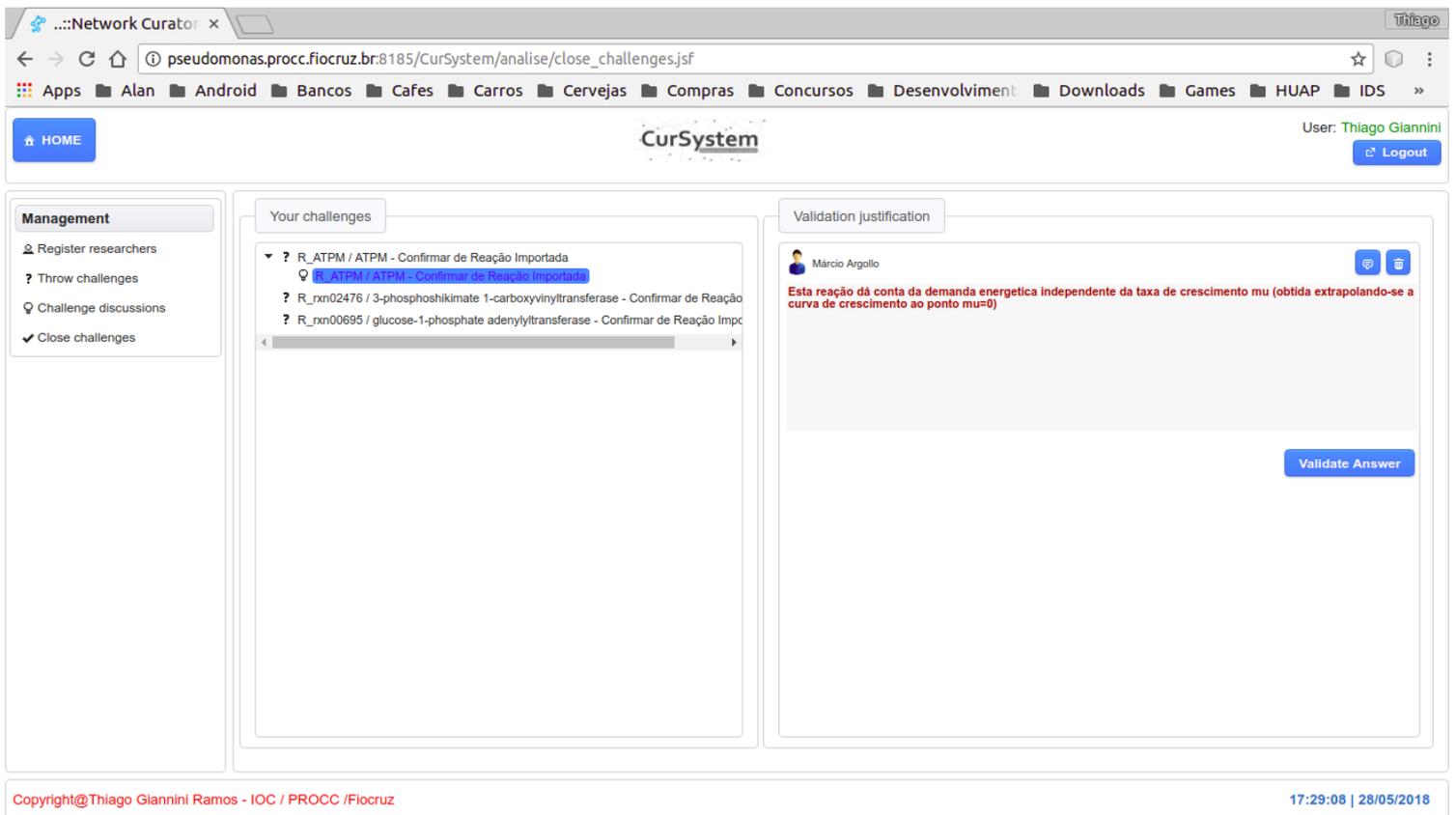


Figura 34 – Página de validação de respostas (unidade de apresentação "Close challenges")

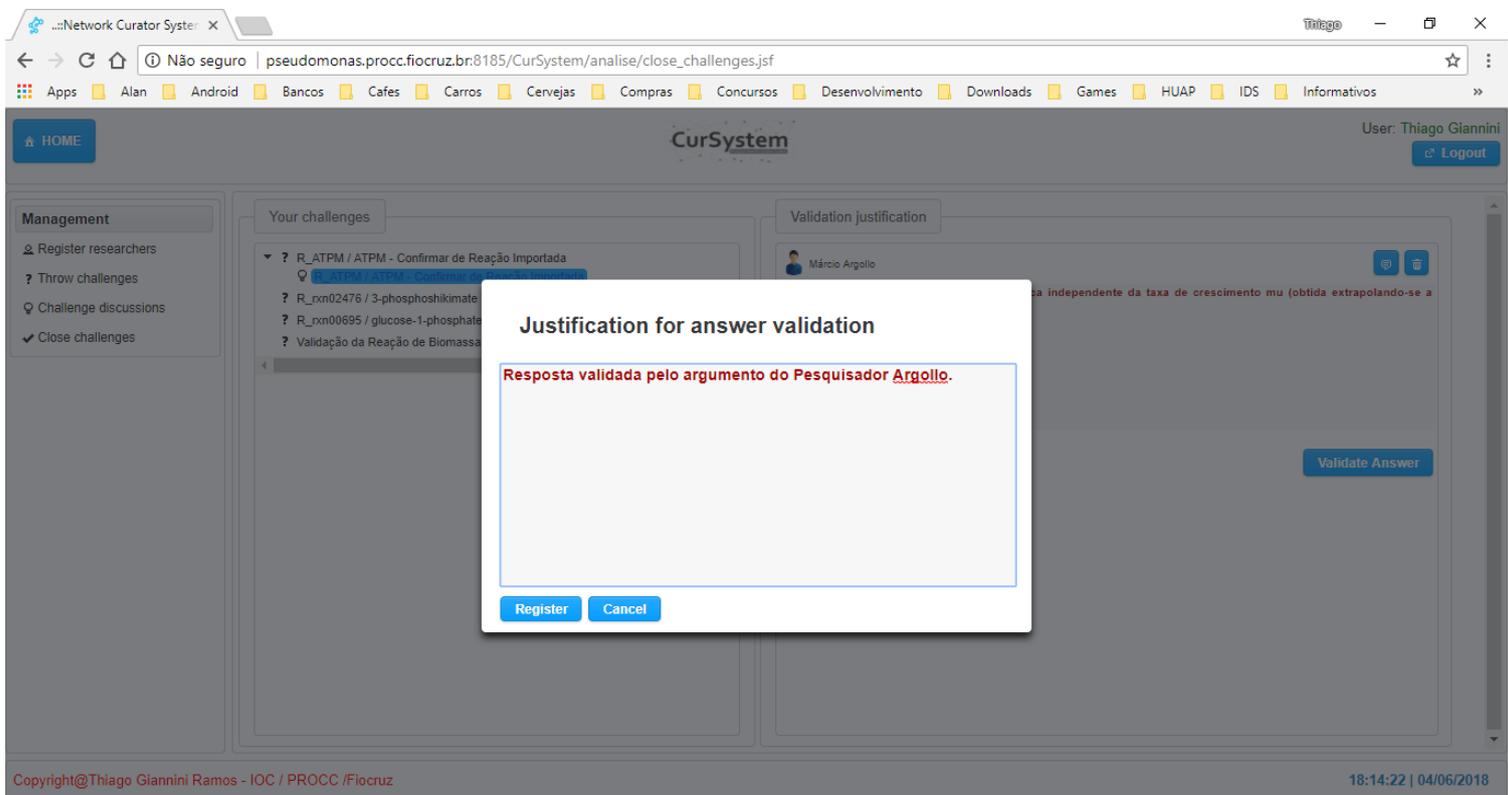


Figura 35 - Janela de diálogo para justificar e validar resposta

3.3 Etapa 3 - Avaliação inicial através de métodos matemáticos e computacionais

Ao fim da execução dos dois passos da metodologia adaptada, é preciso avaliar o modelo reconstruído e verificar se os objetivos foram alcançados. Nesse contexto, o objetivo é a construção de um modelo computacional da rede metabólica da *P. aeruginosa* CCBH4851 que simule a geração de biomassa maior que zero. O NETGen traz em sua implementação um módulo que executa avaliação preliminar do modelo, porém é simples e limitado. Existem outras ferramentas mais sólidas que também se destinam a essa tarefa que foram utilizadas como suporte adicional da avaliação computacional para refinar os resultados.

A avaliação computacional/matemática consiste na utilização de métodos matemáticos implementados em linguagem de programação. A área da matemática que dispõe desses métodos é chamada de programação linear e tem como seu principal objetivo a otimização de problemas de fluxos (7,12,32). A otimização é uma ferramenta precursora para avaliações de fluxos de modelos de redes metabólicas (30).

Para executar avaliação dos modelos gerados por este trabalho, fez-se uso do *FBA* (descrito na seção 1.3.1) (34) e o *Gapfilling* (descrito na seção 1.3.4) da implementação *COBRApy* (seção 1.3.3) do COBRA (14). O *FBA* efetua cálculos dos fluxos das reações levando em consideração um estado estacionário (32). Esse tipo de abordagem nos permite verificar os resultados da reação de geração de biomassa descrita no modelo. O *Gapfilling* auxilia em casos em que a taxa de biomassa gerada da versão do modelo é zero. O *Gapfilling* compara o modelo alvo com o modelo de referência e analisa quais reações poderiam complementar o modelo alvo para simular a geração de biomassa diferente de zero (14). Dentro de um conjunto de soluções possíveis, o método sempre sugere o menor número de reações complementares para geração de biomassa no modelo alvo.

Para a implementação da integração com o NETGen, o *COBRApy* também foi o escolhido (14). A entrada (*input*) para os programas de análise é o arquivo no formato *SBML*. Descritos nas seções anteriores, o resultado (produto) da aplicação da metodologia de reconstrução executada é o modelo alvo em formato *SBML* (37). O *SBML* é o formato padrão do *COBRApy*, portanto, facilita a execução da avaliação preliminar da rede metabólica gerada implementada no NETGen. Na Figura 36 observa-se como um código conciso permite analisar a reação de geração de

biomassa do modelo, assim como os números de reações, metabólitos e genes mapeados na reconstrução (14).

```
testeCCBH4851.py x
1  #!/usr/bin/python
2
3  +import ...
11 pandas.options.display.max_rows = 100
12
13  data_dir = os.getcwd()
14
15  sbml_file = str(sys.argv[1])
16  nomeRede = str(sys.argv[2])
17
18  ccbh = cobra.io.read_sbml_model(sbml_file)
19
20  #print ccbh.reactions.get_by_id("PA01_Biomass")
21
22  print ('----- CCBH4815 -----')
23
24  taxaC = ccbh.optimize().f
25  print ('ID: %s') % nomeRede
26  print ('biomass generated: %f mmol/g DCW/h') % (taxaC)
27  tx_em_horas = log(2.0)/taxaC
28  tx_em_minutos = tx_em_horas * 60
29  print ('biomass produced in hours: %f') % (tx_em_horas)
30  print ('biomass produced in minutes: %f') % (tx_em_minutos)
31  print ('-----Reconstruction numbers-----')
32  print ('reactions: %i') % len(ccbh.reactions)
33  print ('metabolites: %i') % len(ccbh.metabolites)
34  print ('genes: %i') % len(ccbh.genes)
```

Figura 36 - Script em Python utilizando o COBRAPy para análise de modelos metabólicos representados por arquivos em formato SBML

3.4 Base de Conhecimento da *Pseudomonas aeruginosa* CCBH4851

Uma reconstrução de modelos de redes metabólicas não necessariamente precisa de base de dados para armazenar informações (7). Previamente salientou-se que o resultado do processo de reconstrução geralmente é um arquivo representando o modelo gerado, como o SBML (37). Nesse trabalho, optou-se por projetar uma base de dados para guardar as informações de todas reconstruções. Esta base de dados é intitulada de Base de Conhecimento da *P. aeruginosa* CCBH4851. A intenção da construção de tal base é manter as informações mais recentes dos modelos da bactéria e partilhá-las com a comunidade científica.

Quando um modelo está em processo de construção pelo NETGen, a existência das reações do metabolismo é verificada no banco de dados. Em caso positivo, institui-se uma ligação com elas (relacionamentos). Caso contrário, a(s) reação(ões) é(são) inserida(s) e depois a ligação é feita. Essa arquitetura possibilita a geração de uma versão para cada modelo guardado na base de conhecimento, reutilizando todas as outras informações através de referências como reações, metabólitos e associações Gene-Proteína-Reação (GPR) (7). Esse tipo de abordagem aumenta o desempenho do sistema, pois a cada nova reconstrução executada, mais informações serão guardadas na base de conhecimento. Conseqüentemente, o tempo de processamento vai diminuir para novas reconstruções.

Além das informações dos modelos metabólicos, também são armazenados dados relacionados aos componentes do organismo alvo. As informações são: cepa; genoma, gene, proteínas, enzimas e links de referências para outras bases de dados. Outras informações projetadas e agregadas na base de conhecimento são os dados processados pelo CurSystem. Informações de “Desafios”, “Respostas”, “Comentários” e “Usuários” foram adicionadas. A base de conhecimento funciona com uma camada intermediária entre os dois sistemas (Figura 37). O NETGen deposita dados criando modelos, o CurSystem altera dados evoluindo modelos.

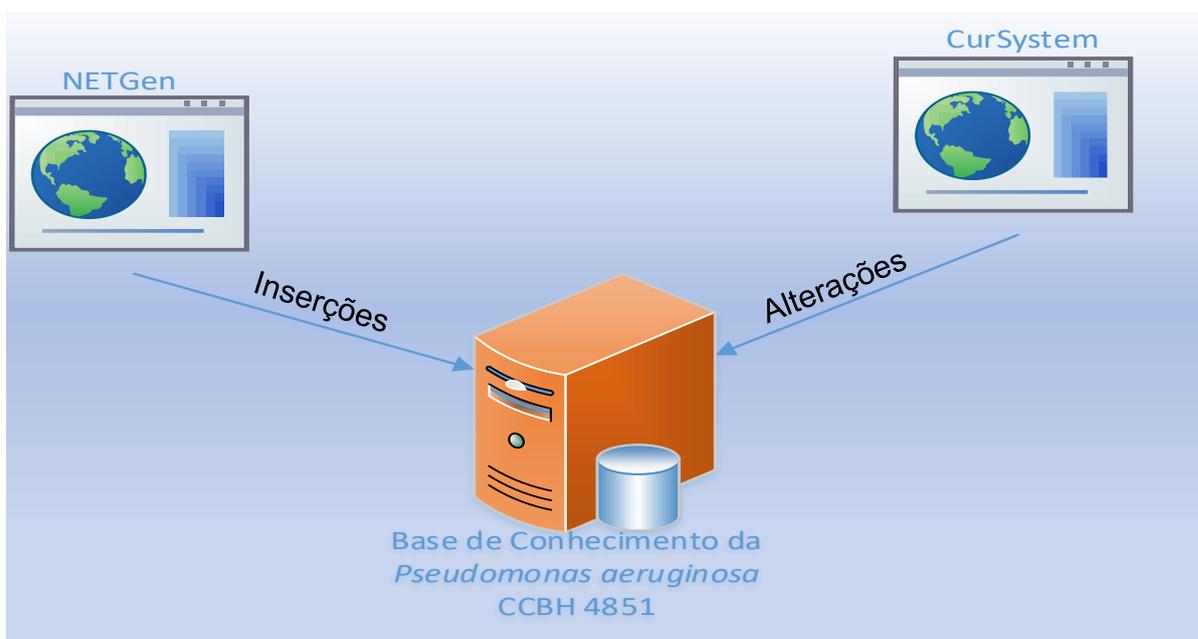


Figura 37 – Arquitetura da interconexão entre os sistemas NETGen, CurSystem e a Base de Conhecimento da *P. aeruginosa* CCBH4851

A ligação entre os dois sistemas é feita através da Entidade “*Reaction*” como se pode observar no Modelo ER (Entidade e Relacionamento) da Base de Conhecimento revelado na Figura 38. Como já descrito na seção 3.2, uma solução candidata à resolução do desafio, quando aprovada pelo responsável, pode validar, modificar ou inserir uma nova reação no modelo, compactuando com sua evolução.

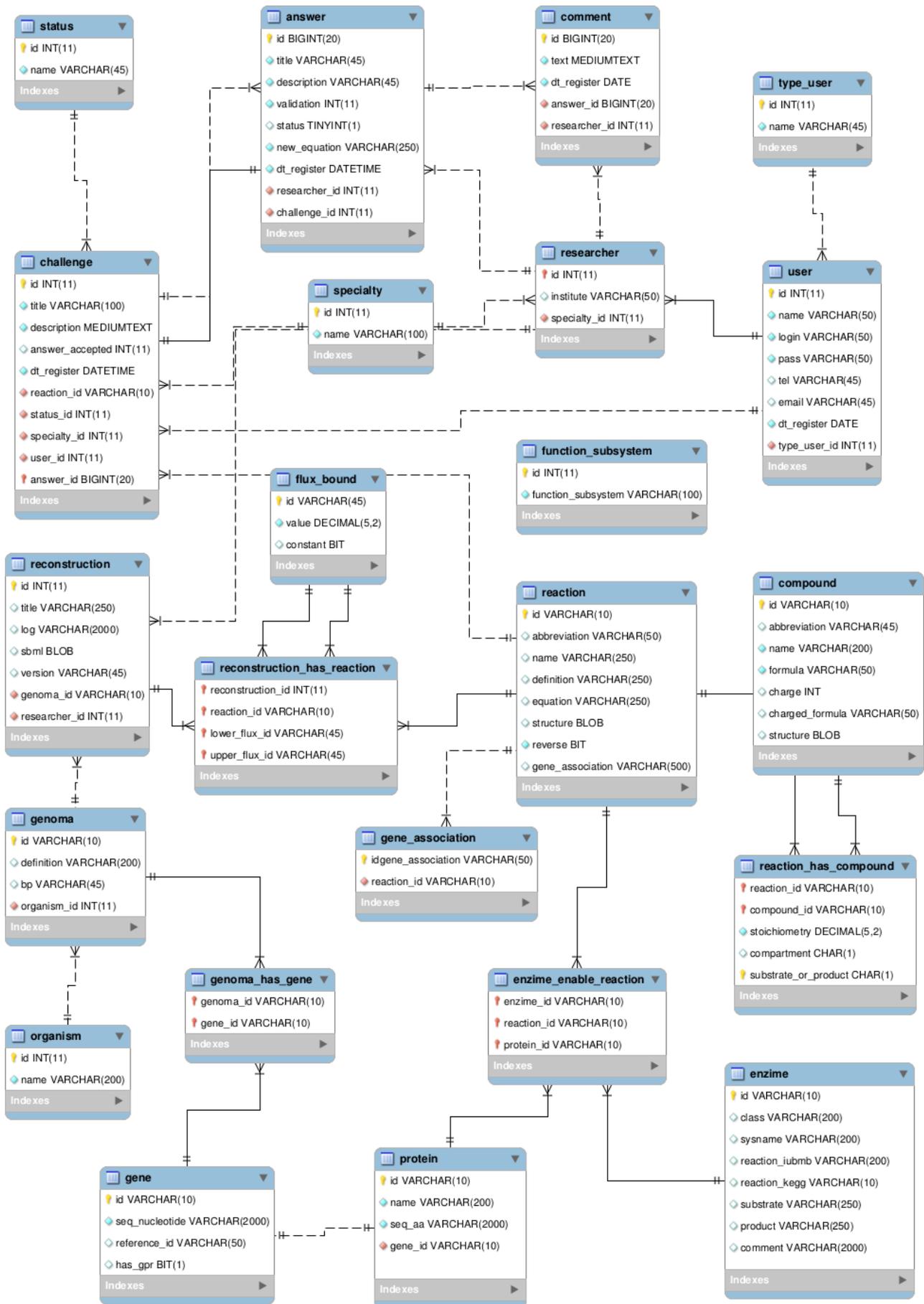


Figura 38 - Modelo de Entidade e Relacionamento (ER) da Base de Conhecimento da *Pseudomonas aeruginosa* CCBH4851

4 RESULTADOS

Os resultados obtidos a partir do uso da metodologia descrita podem ser agrupados em dois conjuntos distintos. Cada conjunto de resultados está relacionado a diferentes etapas do processo de reconstrução da rede metabólica. O primeiro conjunto é fruto da etapa de reconstrução automática da rede, obtida a partir de algumas execuções no sistema NETGen. Este modelo, chamado mCCBH4851_v1, não é capaz de gerar biomassa.

Os demais resultados descrevem uma rede metabólica com geração de biomassa maior que zero, chamada de mCCBH4851_v2. Esta rede é produto das inferências feitas por especialistas a partir da utilização do sistema CurSystem para o aperfeiçoamento da mCCBH4851_v1. Para fins didáticos, ao final deste capítulo será apresentada uma relação de todas reações agrupadas dos subsistemas mCCBH4851_v1 e mCCBH4851_v2.

4.1 Reconstrução Inicial

Entre os modelos que poderiam ser utilizados como referências para a reconstrução inicial da rede metabólica do organismo estudado, havia três opções consolidadas. Em 2008, foi proposto um modelo (12) cujo objetivo consistiu em mapear o metabolismo básico da *Pseudomonas aeruginosa* PAO1 a fim de obter geração de biomassa. No ano de 2017, foi publicado um modelo estendido da *P. aeruginosa* PAO1 (6), cuja intenção foi mapear não apenas o metabolismo básico (geração de biomassa), mas também os mecanismos de resistência. Em paralelo ao modelo da PAO1 de 2017, os mesmos pesquisadores reconstruíram um modelo para a cepa PA14, contemplando seus fatores de virulência (6).

Doravante, os modelos da *P. aeruginosa* PAO1 publicados em 2008 e 2017, serão abordados pelas suas respectivas identificações, mPAO1_v2008 (iMO1056), mPAO1_v2017 (iPAE1146). O modelo da *P. aeruginosa* PA14 também publicado em 2017 será identificado como mP14 (iPAU1129).

Na Tabela 2, observa-se uma comparação dos modelos descritos segundo seus números de genes, reações, metabólitos e biomassa. Note-se que os modelos mPAO1_V2017 e mPA14 apesar de referirem-se a cepas distintas, apresentam valores próximos entre si. Entretanto, tais modelos geram valores bastante distintos

do modelo MPAO1_2008, especialmente no que se refere aos valores de metabólitos e geração de biomassa. Tal diferença se justifica em função da inclusão de novos genes e reações no mapeamento realizado na reconstrução em cada um dos modelos de 2017.

	mPAO1_v2008	mPAO1_v2017	mPA14
Nº. Genes	1054	1148	1132
Nº. Reações	1110	1496	1495
Nº. Metabólitos	879	1284	1286
Biomassa ²	1,047929	15,509635	15.508373

Tabela 2 – Dados dos diferentes modelos reconstruídos da *P. aeruginosa* PAO1 (6,12).

O modelo de referência para a reconstrução da rede metabólica da *P. aeruginosa* CCBH4851 realizada nesse trabalho foi o mPAO1_v2017 (6). O critério de escolha foi definido em função dos objetivos a serem alcançados e da quantidade de informações genéticas mapeadas nos modelos. O mPAO1_v2017 é uma atualização do mPAO1_v2008 que, além de mapear reações de metabólitos precursores de biomassa, adiciona genes para fatores de virulência e resistência a antibióticos, sendo, portanto, o modelo com mais informação genética disponível.

No modelo mCCBH4851_v1, primeiro modelo de rede metabólica reconstruído pelo NETGen, obteve-se o mapeamento de 1446 reações, 1268 metabólitos e 1139 genes (Tabela 3). Seu valor de geração de biomassa foi de 0.0000 mmol/g DCW/h. Este resultado indicou que o modelo gerado não foi capaz de simular o fluxo básico de reações para produção dos metabólitos essenciais para duplicação da bactéria.

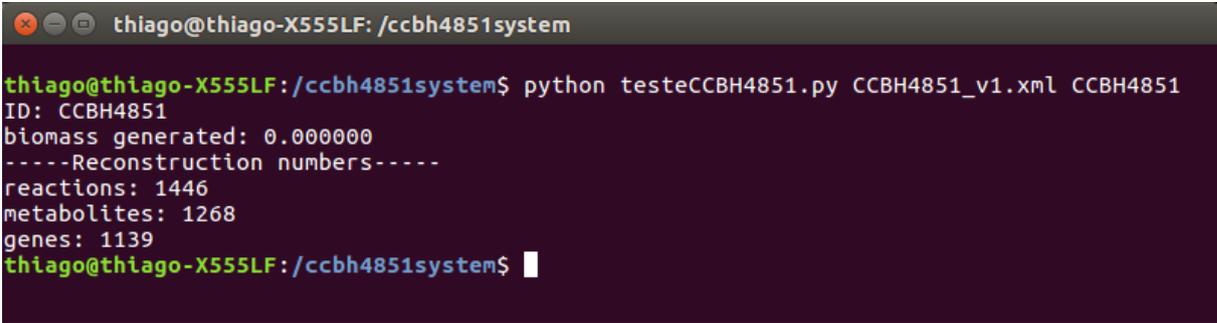
	Genes	Reações	Metabólitos	Biomassa
mCCBH4851_v1	1139	1446	1268	0.00*

*mmol/g DCW/h

Tabela 3 – Dados mapeados da primeira rede metabólica reconstruída da *P. aeruginosa* CCBH4851 no NETGen.

² Medida dada em milimoles por grama de peso seco de célula por hora (mmol/g DCW/h).

Os números extraídos e calculados pelo sistema NETGen (Figura 40) foram validados no pacote *COBRapy* (14). O *script* de cálculo da geração de biomassa é exibido na Figura 36 e os resultados de saída deste *script* são mostrados na Figura 39.



```
thiago@thiago-X555LF: /ccbh4851system
thiago@thiago-X555LF:/ccbh4851system$ python testeCCBH4851.py CCBH4851_v1.xml CCBH4851
ID: CCBH4851
biomass generated: 0.000000
----Reconstruction numbers----
reactions: 1446
metabolites: 1268
genes: 1139
thiago@thiago-X555LF:/ccbh4851system$
```

Figura 39 - Resultados da análise da primeira rede gerada (mCCBH4851_v1) feita no pacote *COBRapy*

Na Figura 40, verifica-se o e-mail com o relatório gerado pelo NETGen reafirmando a igualdade de resultados apontados nas duas análises descritas.

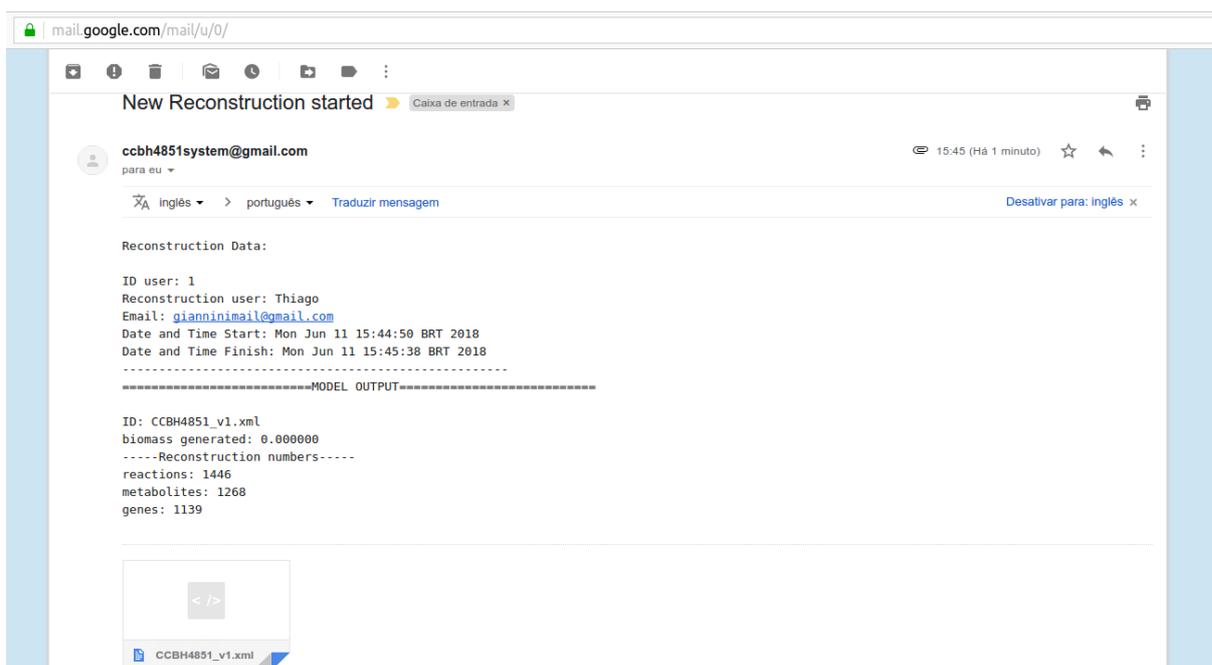


Figura 40 - E-mail com o resultado e análises gerados pelo NETGen

4.2 Gaps para geração de Biomassa e Curadoria Assistida

Os resultados atingidos na etapa anterior permitiram a obtenção de um modelo de rede metabólica da *P. aeruginosa* CCBH4851 mais acurado, o mCCBH4851_2. Descrita na seção 1.3, a análise de lacunas da rede permite solucionar falhas no modelo gerado pelo NETGen, completando o ciclo de reações

que fazem parte do metabolismo básico (7). O *Gapfilling* do *COBRAPy* foi utilizado para indicar quais reações não foram encontradas no modelo de referência (mPAO1_v2017), auxiliando na sinalização de quais reações seriam necessárias para o modelo mCCBH4851_v1 gerar biomassa.

Na Tabela 4 constam as reações que resultaram da análise de lacunas do *COBRAPy*. Tais reações foram consideradas essenciais para que o modelo mCCBH4851_v1 pudesse simular a geração de biomassa.

ID SBML	Associação	EC Number	Nome	Subsistema
R_rxn02476	PA3164	2.5.1.19	3-phosphoshikimate 1-carboxyvinyltransferase	Biossíntese de fenilalanina, tirosina e triptofano
R_rxn00695	-	2.7.7.27	glucose-1-phosphate adenylyltransferase	Metabolismo de amido e sacarose
R_rxn05468	-	-	O2 transport in via diffusion	Transporte de Oxigênio
R_ATPM	-	-	pseudo-reação	Demanda Energética

Tabela 4 - Reações apontadas pela análise de lacunas no Modelo mCCBH4851_v1

Para a segunda reconstrução da rede metabólica, utilizou-se o módulo de inferência de genes implementado e descrito na seção 3.1.1. Somente as reações indicadas pela análise de lacunas do *COBRAPy* foram adicionadas manualmente na segunda reconstrução.

O acréscimo de reações na primeira versão do modelo modificou positivamente a geração de biomassa, porém, o número de metabólitos permaneceu inalterado. Este fato indica que possivelmente novas rotas foram ajustadas para que o fluxo de metabólitos pudesse resultar na geração de biomassa (12).

Os números obtidos no mapeamento foram: 1450 reações, 1268 metabólitos e 1139 genes. A simulação de geração de biomassa resultou em 12.470973 mmol/g

DCW/h. Este valor, quando convertido para taxa de duplicação, resulta em aproximados 3.33 minutos (0.055 horas). O relatório com o resultado encontra-se na Figura 41.

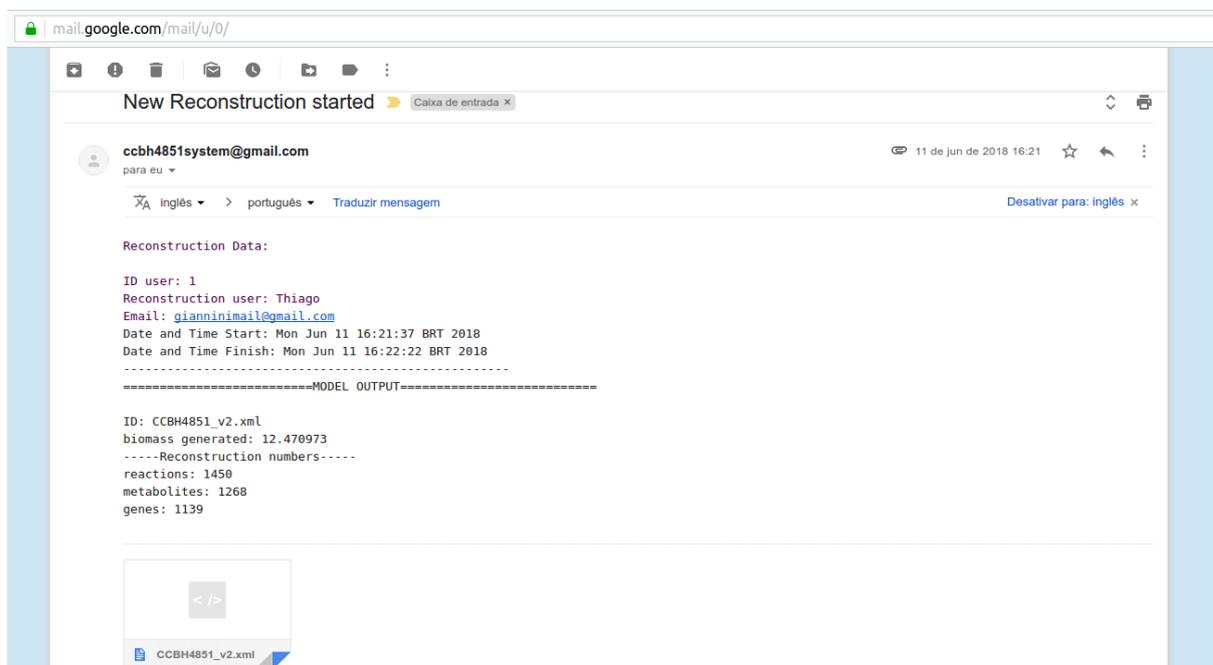


Figura 41 - Relatório da segunda versão gerado do modelo.

Para fins de comprovação dos resultados obtidos pelo NETGen, obteve-se uma contraprova a partir do *COBRAPy*. Tal como o modelo mCCBH4851_v1, o modelo mCCBH4851_v2 foi executado no *COBRAPy* a fim de verificar se seus produtos permaneceriam inalterados. A saída que resultou da execução desse método pode ser visualizada na Figura 42. Observa-se que há uma coincidência (validação) entre os números descritos no relatório do NETGen enviado por e-mail e a saída do *COBRAPy*.

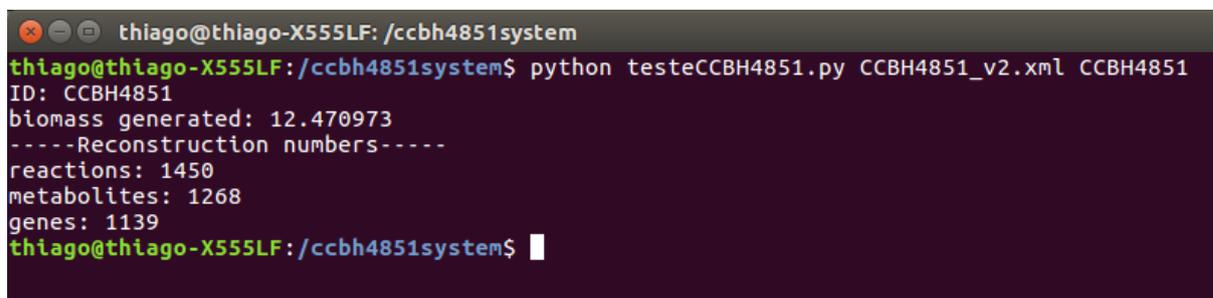


Figura 42 - Resultado do *COBRAPy* utilizando o arquivo *SBML* da segunda versão do modelo (mCCBH4851_v2)

O segundo modelo gerado da rede metabólica (mCCBH4851_v2) mostrou-se mais interessante se comparado ao primeiro. A simulação do modelo apontou um resultado de geração de biomassa acima de zero. A Tabela 5 expõe a diferença entre as duas reconstruções executadas no NETGen. A única diferença entre os dois modelos é de quatro reações, ou seja, as quatro reações inferidas manualmente por indicação da análise de lacunas.

	Genes	Reações	Metabólitos	Biomassa
mCCBH4851_v1	1139	1446	1268	0.00*
mCCBH4851_v2	1139	1450	1268	12.470973*

*mmol/g DCW/h

Tabela 5 - Comparação entre os dois modelos gerados no NETGen

As quatro reações que foram incluídas por inferência no modelo estão em fase de análise por especialistas através do método de análise suportado pelo sistema *WEB* CurSystem. Cada reação foi exposta por desafios lançados no sistema, para que os pesquisadores possam discutir e reafirmar a existência destas reações no modelo. Com isto, espera-se eliminar as lacunas ocasionadas pela reconstrução automática. A seguir, na Figura 43, observa-se a imagem recortada da página de desafios do CurSystem com as reações descritas.

List of challenges						
Id	Title ↕	Status ↕	Project (Network Type) ↕	Owner ▾	Answers	Actions
119	R_ATPM / ATPM - Confirmar de Reação Importada	COMPLETED	Metabolic Network CCBH4851	Thiago Giannini	1	  
120	R_rxn02476 / 3-phosphohikimate 1-carboxyvinyltransferase - Confirmar de Reação Importada	OPEN	Metabolic Network CCBH4851	Thiago Giannini	0	  
121	R_rxn00695 / glucose-1-phosphate adenyltransferase - Confirmar de Reação Importada	OPEN	Metabolic Network CCBH4851	Thiago Giannini	0	  
122	Validação da Reação de Biomassa do Mod. V1 CCBH4851	OPEN	Metabolic Network CCBH4851	Thiago Giannini	0	  
123	rxn05468 / O2 transport in via diffusion - Confirmar de Reação Importada	OPEN	Metabolic Network CCBH4851	Thiago Giannini	0	  

Figura 43 - Reações apontadas na análise de *gap* cadastradas no CurSystem

A reação com *ID* número 119, é um exemplo de reação que foi validada através do desafio respondido no sistema. O desafio encontra-se em fonte de cor verde, indicando que foi resolvido e a reação validada. Almeja-se que outras reações possam ser validadas e até mesmo incrementadas através do CurSystem conforme descrito na seção 3.2.

O arquivo do modelo está disponível para download no portal Pseudomonas.com (<http://pseudomonas.procc.fiocruz.br/index.php/data/ccbh4851->

[metabolic-network/download/6-metabolic-network-data/21-metabolic-network-v2](http://pseudomonas.procc.fiocruz.br/index.php/data/ccbh4851-metabolic-network/download/6-metabolic-network-data/21-metabolic-network-v2)) (8).

Os dados gerados e armazenados na base de conhecimento até o momento também estão disponíveis no portal. A princípio, não foi desenvolvida uma interface *web* para acesso aos dados. Foram criadas planilhas em *Excel* com todas as informações separadas por abas, seguindo a própria estruturação do banco de dados. Cada aba da planilha representa uma tabela na base de conhecimento. O arquivo pode ser acessado pelo link <http://pseudomonas.procc.fiocruz.br/index.php/data/ccbh4851-metabolic-network/download/6-metabolic-network-data/22-knowledge-base> (8).

4.3 Modelos mPAO1_v2008 x mPAO1_v2017 x mCCBH4851_v2

A reconstrução do modelo mPAO1_v2008 obteve o mapeamento de 1056 genes, 883 reações e 760 metabólitos (12). A atualização que resultou no modelo mPAO1_v2017 (6) acrescentou 90 genes, 610 reações e 524 metabólitos. A reconstrução executada neste trabalho resultou em um modelo com poucas diferenças em relação ao modelo de referência. O mCCBH4851_v2 alcançou os seguintes números no mapeamento: 1139 genes, 1450 reações e 1268 metabólitos e 576 EC numbers. A Tabela 6 apresenta a comparação dos dados mapeados de todos os modelos descritos.

	Genes	Reações	Metabólitos	EC numbers	Biomassa
mCCBH4851_v2	1139	1450	1268	576	12.470973*
mPAO1_v2008	1056	883	760	502	1,047929*
mPAO1_v2017	1146	1493	1284	674	15,509635*

*mmol/g DCW/h

Tabela 6 - Dados comparativos entre os modelos mCCBH4851_v2, mPAO1_v2008 e mPAO1_v2017 (6,12).

4.4 Reações mapeadas e agrupadas por subsistema

A maior parte das reações mapeadas foi identificada dentro do subsistema a qual pertence. Apenas doze reações tiveram seu subsistema desconhecido. Na Tabela 7 consta a lista de reações divididas por subsistemas.

Subsistema	Nº de Reações
valine, leucine and isoleucine degradation	17
purine metabolism	61
tryptophan metabolism	3
glyoxylate and dicarboxylate metabolism	13
nitrogen metabolism	13
alginate synthesis	3
one carbon pool by folate	3
none	12
terpenoid biosynthesis	8
aminoacyl-trna biosynthesis	7
aminobenzoate degradation	1
methane metabolism	1
pentose and glucuronate interconversions	2
peptidoglycan biosynthesis	7
cysteine and methionine metabolism	20
cyanoamino acid metabolism	1
amino sugar and nucleotide sugar metabolism	9
galactose metabolism	3
lysine degradation	3
glutamate metabolism	8
glycerolipid metabolism	90
glycerophospholipid metabolism	56
streptomycin biosynthesis	3
biotin metabolism	5
quorum sensing	8
lipopolysaccharide biosynthesis	37
phenylalanine, tyrosine and tryptophan biosynthesis	16
fructose and mannose metabolism	10
rhamnolipid synthesis	3
beta-alanine metabolism	2
biosynthesis of steroids	4
oxidative phosphorylation	19
pentose phosphate pathway	20
ascorbate and aldarate metabolism	1
valine, leucine and isoleucine biosynthesis	8
pyrimidine metabolism	34
porphyrin and chlorophyll metabolism	27

pantothenate and coa biosynthesis	14
exchange	171
butanoate metabolism	15
starch and sucrose metabolism	6
sulfur metabolism	6
terpenoid backbone biosynthesis	1
glycine, serine and threonine metabolism	20
benzoate degradation via coa ligation	1
carbohydrate metabolism	1
selenoamino acid metabolism	1
biosynthesis of siderophore group nonribosomal peptides	6
methionine metabolism	7
nicotinate and nicotinamide metabolism	10
penicillin and cephalosporin biosynthesis	1
propanoate metabolism	9
thiamine metabolism	6
alanine, aspartate and glutamate metabolism	11
arginine and proline metabolism	27
histidine metabolism	13
vitamin b6 metabolism	14
benzoate degradation	10
carbon metabolism	4
glycolysis / gluconeogenesis	18
biomass	1
glutathione metabolism	7
fatty acid metabolism	40
folate biosynthesis	14
citrate cycle (tca cycle)	15
transport	223
lysine biosynthesis	11
riboflavin metabolism	8
phenazine biosynthesis	8
metabolism of terpenoids and polyketides	1
glycerolipid synthesis	10
ubiquinone biosynthesis	21
fatty acid biosynthesis	147
phosphonate and phosphinate metabolism	2
pyruvate metabolism	17
urea cycle and metabolism of amino groups	7
tyrosine metabolism	8

Tabela 7 - Relação de reações mapeadas por subsistema do modelo mCCBH4851_v2

As reações que não tiveram seu subsistema identificado foram relacionadas na Tabela 8. Essas reações também foram incluídas como desafios no CurSystem para identificação por parte dos pesquisadores na curadoria da rede.

Reações com subsistema desconhecido	Status (CurSystem)
R_rxn02831	Em avaliação
R_rxn12206	Em avaliação
R_rxn04093	Em avaliação
R_rxn06216	Em avaliação
R_rxn00006	Em avaliação
R_rxn00206	Em avaliação
R_rxn13819	Em avaliação
R_rxn13810	Em avaliação
R_rxn06831	Em avaliação
R_rxn02875	Em avaliação
R_rxn05064	Em avaliação
R_ATPM	Avaliada

Tabela 8 - Relação de reações com subsistema desconhecido

5 DISCUSSÃO

Esquemáticamente, os resultados obtidos foram divididos em duas seções em razão de suas naturezas bastante distintas. A metodologia desenvolvida é considerada um resultado para a ciência computacional e será discutida na seção 5.1, analisando a implementação dos softwares e seus desempenhos ao processar as informações. Separadamente, na seção 5.2 serão discutidas as informações acerca da versão atual do modelo da rede metabólica da *Pseudomonas aeruginosa* CCBH4851.

5.1 Aspectos Computacionais

O primeiro aspecto a ser destacado refere-se à aplicabilidade da metodologia que foi empregada neste trabalho. Ambos sistemas adotados como ferramentas para a aplicação da metodologia foram estruturados para servirem a um amplo espectro de utilização. O NETGen destina-se a reconstruções de diferentes organismos e o CurSystem pode ser empregado em diferentes projetos que envolvem refinamento de informações de forma descentralizada por várias pessoas.

O sistema NETGen foi desenvolvido para dar suporte a reconstruções de modelos de redes metabólicas de qualquer tipo de organismo. O sistema não se restringe a cepas de *Pseudomonas aeruginosa*, sequer a bactérias. Os dois requisitos básicos impostos pela metodologia são as entradas de arquivos digitais que representem o organismo alvo e o organismo referência. Como descrito na seção 3.1.1, tais entradas do NETGen são obrigatórias. A primeira é a anotação no formato digital *GBK*, a segunda é um arquivo em formato digital do tipo *SBML* (37,52).

O sistema CurSystem também teve sua importância e generalidade de utilização confirmada. O sistema não está sendo utilizado somente para a curadoria da rede metabólica reconstruída neste trabalho pelo NETGen, mas também em outros projetos que necessitam de interlocuções entre pesquisadores. Como exemplo, menciona-se o projeto que está sendo desenvolvido pelo grupo de modelagem do IOC/Fiocruz que está trabalhando na modelagem da Rede Regulação Gênica da *Pseudomonas aeruginosa* CCBH4851. A Figura 44 mostra detalhes de desafios que foram criados e solucionados para este projeto-

Id	Title	Status	Project (Network Type)	Owner	Answers	Actions
31	Gene "hur"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
32	Gene "hif"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
33	Gene "pfa"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
34	Gene "acpP2"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
35	Gene "hifA"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
36	Gene "hifB"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
38	Gene "kinU"	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
41	Gene "adca"	OPEN	Gene Regulation Network CCBH4851	Ana Nascimento	1	[Icons]
43	Gene tpsB	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
44	Gene tpsA	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
45	Gene tpbR	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]
46	Gene PA4625	COMPLETED	Gene Regulation Network CCBH4851	Fernando Medeiros	1	[Icons]

Figura 44 - Desafios cadastrados para a Modelagem da Rede de Regulação Gênica da *P. aeruginosa* CCBH8451

5.1.1 Desempenho de processamento e Importância da base de conhecimento

O processamento executado pelo NETGen foi medido através de armazenamento de *log* (histórico) de dados, processo no qual registram-se os processamentos relevantes em um sistema computacional. Para a realização deste procedimento, delimitam-se dois marcos temporais, um no início e outro ao final da execução de todas as rotinas de uma reconstrução. Essas informações registradas em formato com data e hora são enviadas por e-mail para o usuário solicitante, juntamente com os dados da reconstrução.

A execução de uma reconstrução pôde ser classificada segundo duas tendências (vertentes) de desempenho. A reconstrução do organismo alvo pela primeira vez no sistema apresentou processamento completo das etapas mais demorado. Comparativamente, depois da primeira execução, todas as outras procederam de maneira mais rápida.

A reconstrução de um organismo alvo em relação ao qual se tem pouca ou nenhuma informação na base de conhecimento leva em média duas horas. Isto foi verdadeiro também para a reconstrução do modelo mCCBH4851_v1, tal como pode ser observado na Figura 45.

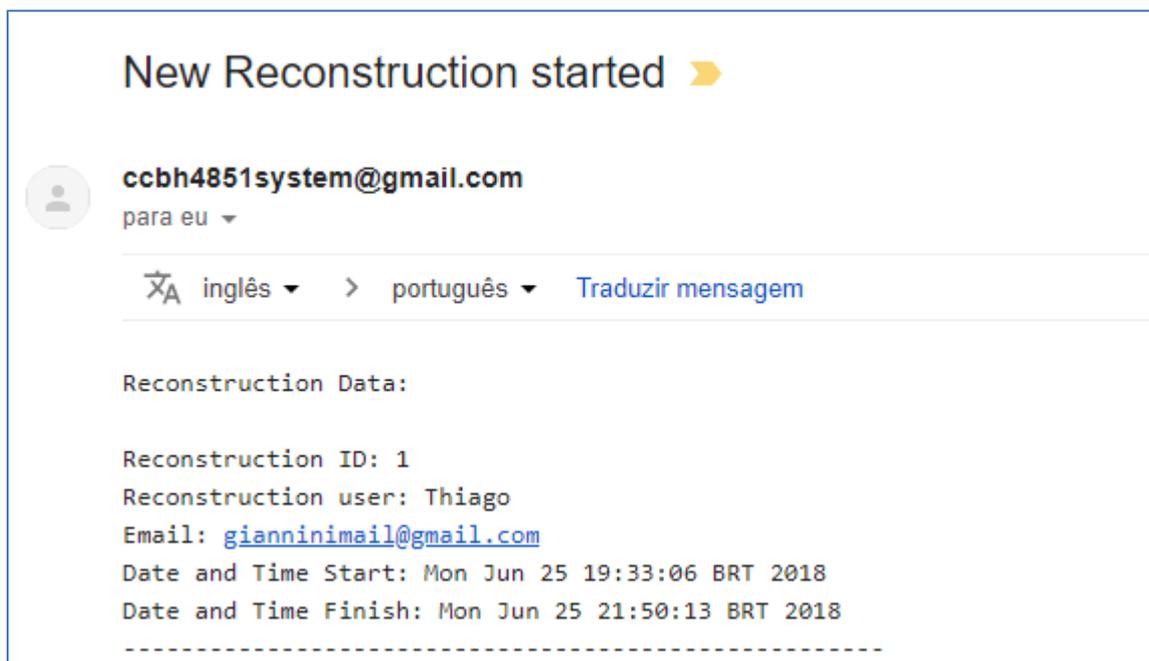


Figura 45 - Resultado de execução de uma reconstrução com pouca ou nenhuma informação na Base de Conhecimento

Este tempo de processamento elevado se justifica devido à ausência de qualquer tipo de informação do modelo na base de conhecimento. Graças à indisponibilidade de dados, o sistema opera buscas externas na base de conhecimento do KEGG. Cada reação mapeada inclusa no modelo pelo módulo do NETGen requereu a utilização do serviço *web* da plataforma para buscar informações adicionais de enzimas (26). Graças ao reaproveitamento dos registros da base possibilitado pela estrutura de versionamento desenvolvida no sistema, as reconstruções seguintes acontecem em tempo consideravelmente menor. Isto se confirma também nas reconstruções posteriores executadas para a *P. aeruginosa* CCBH4851. O modelo mCCBH4851_v2 levou em torno de um minuto para ser gerado, pois a maioria das informações sobre suas reações já estavam registradas na base de conhecimento gerada pela reconstrução anterior. Trata-se de um sistema de retroalimentação: cada nova reconstrução gera o armazenamento de informações que são reaproveitadas posteriormente por novas reconstruções, diminuindo a necessidade de buscas adicionais em plataformas externas. O resultado pode ser conferido na Figura 46.

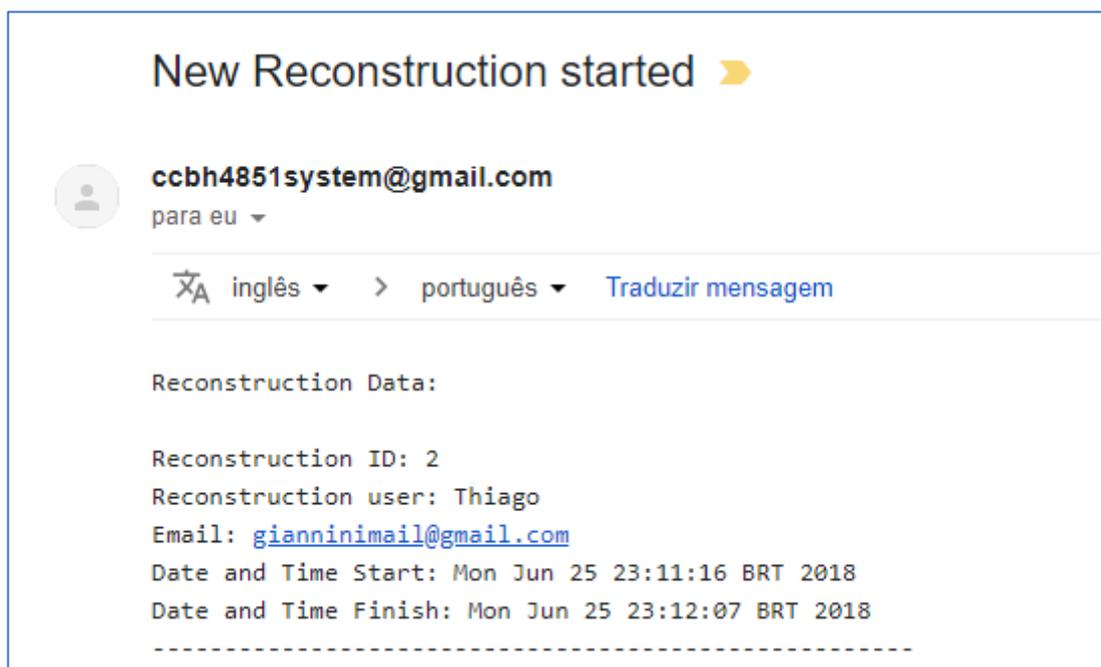


Figura 46 - Resultado de execução de uma reconstrução com a maioria das informações contidas na Base de Conhecimento

A criação de uma base de conhecimento própria para as reconstruções executadas pelo NETGen tinha como o propósito inicial a gênese de um repositório de dados para a cepa CCBH4851, objeto de estudo do grupo de pesquisa no qual este trabalho se insere. Porém, no decorrer do desenvolvimento e preenchimento dos dados, a criação desta base se mostrou primordial para o bom desempenho das várias execuções realizadas a fim de obter o melhor modelo da bactéria. Do ponto de vista computacional, tais diferenças de desempenho são importantes e apontam para a relevância da criação da base de conhecimento.

5.2 Aspectos do Modelo Biológico da *P. aeruginosa* CCBH4851

Esta seção aborda aspectos relevantes acerca do modelo, informações que influenciaram diretamente na reconstrução e as dificuldades impostas por todas as reconstruções metabólica abordadas neste trabalho (6,12,36).

Em relação às reconstruções metabólicas empreendidas pelos trabalhos realizados em 2008 e 2017 (6,7,12,36) este trabalho representa uma reconstrução inicial (*draft*), uma vez que necessitou delimitar um recorte nas etapas do protocolo *T&P* em função dos prazos impostos pela realização de uma pós-graduação.

Comparativamente, os demais trabalhos citados puderam executar mais etapas do protocolo *T&P* (7), tendo como consequência a obtenção de resultados mais abrangentes do que os aqui apresentados. O exíguo tempo de trabalho impôs limites à implementação de todas as etapas do protocolo *T&P*, obrigando a formular uma adaptação do mesmo, por essa razão, pode-se afirmar que a elaboração da metodologia empregada neste trabalho foi fundamental para o sucesso da reconstrução do modelo da *P. aeruginosa* CCBH4851. Atualmente, o modelo encontra-se em fase de curadoria, mas através da aplicação do método atingiu-se uma versão que simula a geração de biomassa maior que zero, atendendo aos objetivos deste trabalho. Isto só foi possível porque toda informação contida no modelo representa fortemente as relações Gene-Proteína-Reação caracterizadas pelo genoma.

5.2.1 Sobre a escolha do Modelo de Referência mPAO1_v2017

A modelagem metabólica é área emergente, embora já possua muitas informações disponíveis. Entretanto, nem sempre tais informações são de fácil análise e validação (7,15,28). Os modelos disponíveis para serem utilizados como referências não seguem um padrão de nomenclaturas para suas reações e metabólitos (6,12,36). O modelo publicado em 2008 (12) e os de 2017 (6) estão disponíveis no mesmo formato, porém em diferentes versões: o SBML level 3 (57) e SBML level 2 (58), respectivamente. Há ainda um modelo publicado em 2015 (36), disponível apenas em formato proprietário do MATLAB (31).

O modelo mPAO1_2017 apresenta algumas características interessantes, dentre elas destacam-se a escolha da cepa, o número de informações geradas comparativamente a outros modelos e a validação de seus dados pela comunidade científica.

A cepa PAO1 é a mais amplamente utilizada pela comunidade científica quando se trata de estudos de modelagem e reconstrução metabólica da *P. aeruginosa* (6). É não apenas a cepa que abrange mais versões de modelos metabólicos disponíveis em bases de dados científicos, como também aquela cujos modelos ensejaram a publicação de análises posteriores. Por esses motivos, considera-se que a *Pseudomonas aeruginosa* PAO1 possui modelos sólidos (curados e validados) (6,12,36). Assim, ainda que não exista um padrão de representação dos modelos e formato de arquivos os dados disponíveis são

considerados confiáveis, o que facilita o reaproveitamento para outras reconstruções de redes metabólicas (9).

Diante desse cenário e dificuldades apresentadas, como referência para este trabalho optou-se pela adoção do modelo mais atual dessa cepa, publicado em 2017 por Bartell et al (6). Além disso, acresce-se o fato de que a versão publicada em 2017 é a que tem a maior quantidade de informações mapeadas em um modelo de *Pseudomonas aeruginosa* PAO1 (6), com o maior número de genes, proteínas, reações, e *EC Numbers* mapeados. As diferenças podem ser visualizadas na Figura 47.

Organism	PAO1		
<i>Genome size</i>	6.3 Mbp		
<i>GC content</i>	66.60%		
<i>CDS</i>	5716		
Model	iMO1056	iMO1086	mPAO1
<i>Genes</i>	1056	1086	1146
<i>Unique EC identifiers</i>	502	NA	674
<i>Metabolites</i>	760	1021	1284
Cytoplasmic	760	916	1083
Extracellular	NA	105	201
<i>Reactions</i>	883	1031	1493
Metabolic	750	898	1083
Transport	133	133	238
Exchange	NA	105	172
<i>Gene–protein–reaction associations</i>			
Single gene			800
Multi-gene	839	839	471
None	44	47	49

Figura 47 – Dados comparativos das diferentes reconstruções da *P. aeruginosa* PAO1 (iMO1056/2008, iMO1086/2015 e mPAO1/2017) retirado de (6).

5.2.2 Reações de Biomassa (pseudo-reações)

A reação de biomassa é a principal reação de um modelo metabólico (seção 1.2) (15). As reações de biomassa nos modelos revisados foram inferidas a partir da reformulação de reações presentes no modelo da reconstrução metabólica da *E. coli* (16). A criação de uma equação de biomassa para um organismo demanda muitos experimentos biológicos e tempo, sendo por isso muito custosa financeiramente (19,22). A formulação da pseudo-reação de biomassa depende do conhecimento da

composição celular e dos requisitos básicos de energia para que a célula consiga gerar o conteúdo da biomassa a partir dos seus metabólitos precursores (22).

Em função das dificuldades impostas na elaboração e formulação de pseudo-reação de biomassa própria para a *P. aeruginosa* CCBH4851, fez-se necessário buscar uma estratégia para a inclusão de uma reação no modelo. Devido ao alto grau de homologia entre as cepas, considerou-se que seria possível seguir o mesmo paradigma utilizados nas demais reconstruções (6,12,36). Assim, para a reconstrução da *P. aeruginosa* CCBH4851 (mCCBH4851_v2) importou-se uma das reações disponíveis no modelo mPAO1_v2017 (6). Para a escolha da reação, executou-se a análise *FBA* testando cada uma das pseudo-reações disponíveis no modelo publicado em Bartell et al (6). Somente a reação denominada “R_PAO1_Biomass” funcionou. Ao integrá-la ao mapa metabólico da *P. aeruginosa* CCBH4851 e reconstruir rede, a análise de balanço de fluxo do modelo mCCBH4851_v2 apresentou o fluxo de biomassa maior que zero.

O fluxo de biomassa apontado pelo modelo mCCBH4851_v2 não representa uma realidade obtida em experimentos *in vitro* da bactéria. Entretanto, o modelo resultante atende ao objetivo inicial do projeto. Uma reconstrução é considerada suficiente quando minimamente é capaz de representar as vias metabólicas que possibilitam a geração de biomassa positiva (15). Ao avaliar as publicações da própria *P. aeruginosa* PAO1, observa-se que a taxa de biomassa gerada é variável (6,12,36). O CCBH4851_v2 gera uma taxa acima da versão do modelo 2008. Entretanto, em relação ao modelo aprimorado da versão 2017 (6), houve uma redução do valor. A expectativa é que no decorrer da curadoria da rede, o modelo seja refinado ao ponto de gerar uma taxa de biomassa compatível aos experimentos *in vitro* (59).

5.2.3 Exemplo de análise com o Modelo mCCBH4851_v2

Uma análise de uso do modelo reconstruído (mCCBH4851_v2), foi realizada através do sistema *web FindTargetsWeb* (9). O sistema parte de um modelo metabólico que tenha a taxa de crescimento acima de zero e efetua uma busca de possíveis alvos terapêuticos. Como se pode confirmar na Figura 48, o modelo submetido é aceito como válido pelo sistema, o que comprova que sua geração de biomassa é positiva.

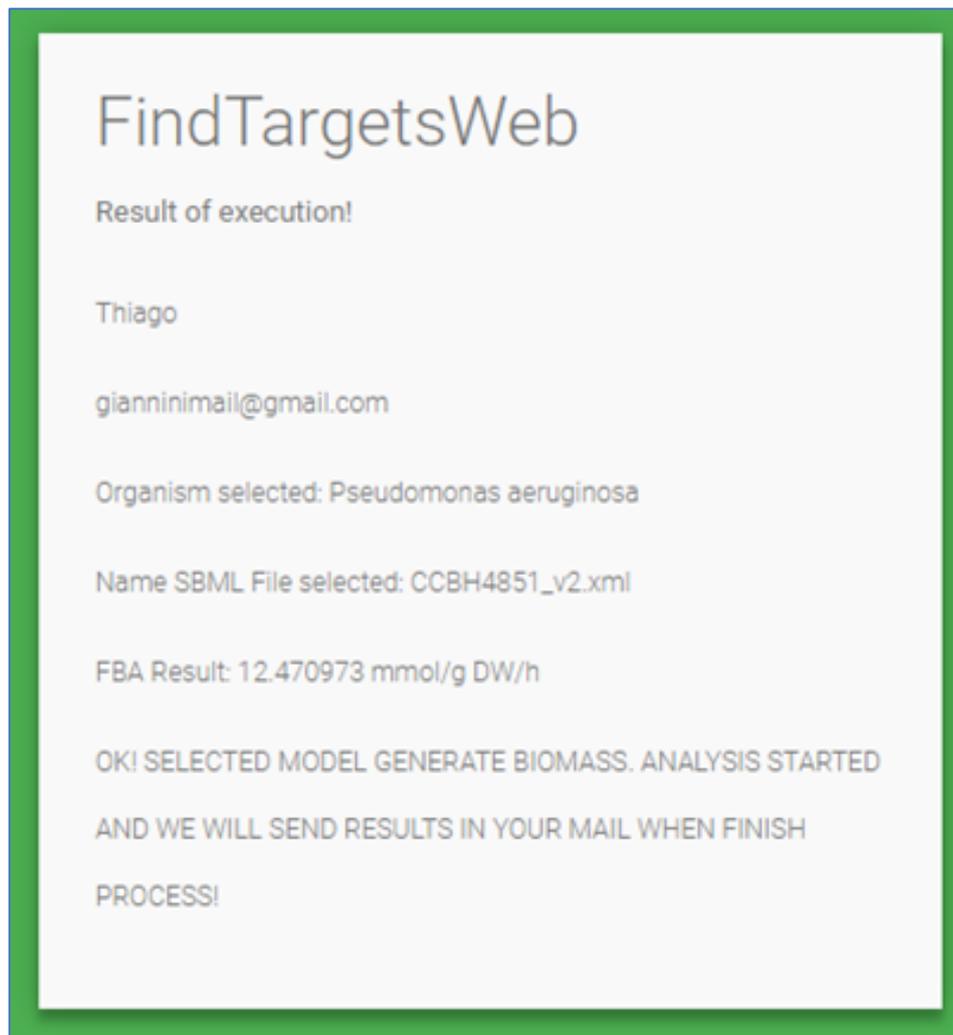


Figura 48 - Análise executada no *FindTargetsWeb* a partir do modelo reconstruído mCCBH4851_v2 (9)

Após se concluir a execução da análise do sistema, é enviado um e-mail com as indicações de alvos terapêuticos listadas em uma planilha em anexo. O e-mail exprime o resultado em formato de relatório, como se pode observar na Figura 49. A saída mais detalhada do resultado pode ser vista no apêndice E.

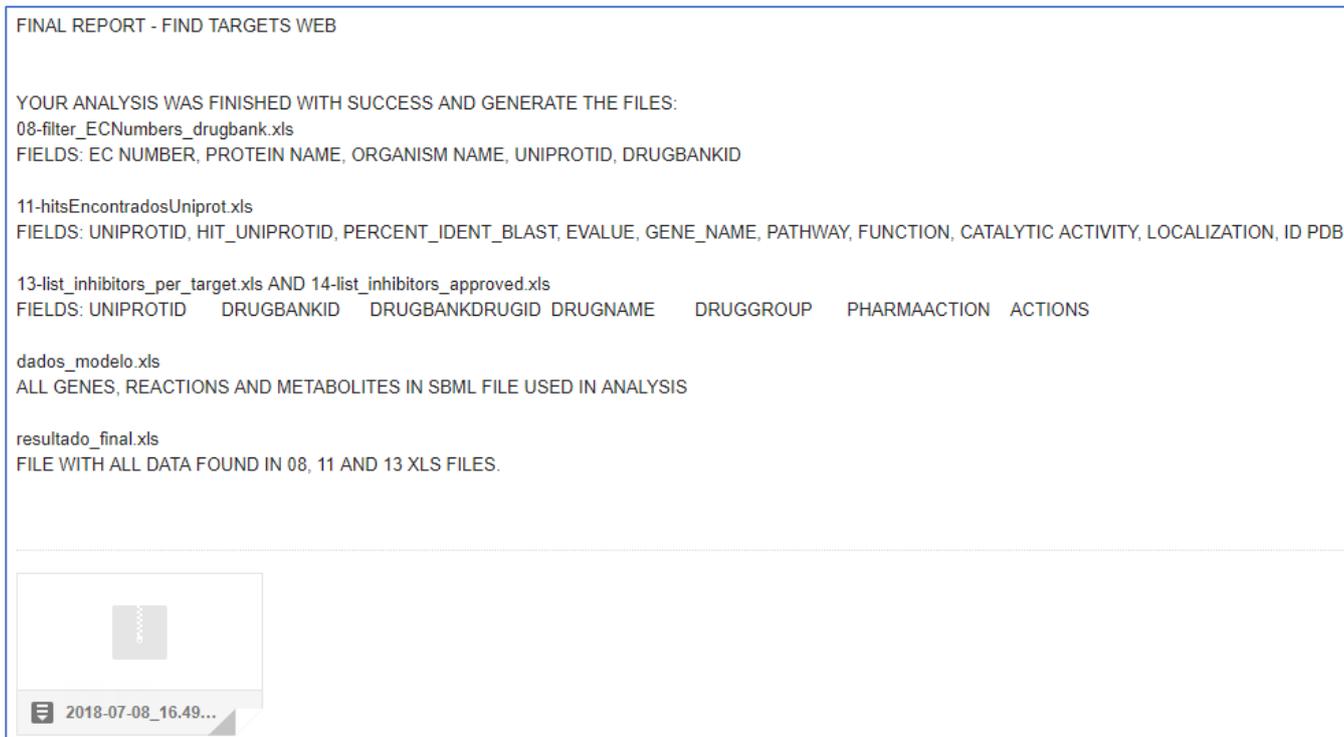


Figura 49 - E-mail de resultado do processamento da análise do FindTargetsWeb através do modelo de rede metabólica mCCBH4851_v2

5.2.4 Validação de reações inferidas e dificuldades encontradas

A reconstrução de redes metabólicas é uma área de conhecimento multidisciplinar. Este fato tanto impõe restrições (15) que dificultaram a apuração de respostas para refinamento do modelo mCCBH4851_v2, quanto expande as possibilidades de pesquisa. Atualmente, empenha-se para confirmar a existência das quatro reações (Tabela 4) inferidas manualmente no modelo. Tais reações foram: “R_rxn02476”, “R_rxn00695”, “R_rxn05468” e “R_ATPM”. A validação destas reações exige que se as considere em separado, avaliando-as caso a caso.

A primeira reação analisada foi a “R_ATPM”, descrita como reação de demanda energética (pseudo-reação). Para validar essa reação foi necessário proceder a uma revisão de literatura e busca de informações nos modelos reconstruídos para a cepa de *Pseudomonas aeruginosa* PAO1 (6,12,36). Os dados da reação constam na Tabela 9.

Modelo	Equação	Subsistema	ID no modelo
mPAO1_v2008	$[c] : \text{atp} + \text{h}_2\text{o} \rightarrow \text{adp} + \text{h} + \text{pi}$	Demanda de energia externa	ATPM
mPAO1_v2015	$[c] : \text{atp} + \text{h}_2\text{o} \rightarrow \text{adp} + \text{h} +$	Demanda de energia	ATPM

	pi	externa	
mPAO1_v2017	H2O + ATP -> H+ + ADP + Phosphate	-	ATPM
PA14	H2O + ATP -> H+ + ADP + Phosphate	-	ATPM

Tabela 9 - Diferentes modelos que contém a reação de demanda energética "R_ATPM"

Todos os quatro modelos analisados têm a reação catalogada, mas somente as versões publicadas em 2008 e 2015 contêm dados sobre sua existência. Nos mapas metabólicos, a reação foi definida como demanda externa de ATP (6,12, 36). Trata-se, na verdade, de uma pseudo-reação criada para suprir a demanda de ATP. Nos modelos publicados em 2017 (mPAO1_v2017 e PA14) não há referência a bases de dados externas. Entretanto, essa pseudo-reação encontra-se catalogada em modelos mais antigos sob o *EC number* 3.6.1.15 (12,36). Então, por comprovação de literatura (25), a reação "R_ATPM" foi validada para o modelo reconstruído (mCCBH4851_v2) neste trabalho. Essa discussão também pode ser vista no sistema CurSystem, inserida no desafio correspondente a esta reação.

Outra reação analisada de maneira semelhante foi a reação denominada "R_rxn05468" (O₂ transport *in via diffusion*). Esta reação permite o transporte de oxigênio do meio externo para dentro da célula. Todos os modelos revisados (6,12,36) possuem a reação. A revisão de literatura justifica sua inclusão em modelos que representem um organismo que necessita de oxigênio em seu metabolismo.

A reação cujo ID é "R_rxn00695" (glucose-1-phosphate adenylyltransferase) ainda se encontra em processo de validação. Sua permanência no modelo também deve ser justificada em função da literatura e dos dados existentes em modelos publicados (25). Por análise de ortologia, observa-se que possivelmente há um gene associado à esta reação no genoma da *P. aeruginosa* CCBH4851. O gene tem o ID NP_250713.1 e foi anotado com o *EC number* número 2.7.7.9. A enzima UTP-glucose-1-phosphate uridylyltransferase, expressa por este gene, contém a diferença de uma molécula de UDP em vez de ATP na equação da reação. A Figura 50 mostra um recorte da anotação do genoma com os dados completos do gene.

```

CDS      complement(3658798..3659637)
         /translation="MIKKCLFPAAGYGTRFLPATKAMPKEMLPVVNKPLIQYAVEEAL
         EAGLSEIGIVTGRGKRSLEDHFDISYELEHQIRNTDKEKYLVGIRRLIDECTFAYTRQ
         VEMKGLGHAILTGRPLIGDEPFAVVLADDLCLNLEGDSVLKQMVKLYNQFRCSIVAIQ
         EVPPEETNKYGVVIAGEMIRDDIFRVNTMVEKPKPEEAPSNLAIIGRYILTPDIFDLIE
         QTEPGKGGEIQITDALMKQAQDGCVLAYKFKGKRFDGCSAEGYIEATNFCYENLYKTG
         KAH"
         /inference="COORDINATES: similar to AA
         sequence:RefSeq:NP_250713.1"
         /gene="galU"
         /codon_start=1
         /transl_table=11
         /product="UTP-glucose-1-phosphate uridylyltransferase"
         /locus_tag="PA4851_16835"
         /EC_number="2.7.7.9"
gene     complement(3659685..3661046)
         /locus_tag="PA4851_16840"

```

Figura 50 - Dados completos do gene com ID NP_250713.1 da última versão do genoma da *P. aeruginosa* CCBH4851

A busca por ortólogos mostrou um alto grau de similaridade com o gene ARU36339.1 (Caur_2523) do genoma da bactéria *Chloroflexus aurantiacus* (KEGG ID: T00639) (25). O resultado da análise pode ser visto na Figura 51.

```

#####
# List of CCBH genes appearing as orthologue pairs of KEGG genes promoting PA01 gapfilling rxn #
# GLGC (EC number 2.7.7.27) (reciprocal best hits with evaluate<1e-05) #
# -----#
# [iPAE1146] R_rxn00695 (glucose-1-phosphate adenyltransferase) #
# H+ + ATP + Glucose-1-phosphate <=> Ppi + ADPglucose #
# -----#
# [iMO1063] GLGC (glucose-1-phosphate adenyltransferase) #
# atp + glp + h --> adpglc + ppi #
# ** Blasted for the ecoli version (NT02EC4565=tigr locus name) and didn't find it in system ** #
# ***** Possibly NT03PA2256, GalU gene ***** #
#####

=====
(1) Gene ARU33500.1
===== gbk info =====
/inference="COORDINATES: similar to AA sequence:RefSeq:NP_250713.1"
/note="Derived by automated computational analysis using gene prediction method: Protein Homology."
/product="UTP--glucose-1-phosphate uridylyltransferase"
/protein_id="ARU33500.1"
/db_xref="GI:1199091979"
/translation="MIKKCLFPAAGYGTRFLPATKAMPKEMLPVVNKPLIQYAVEEAL
EAGLSEIGIVTGRGKRSLEDHFDISYELEHQIRNTDKEKYLVGIRRLIDECTFAYTRQ
VEMKGLGHAILTGRPLIGDEPFAVVLADDLCLNLEGDSVLKQMVKLYNQFRCSIVAIQ
EVPPEETNKYGVVIAGEMIRDDIFRVNTMVEKPKPEEAPSNLAIIGRYILTPDIFDLIE
QTEPGKGGEIQITDALMKQAQDGCVLAYKFKGKRFDGCSAEGYIEATNFCYENLYKTG
KAH"

===== gbk info =====
* Closest KEGG orthologue: (cau:Caur_2523) Eval=6e-14
  To see the gene: https://www.genome.jp/dbget-bin/www\_bget?cau:Caur\_2523
* CCBH orthologues:
  (gnl|ccbh4851|ARU35951.1, eval=3e-22) (gnl|ccbh4851|ARU36339.1, eval=1e-08) (gnl|ccbh4851|ARU36966.1,

```

Figura 51 - Resultado da análise de Ortologia via BLASTp para o gene NP_250713.1 do genoma da *P. aeruginosa* CCBH4851

Então, realizou-se uma busca no KEGG (https://www.genome.jp/dbget-bin/www_bget?cau:Caur_2523) (25) pelo gene denominado “Caur_2523” para verificar qual a proteína codificada. O banco de dados apontou para a enzima glucose-1-phosphate adenylyltransferase com o EC *number* número 2.7.7.27, a mesma identificação da reação inferida manualmente na segunda versão do modelo mCCBH4851_v2. Estas análises apontam para informações convergentes, indicando que possivelmente o gene NP_250713.1 representa a reação R_rxn00695. Porém, ainda que estas conclusões sejam plausíveis, não se pode afirmá-las de maneira contundente sem que haja validação de um especialista em bioquímica.

A reação com ID “R_rxn02476”, denominada 3-phosphoshikimate 1-carboxyvinyltransferase (Biossíntese de fenilalanina, tirosina e triptofano) está no *pipeline* de curadoria do modelo, com sua análise prevista para iniciar em breve. Os processos de curadoria descritos acima expõem como a resolução dos problemas (*gaps*) do modelo metabólico demanda tempo e rigor analítico. É um processo laborioso, que pode despende alguns anos para solucionar todas as inconsistências de dados. Apesar disso, as soluções para o modelo estão progredindo.

6 CONCLUSÕES

A interdisciplinaridade inerente à execução deste trabalho elevou o grau de dificuldade de cobrir todos aspectos envolvidos na modelagem de redes metabólicas em escala genômica. Este tipo de pesquisa é uma proposta relativamente nova para o estudo sistêmico celular (15). O nível de dificuldade pode ser depreendido pela necessidade de utilização de diferentes ferramentas computacionais, pela falta de padronização das bases de dados, das diversas tecnologias envolvidas e até das maneiras de representar as informações acerca da área. Contudo, julga-se que o trabalho produziu conhecimento relevante no âmbito da Biologia Computacional, produzindo resultados que podem contribuir para o desenvolvimento da área. Outro aspecto importante a ser destacado é o potencial do trabalho a ser explorado futuramente, com possibilidade de evolução do método e das ferramentas desenvolvidas, e, principalmente, do modelo de rede metabólica gerado.

O método de reconstrução proporcionou o acompanhamento e aprendizado consistente de todo o processo de reconstrução em escala genômica. O trabalho produziu um modelo de rede metabólica inexistente para um microrganismo e também a metodologia que subsidiou a reconstrução do modelo. Esta metodologia envolve dois sistemas *web* completos e podem ser considerados também como produtos para a comunidade científica. Os sistemas estão em pleno funcionamento e pesquisadores que os utilizam estão enviando sugestões de melhorias a serem avaliadas e implantadas posteriormente. Algumas contribuições de melhorias recebidas podem ser citadas, tais como: 1) a expansão do NETGen para busca em outras bases de dados, 2) a execução de reconstruções com adoção de mais de um modelo de referência e 3) permissão para recebimento de mais de um formato de arquivo de entrada e saída. Para o CurSystem, há também melhorias previstas, tais como a apresentação de alterações da rede através de representações gráficas, o armazenamento do modelo em banco de dados não relacional, como o *NoSQL*, entre outras melhorias de interface já sinalizadas.

O modelo resultante confirmou que é possível reconstruir uma rede metabólica a partir de um subconjunto de dados de outros modelos metabólicos. Ou seja, a partir da metodologia realizada e de um modelo de referência, foi possível executar uma reconstrução em escala genômica. No processo de curadoria, há expectativa de que o modelo seja incrementado e cada vez mais aperfeiçoado. Cabe

ressaltar que o modelo mCCBH4851_v2 foi considerado final do ponto de vista da execução deste trabalho. Porém, na verdade, representa apenas mais um passo dado em um longo caminho a ser percorrido a fim de que se obtenham reconstruções cada vez mais acuradas em relação ao organismo vivo.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- (1) Silveira M, Albano R, Asensi M, & Assef APC. The draft genome sequence of multidrug-resistant *Pseudomonas aeruginosa* strain CCBH4851, a nosocomial isolate belonging to clone SP (ST277) that is prevalent in Brazil. *Memórias do Instituto Oswaldo Cruz*. 2014; 109(8): 1086-1087.
- (2) Anvisa. Segurança do Paciente e Qualidade em Serviços de Saúde. *Bol Inf*. 2014;1:1–12.
- (3) OMS. La OMS publica la lista de las bacterias para las que se necesitan urgentemente nuevos antibióticos [Internet]. 27 de febrero. 2017 [cited 2018 Mar 1]. p.1. Available from: <http://www.who.int/mediacentre/news/releases/2017/bacteria-antibiotics-needed/es/>
- (4) Vallet-Gely I, Boccard F. Chromosomal Organization and Segregation in *Pseudomonas aeruginosa*. *PLoS Genet*. 2013; 9(5): e1003492.
- (5) Verli H (org). *Bioinformática: da Biologia à Flexibilidade Molecular*. 1.Ed. São Paulo, Brasil. SBBq; 2014.
- (6) Bartell JA, Blazier AS, Yen P, Thogersen JC, Jelsbak L, Goldberg JB, et al. Reconstruction of the metabolic network of *Pseudomonas aeruginosa* to interrogate virulence factor synthesis. *Nature Communications*. 2017; 8. doi: 10.1038 / ncomms14631.
- (7) Thiele I, Palsson B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*. 2010; 5(1): 93-121.
- (8) Modelagem computacional da *P. aeruginosa* CCBH 4851. [Homepage]. [acesso em 2017 ago. 05] Disponível em: <http://pseudomonas.procc.fiocruz.br/>.

- (9) Merigueti TC. Identificação de Alvos Terapêuticos para a Bactéria Multirresistente *P. aeruginosa* CCBH4851 através da Análise de Redes Metabólicas. Rio de Janeiro: Instituto Oswaldo Cruz / Fiocruz; 2018.
- (10) Trabulsi L, Alterthum F. Microbiologia. 4.ed. Atheneu Editora; 2008. 780 p.
- (11) Teixeira JOG. Infecção da Corrente Sanguínea causada por *Pseudomonas aeruginosa* Resistente aos Carbapenêmicos: Fatores Associados a Mortalidade e Influência da Terapia combinada com Polimixina B e Imipenem. São Paulo: Escola Paulista de Medicina; 2011.
- (12) Oberhardt MA, Puchałka J, Fryer KE, Martins dos Santos VA, Papin JA. Genome-scale metabolic network analysis of the opportunistic pathogen *Pseudomonas aeruginosa* PAO1. *J Bacteriol.* 2008;190(8) : 2790–803.
- (13) Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet.* 2014;15:107 EP. Review Article.
- (14) Ebrahim A, Lerman JA, Palsson B, Hyduke DR. COBRApy: COntstraints-Based Reconstruction and Analysis for Python. *BMC Systems Biology.* 2013; 7:74. <https://doi.org/10.1186/1752-0509-7-74>.
- (15) Menezes MA. Metabolic Models: From DNA to Physiology (and Back). In: Alves Barbosa da Silva F., Carels N., Paes Silva Junior F. (eds) Theoretical and Applied Aspects of Systems Biology. Computational Biology, vol 27. Springer, Cham; 2018.
- (16) Sousa FL, Hordijk W, Steel M, Martin WF. Autocatalytic sets in *E. coli* metabolism. *J Syst Chem.* 2015;6(1):4. 9[PII].
- (17) Kohanski MA, Dwyer DJ, Collins JJ. How antibiotics kill bacteria: from targets to networks. *Nat Rev Microbiol.* 2010;8(6):423–35. 20440275[pmid].

- (18) Niedenfhr S, Wiechert W, Katharina NH. How to measure metabolic fluxes: a taxonomic guide for 13c fluxomics. *Curr Opin Biotechnol.* 2015;34(Supplement C):82–90. Systems biology Nanobiotechnology.
- (19) Joyce AR, Palsson B. Predicting gene essentiality using genome-scale in silico models. In: Osterman AL, Gerdes SY, editors. *Microbial gene essentiality: protocols and bioinformatics.* Totowa: Humana Press; 2008. p. 433–57.
- (20) Walsh JR, Schaeffer ML, Zhang P, Rhee SY, Dickerson JA, Sen TZ. The quality of metabolic pathway resources depends on initial enzymatic function assignments: a case for maize. *BMC Syst Biol.* 2016;10:129. 369[PII].
- (21) Smith CA, Neidhardt FC, Ingraham JL, Schaechter M. *Physiology of the bacterial cell: a molecular approach.* Sunderland: Sinauer Associates; 1990. p. 507; 43:95. ISBN: 0878936084; 2010;20:124–5.
- (22) Mendoza SN, Can PM, Contreras N, Ribbeck M, Agosn E. Genome-scale reconstruction of the metabolic network in *Oenococcus oeni* to assess wine malolactic fermentation. *Front Microbiol.* 2017;8:534. 28424673[pmid].
- (23) Silva WMC. Método para reconstrução in silico de redes metabólicas de fungos: um estudo de caso para o *Paracoccidioides lutzii*. 2014. xii, 70 f., il. Dissertação (Mestrado em Informática) – Universidade de Brasília, Brasília.
- (24) TORTORA GJ, FUNKE BR, CASE CL. *Microbiologia.* 10.ed. Porto Alegre, Brasil: Artmed Editora SA; 2012.
- (25) KEGG: Kyoto Encyclopedia of Genes and Genomes. [Homepage]. [acesso em 2017 ago. 05] Disponível em: <http://www.kegg.jp/>.
- (26) Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: Kyoto encyclopedia of genes and genomes. Vol. 27, *Nucleic Acids Research.* 1999. p. 29–34.

- (27) IUBMB. International Union of Biochemistry and Molecular Biology. [Homepage]. [acesso em 2018 jan. 15] Disponível em: <http://www.iubmb.org/>.
- (28) Karr JR, et al. A Whole-Cell Computational Model Predicts Phenotype From Genotype. *CELL*. 2012; 150(2): 389-401.
- (29) Covert M. Fundamentals of Systems Biology: From Synthetic Circuits to Whole-cell Models. Press C, editor. ISBN: 9781449369330; 2015. 367 p.
- (30) Reed JL, Patel TR, Chen KH, et al. Systems approach to refining genome annotation. *Proceedings of the National Academy of Sciences of the United States of America*. 2006;103(46):17480-17484. doi:10.1073/pnas.0603364103.
- (31) Hyduke D, Hyduke D, Schellenberger J, Que R, Fleming R, Thiele I, et al. COBRA Toolbox 2.0. *Protoc Exch*. 2011;
- (32) Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature reviews Microbiology*. 2012;10(4):291-305. doi:10.1038/nrmicro2737.
- (33) Heirendt L, Thiele I, Fleming RMT. DistributedFBA.jl: High-level, high-performance flux balance analysis in Julia. *Bioinformatics* 2017. 2016; btw838. doi - 10.1093/bioinformatics/btw838.
- (34) Orth JD, Thiele I, Palsson BØ. What is flux balance analysis?. *Nature biotechnology*. 2010; 28(3), 245-248.
- (35) Feist AM, Palsson BO. The Biomass Objective Function. *Current opinion in microbiology*. 2010;13(3):344-349. doi:10.1016/j.mib.2010.03.003.
- (36) Vital-Lopez FG, Reifman J, Wallqvist A. Biofilm Formation Mechanisms of *Pseudomonas aeruginosa* Predicted via Genome-Scale Kinetic Models of Bacterial Metabolism. *PLoS Comput Biol*. 2015; 11(10): e1004452. doi:10.1371 / journal.pcbi.1004452

- (37) Hucka M, Hucka M, Bergmann F, Hoops S, Keating S, Sahle S, et al. The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *bioinformatics*. 2003; 19(4): 524-31.
- (38) Pseudomonas GenomeDB (PseudoCAP). [Homepage]. [acesso em 2017 ago. 05] Disponível em: <http://www.pseudomonas.com>.
- (39) Qian K, Allen R, Gan M, Brown R. *Desenvolvimento Web Java*. 1.ed. Rio de Janeiro, Brasil: LTC; 2010.
- (40) Hucka M, Hucka M, Bergmann F, Hoops S, Keating S, Sahle S, e outros. Linguagem de marcação de biologia de sistemas (SBML): Especificação de linguagem para o nível 3 da versão 1 do núcleo. *Nat Preced*. 2010;
- (41) Winsor GL, Lam DKW, Fleming L, et al. Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. *Nucleic Acids Research*. 2011; 39(Database issue):D596-D600. doi:10.1093/nar/gkq869.
- (42) Gonçalves E. *Desenvolvendo Aplicações WEB com JSP, Servlets, Java Server Faces, Hibernate, EJB 3 Persistence e Ajax*. 1.ed. Ciência Moderna; 2007.
- (43) Ahmed KZ, Umrysh CE. *Developing Enterprise Java Applications with J2EE and UML*. 1.ed. Addison-Wesley; 2002.
- (44) Deitel H, Deitel P. *Java: Como Programar*. 6.ed. Prentice Hall; 2005.
- (45) Prlić A, Yates A, Bliven SE, Rose PW, Jacobsen J, Troshin PV, et al. BioJava: an open-source framework for bioinformatics in 2012. *Bioinformatics*. 2012; 28 : 2693–2695.
- (46) Hedley J. *jsoup: Java HTML Parser. Cookbook*; 2017; [Homepage]. [acesso em 2018 fev. 05] Disponível em: <https://jsoup.org/cookbook/>.

- (47) Dräger A, Rodriguez N, Dumousseau M, Dörr A, Wrzodek C, Novère NL, et al. JSBML: a flexible Java library for working with SBML. *Bioinformatics*; 2011; 27 : 2167–2168.
- (48) Weiss MA. *Data Structures and Algorithm Analysis in Java*, 3.ed. Florida, US: Pearson; 2012.
- (49) BRENDA enzyme database. [Homepage]. [acesso em 2017 ago. 05] Disponível em: <http://www.brenda-enzymes.org/index.php>.
- (50) Copeland WB, Bartley BA, Chandran D, Galdzicki M, Kim KH, Sleight SC, et al. Computational tools for metabolic engineering. *Metab Eng*; 2012; 14(3) : 270-80.
- (51) Qian K. et al. *Desenvolvimento Web Java*. 1.ed. São Paulo, Brasil: LTC; 2010.
- (52) McEntyre J, Ostell J. *The NCBI Handbook*. Bethesda (MD): National Center for Biotechnology Information (US); 2002; [Homepage]. [acesso em 2018 jan. 15] Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK21101/>.
- (53) Harold ER. *JavaMail API: Sending and Receiving Email with Java*. 1.ed. USA: O'Reilly Media; 2013.
- (54) Barbosa SDJ, Silva BS. *Interação Humano-Computador*. 1.ed. Rio de Janeiro, Brasil: Elsevier; 2010.
- (55) Tahaghoghi S, Williams H. *Aprendendo MySQL*. 1.ed. Rio de Janeiro, Brasil. Alta Books; 2007.
- (56) MySQL AB. *MySQL 5.6 Reference Manual*; 2018. 1-20.
- (57) Hucka M, Bergmann FT, Hoops S, et al. The Systems Biology Markup Language (SBML): Language Specification for Level 3 Version 1 Core. *Journal of integrative bioinformatics*. 2015;12(2):266. doi:10.2390/biecoll-jib-2015-266.

(58) Finney A, Hucka M. Systems biology markup language: Level 2 and beyond. *Biochemical Society Transactions*. 2003; 31 (6) 1472-1473; DOI: 10.1042/bst0311472.

(59) LaBauve AE, Wargo MJ. Growth and laboratory maintenance of *Pseudomonas aeruginosa*. *Curr Protoc Microbiol*. 2012;(SUPPL.25).

(60) OMS. OMS. In: El primer informe mundial de la OMS sobre la resistencia a los antibióticos pone de manifiesto una grave amenaza para la salud pública em todo el mundo. 2014. p. 5.

(61) Kazakiewicz D, Karr JR, Langner KM, Plewczynski D. A combined systems and structural modeling approach repositions antibiotics for *Mycoplasma genitalium*. *Comput Biol Chem pii*. 2015; S1476-9271(15)30089-X, 51 Part B: 91-97.

(62) Ma H, Zeng AP. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics*; 2003, 19(2):270–277.

(63) Feofiloff P, Kohayakawa Y, Wakabayashi Y. Uma Introdução Sucinta à Teoria dos Grafos. 1.ed. São Paulo, Brasil: Sociedade Brasileira de Matemática; 2004.

(64) Palsson B. METABOLIC SYSTEMS BIOLOGY. *FEBS letters*. 2009;583(24):3900-3904. doi:10.1016/j.febslet.2009.09.031.

(65) Nelson DL, Cox MC. Principles of biochemistry. W. H. Freeman, 4 edition. Viii; 2005, 1, 6, 50, 51.

(66) Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, Walter P. Molecular biology of the cell. 500 Tips. New York: Garland Science; 2014.

(66) Bitbucket Cloud documentation. [Homepage]. [acesso em 2018 maio. 05] Disponível em: <https://confluence.atlassian.com/bitbucket/bitbucket-cloud-documentation-221448814.html>.

(67) What is Git. [Homepage]. [acesso em 2018 maio. 05] Disponível em:
<https://www.atlassian.com/git/tutorials/what-is-git>.

APÊNDICES E/OU ANEXOS

Nesta seção disponibilizamos matérias adicionais e informações consideradas importantes, e que complementam este trabalho. Na Apêndice A expomos as equações de biomassas do modelo de referência e o do novo modelo criado. No Apêndice B disponibilizamos os links para download do código fonte dos Sistemas WEB desenvolvidos, pois são mais de 9.000 linhas de códigos para cada um. Seria inviável colocar o código na íntegra nesta seção. No Apêndice C, colocamos a referência para download do manual do CurSystem, pois foi feito para usuários de outros projetos utilizarem o sistema. E por último, o Apêndice D, na qual, incluímos algumas análises feitas com a segunda versão do modelo da rede gerado (mCCBH4851_v2).

7.1 Apêndice A – Equação de Biomassa

A reação de biomassa utilizada para a reconstrução da rede metabólica da *P. aeruginosa* CCBH4851 foi importada do modelo da *P. aeruginosa* PAO1 versão 2017 (iPAE1146). O modelo tinha quatro reações disponíveis, nomeados respectivamente de “R_PAO1_Biomass”, “R_PA_Biomass_v13ub”, “R_PA_Biomass_v13” e “R_PA_Biomass_v4”. A reação de biomassa escolhida foi a “P_PAO1_Biomass”.

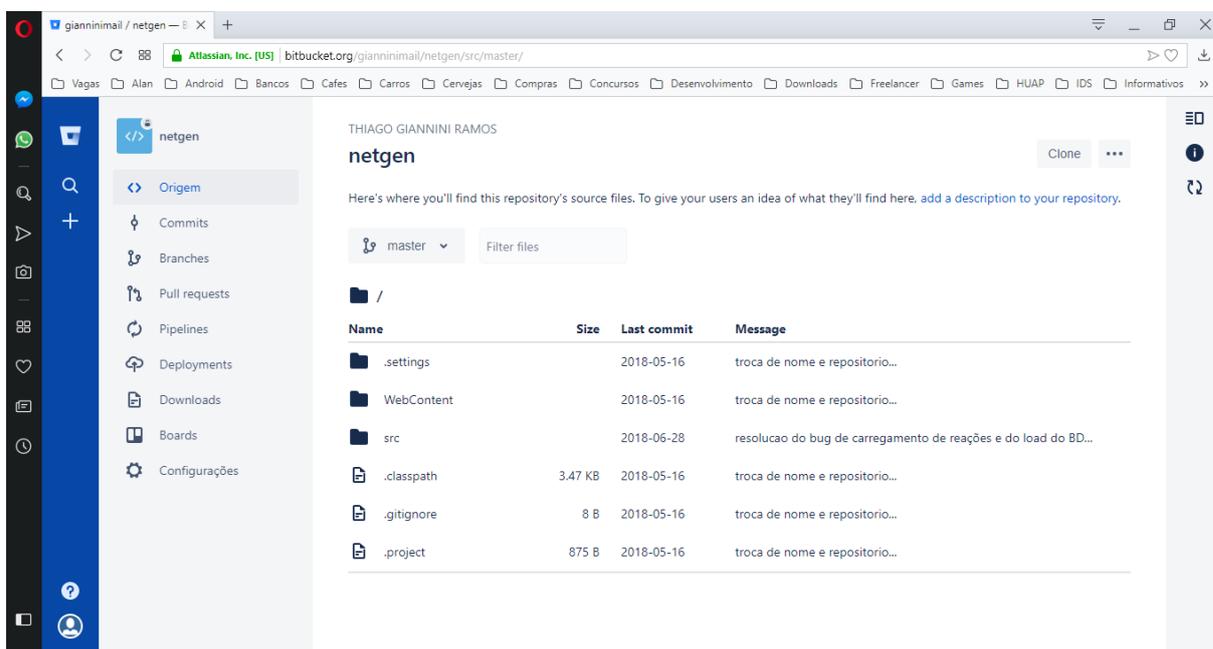
Equação da reação de biomassa utilizada como referência para reconstrução dos modelos mCCBH4851 e mCCBH4851_v2:

PAO1_Biomass: 0.0001 cJB00125_c + 0.0013602 cPY00124_c + 0.0030243 cPY00129_c + 0.00079222 cPY00132_c + 0.0015844 cPY00135_c + 0.00053681 cPY00138_c + 0.00097843 cPY00140_c + 59.81 cpd00001_c + 59.81 cpd00002_c + 0.10147 cpd00018_c + 0.22711 cpd00023_c + 0.0001 cpd00028_c + 0.31517 cpd00033_c + 0.43505 cpd00035_c + 0.10702 cpd00039_c + 0.19852 cpd00041_c + 0.20202 cpd00046_c + 0.28478 cpd00051_c + 0.15885 cpd00053_c + 0.20628 cpd00054_c + 0.075597 cpd00060_c + 0.055547 cpd00065_c + 0.13262 cpd00066_c + 0.094677 cpd00069_c + 0.037522 cpd00084_c + 0.099817 cpd00091_c + 0.46473 cpd00107_c + 0.56945 cpd00118_c + 0.080962 cpd00119_c + 0.0001 cpd00125_c + 0.19855 cpd00126_c + 0.18967 cpd00129_c + 0.098535 cpd00132_c + 2.5e-05 cpd00155_c + 0.25796 cpd00156_c + 0.15587 cpd00161_c + 0.03128 cpd00206_c + 0.016114 cpd00294_c + 0.03128 cpd00296_c + 0.016114 cpd00298_c + 0.15571 cpd00322_c + 0.0078591 cpd15428_c + 0.0041649 cpd15429_c + 0.00037815 cpd15430_c + 0.011889 cpd15431_c + 0.047824 cpd15531_c + 0.020287 cpd15532_c + 0.0020692 cpd15533_c + 0.068198 cpd15534_c + 0.012557 cpd15538_c + 0.0037982 cpd15539_c + 0.00042166 cpd15540_c + 0.011557 cpd15541_c + 0.053367 cpd15665_c + 0.025523 cpd17066_c --> 59.81 cpd00008_c + 59.81 cpd00009_c + 0.69664 cpd00012_c + 59.81 cpd00067_c

7.2 Apêndice B – Código fonte dos sistemas

Ao longo de todo o ciclo de desenvolvimento dos dois Sistemas WEB foi necessário trabalhar com a plataforma Bitbucket. A plataforma disponibiliza serviços de repositórios de projetos de software, bem como, um sistema de controle de versionamento de código. O Bitbucket suporta os sistemas Git e Mercurial, e no contexto deste trabalho, escolheu-se o Git. Tanto o NETGen, como o CurSystem estão disponíveis para download. Abaixo segue os links para clonagem através do sistema Git (66,67).

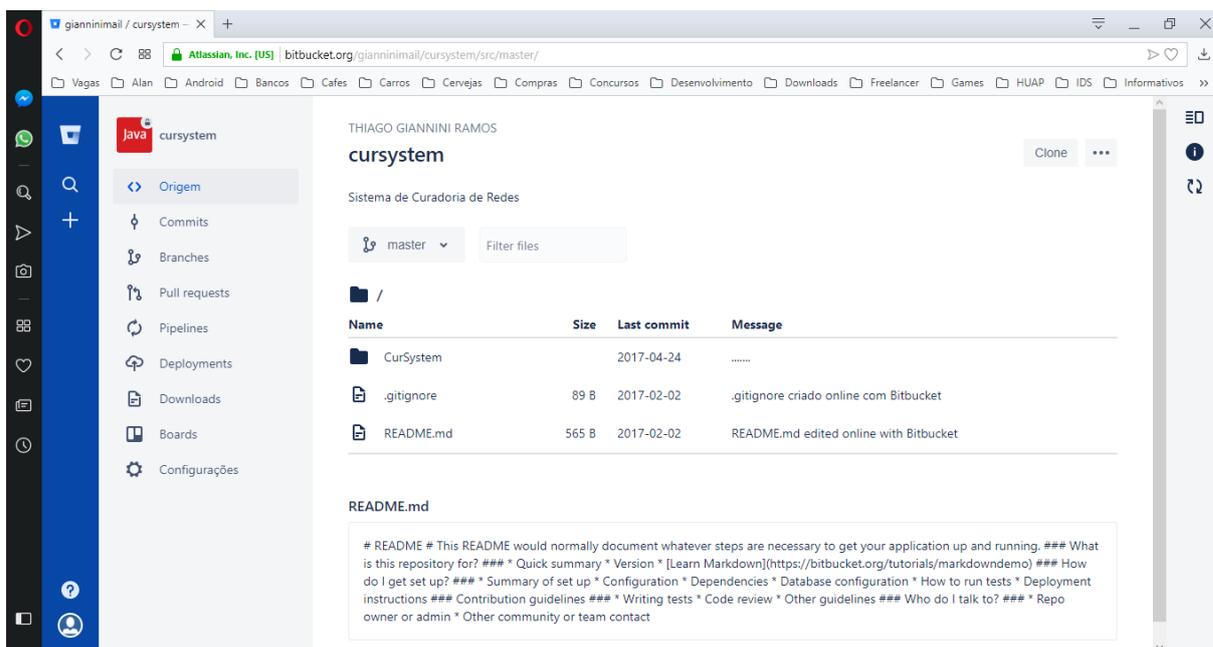
NETGen : git clone <https://gianninimail@bitbucket.org/gianninimail/netgen.git>.



The screenshot shows the Bitbucket web interface for the repository 'netgen' by Thiago Giannini Ramos. The repository is located at `bitbucket.org/gianninimail/netgen/src/master/`. The interface includes a sidebar with navigation options like 'Origem', 'Commits', 'Branches', 'Pull requests', 'Pipelines', 'Deployments', 'Downloads', 'Boards', and 'Configurações'. The main content area displays the repository's structure with a table of files and folders:

Name	Size	Last commit	Message
./			
./settings		2018-05-16	troca de nome e repositório...
WebContent		2018-05-16	troca de nome e repositório...
src		2018-06-28	resolucao do bug de carregamento de reações e do load do BD...
./classpath	3.47 KB	2018-05-16	troca de nome e repositório...
./gitignore	8 B	2018-05-16	troca de nome e repositório...
./project	875 B	2018-05-16	troca de nome e repositório...

CurSystem: git clone <https://gianninimail@bitbucket.org/gianninimail/cursystem.git>.



The screenshot shows the Bitbucket web interface for the repository 'cursystem' by Thiago Giannini Ramos. The repository is located at `bitbucket.org/gianninimail/cursystem/src/master/`. The interface includes a sidebar with navigation options like 'Origem', 'Commits', 'Branches', 'Pull requests', 'Pipelines', 'Deployments', 'Downloads', 'Boards', and 'Configurações'. The main content area displays the repository's structure with a table of files and folders:

Name	Size	Last commit	Message
./			
CurSystem		2017-04-24
./gitignore	89 B	2017-02-02	./gitignore criado online com Bitbucket
README.md	565 B	2017-02-02	README.md edited online with Bitbucket

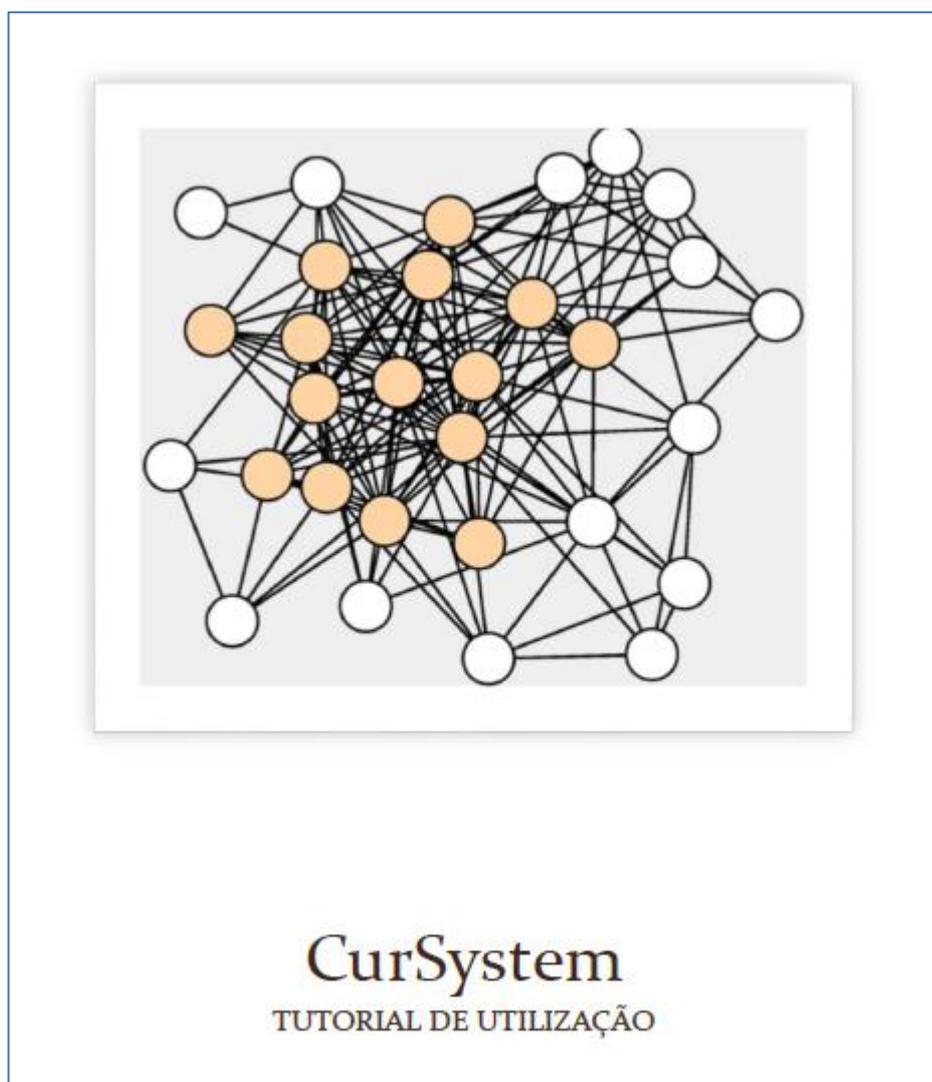
Below the table, the README.md content is displayed:

```
# README # This README would normally document whatever steps are necessary to get your application up and running. ## What is this repository for? ## * Quick summary * Version * [Learn Markdown](https://bitbucket.org/tutorials/markdowndemo) ## How do I get set up? ## * Summary of set up * Configuration * Dependencies * Database configuration * How to run tests * Deployment instructions ## Contribution guidelines ## * Writing tests * Code review * Other guidelines ## Who do I talk to? ## * Repo owner or admin * Other community or team contact
```

7.3 Apêndice C - Manual do CurSystem

No processo de implantação do sistema CurSystem, obteve-se a solicitação de outros projetos para utilização do Sistema de Curadoria Assistida. Foi necessário a elaboração de um material didático para auxiliar os pesquisadores, diante da necessidade, construiu-se um manual do sistema, que se encontra disponível para download no link <http://pseudomonas.procc.fiocruz.br/images/Tutorial.pdf>.

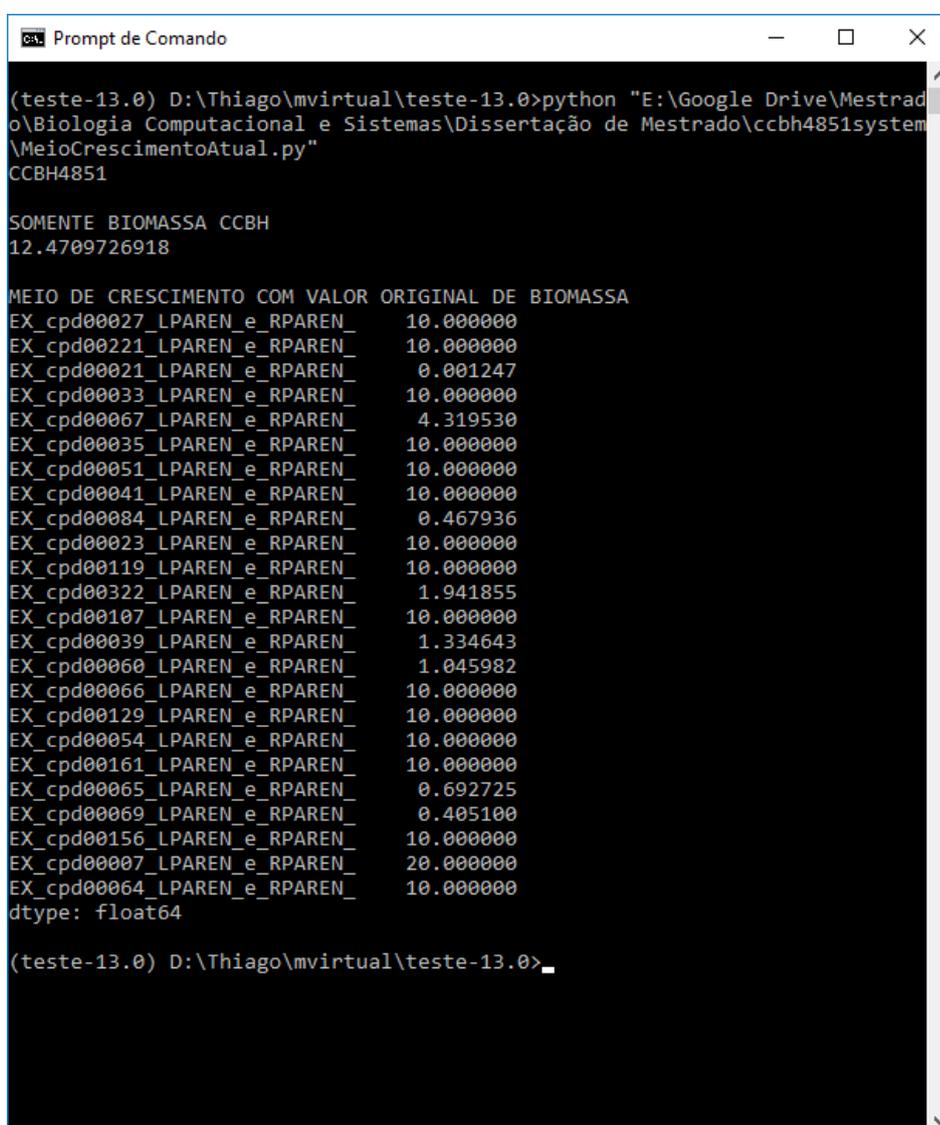
Manual do CurSystem



7.4 Apêndice D - Análise Complementar da Rede Metabólica mCCBH4851_v2

Para complementar a análise preliminar realizada pelo próprio NETGen e levantar mais informações biológicas acerca da segunda versão do modelo, o modelo foi submetido a outros métodos de análise da ferramenta *COBRAPy*.

A primeira análise realizada foi o cálculo do meio mínimo mantendo todas reações de *uptake* mapeadas do modelo mPAO1_v2017 para simulação de geração de biomassa atual no modelo mCCBH4851_v2 (14). O resultado da execução da análise indicativa é mostrado na Figura 52. Pode-se observar a simulação do meio de crescimento mapeado pela reconstrução e a lista de reações e seus fluxos mínimos para o valor atual dado pelo modelo.



```
(teste-13.0) D:\Thiago\mvirtual\teste-13.0>python "E:\Google Drive\Mestrado\Biologia Computacional e Sistemas\Dissertação de Mestrado\ccbh4851system\MeioCrescimentoAtual.py"
CCBH4851

SOMENTE BIOMASSA CCBH
12.4709726918

MEIO DE CRESCIMENTO COM VALOR ORIGINAL DE BIOMASSA
EX_cpd00027_LPAREN_e_RPAREN_ 10.000000
EX_cpd00221_LPAREN_e_RPAREN_ 10.000000
EX_cpd00021_LPAREN_e_RPAREN_ 0.001247
EX_cpd00033_LPAREN_e_RPAREN_ 10.000000
EX_cpd00067_LPAREN_e_RPAREN_ 4.319530
EX_cpd00035_LPAREN_e_RPAREN_ 10.000000
EX_cpd00051_LPAREN_e_RPAREN_ 10.000000
EX_cpd00041_LPAREN_e_RPAREN_ 10.000000
EX_cpd00084_LPAREN_e_RPAREN_ 0.467936
EX_cpd00023_LPAREN_e_RPAREN_ 10.000000
EX_cpd00119_LPAREN_e_RPAREN_ 10.000000
EX_cpd00322_LPAREN_e_RPAREN_ 1.941855
EX_cpd00107_LPAREN_e_RPAREN_ 10.000000
EX_cpd00039_LPAREN_e_RPAREN_ 1.334643
EX_cpd00060_LPAREN_e_RPAREN_ 1.045982
EX_cpd00066_LPAREN_e_RPAREN_ 10.000000
EX_cpd00129_LPAREN_e_RPAREN_ 10.000000
EX_cpd00054_LPAREN_e_RPAREN_ 10.000000
EX_cpd00161_LPAREN_e_RPAREN_ 10.000000
EX_cpd00065_LPAREN_e_RPAREN_ 0.692725
EX_cpd00069_LPAREN_e_RPAREN_ 0.405100
EX_cpd00156_LPAREN_e_RPAREN_ 10.000000
EX_cpd00007_LPAREN_e_RPAREN_ 20.000000
EX_cpd00064_LPAREN_e_RPAREN_ 10.000000
dtype: float64

(teste-13.0) D:\Thiago\mvirtual\teste-13.0>_
```

Figura 52 - Resultados obtidos no COBRAPy através de análise de meio mínimo do modelo mCCBH4851_v2

No intuito de validar o resultado da simulação do *COBRAPy*, todas as reações de *uptake* que NÃO estão na relação foram retiradas do modelo, em seguida, executou-se uma nova simulação de geração de biomassa, deixando somente as reações de *uptake* conforme a Tabela 10.

Id SBML	Substrato	V. Fluxo
EX_cpd00027_LPAREN_e_RPAREN_	D-Glicose	10.000000
EX_cpd00221_LPAREN_e_RPAREN_	D-Lactato	10.000000
EX_cpd00021_LPAREN_e_RPAREN_	Ferro	0.001247
EX_cpd00033_LPAREN_e_RPAREN_	Glicina	10.000000
EX_cpd00067_LPAREN_e_RPAREN_	Hidrogênio	4.319530
EX_cpd00035_LPAREN_e_RPAREN_	L-Alanina	10.000000
EX_cpd00051_LPAREN_e_RPAREN_	L-Arginina	10.000000
EX_cpd00041_LPAREN_e_RPAREN_	L- Aspartato	10.000000
EX_cpd00084_LPAREN_e_RPAREN_	L-Cisteína	0.467936
EX_cpd00023_LPAREN_e_RPAREN_	L-Glutamato	10.000000
EX_cpd00119_LPAREN_e_RPAREN_	L-Histidina	10.000000
EX_cpd00322_LPAREN_e_RPAREN_	L-Isoleucina	1.941855
EX_cpd00107_LPAREN_e_RPAREN_	L-Leucina	10.000000
EX_cpd00039_LPAREN_e_RPAREN_	L-Lisina	1.334643
EX_cpd00060_LPAREN_e_RPAREN_	L-Metionina	1.045982
EX_cpd00066_LPAREN_e_RPAREN_	L-Fenilalanina	10.000000
EX_cpd00129_LPAREN_e_RPAREN_	L-Prolina	10.000000
EX_cpd00054_LPAREN_e_RPAREN_	L-Serina	10.000000
EX_cpd00161_LPAREN_e_RPAREN_	L-Treonina	10.000000
EX_cpd00065_LPAREN_e_RPAREN_	L-Triptofano	0.692725
EX_cpd00069_LPAREN_e_RPAREN_	L-Tirosina	0.405100
EX_cpd00156_LPAREN_e_RPAREN_	L- Valina	10.000000
EX_cpd00007_LPAREN_e_RPAREN_	Oxigênio	20.000000
EX_cpd00064_LPAREN_e_RPAREN_	Ortina	10.000000

Tabela 10 - Reações de *uptake* e fluxos mínimos para geração de Biomassa no valor de 12.470970 mmol/g DCW/h indicadas na análise de meio mínimo

O resultado da simulação não mostrou variação, o valor original de biomassa foi mantido. É possível comparar na Figura 53, os valores das reações de *uptake* são alterados e o valor gerado de biomassa é mantido, fato que valida a simulação do modelo extraída pela ferramenta.

```

ca. Prompt de Comando
BIOMASSA CCBH
12.4709726918
0.0555808434267

MEIO DE CRESCIMENTO ATUAL
{'EX_cpd00001_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00067_LPAREN_e_RPAREN_': 10.0, 'EX_cpd0003
9_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00322_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00048_LPAREN_e_R
PAREN_': 10.0, 'EX_cpd00007_LPAREN_e_RPAREN_': 20.0, 'EX_cpd00254_LPAREN_e_RPAREN_': 10.
0, 'EX_cpd00021_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00051_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00
107_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00084_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00011_LPAREN_e
_RPAREN_': 10.0, 'EX_cpd00205_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00209_LPAREN_e_RPAREN_': 1
0.0, 'EX_cpd00064_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00035_LPAREN_e_RPAREN_': 10.0, 'EX_cpd
00060_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00023_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00161_LPAREN
_e_RPAREN_': 10.0, 'EX_cpd00013_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00066_LPAREN_e_RPAREN_':
10.0, 'EX_cpd00027_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00041_LPAREN_e_RPAREN_': 10.0, 'EX_c
pd00009_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00054_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00071_LPAR
EN_e_RPAREN_': 10.0, 'EX_cpd00069_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00033_LPAREN_e_RPAREN
_': 10.0, 'EX_cpd00156_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00119_LPAREN_e_RPAREN_': 10.0, 'EX
_cpd00129_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00065_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00221_LP
AREN_e_RPAREN_': 10.0}

MEIO DE CRESCIMENTO COM REDUCAO DE UPTAKES \{(COBRAPy\)
{'EX_cpd00067_LPAREN_e_RPAREN_': 4.31953, 'EX_cpd00039_LPAREN_e_RPAREN_': 1.334643, 'EX_
cpd00322_LPAREN_e_RPAREN_': 1.941855, 'EX_cpd00007_LPAREN_e_RPAREN_': 20.0, 'EX_cpd00021
_LPAREN_e_RPAREN_': 0.001247, 'EX_cpd00051_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00107_LPAREN
_e_RPAREN_': 10.0, 'EX_cpd00084_LPAREN_e_RPAREN_': 0.467936, 'EX_cpd00064_LPAREN_e_RPAREN
_': 10.0, 'EX_cpd00035_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00060_LPAREN_e_RPAREN_': 1.045982
, 'EX_cpd00023_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00161_LPAREN_e_RPAREN_': 10.0, 'EX_cpd000
66_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00027_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00041_LPAREN_e
_RPAREN_': 10.0, 'EX_cpd00054_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00069_LPAREN_e_RPAREN_': 0.
4051, 'EX_cpd00033_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00156_LPAREN_e_RPAREN_': 10.0, 'EX_cp
d00119_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00129_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00065_LPARE
N_e_RPAREN_': 0.692725, 'EX_cpd00221_LPAREN_e_RPAREN_': 10.0}

BIOMASSA MEIO MINIMO BIOMASSA ATUAL
12.4709708839
0.0555808514841 hours
(teste-13.0) D:\Thiago\mvirtual\teste-13.0>_

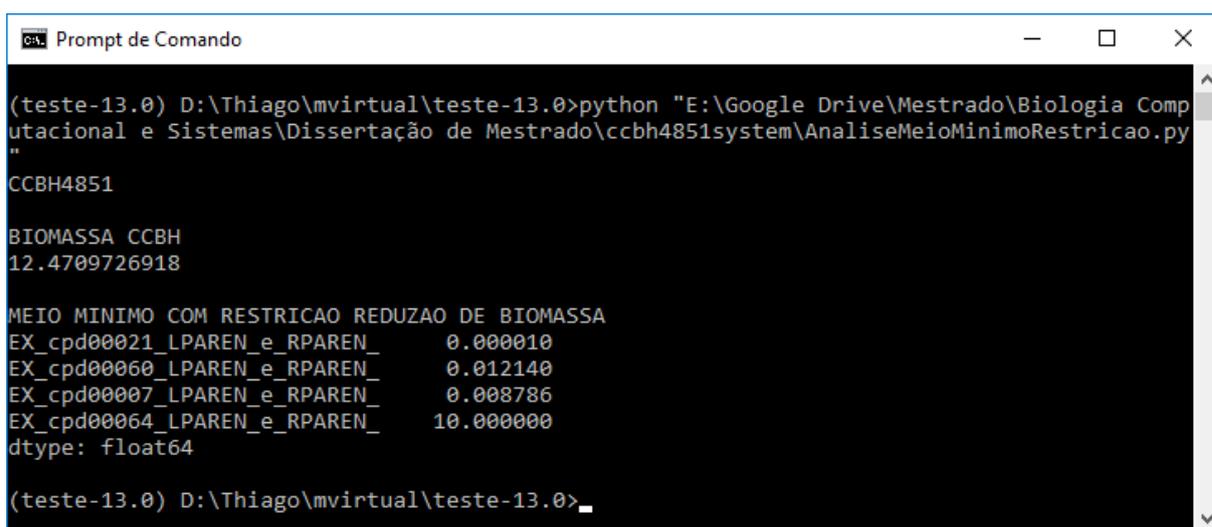
```

Figura 53 – Resultado da nova Simulação do modelo utilizando somente reações de uptake indicadas pelo *COBRAPy*

Uma segunda análise foi proposta, simular quais seriam os metabólitos e seus fluxos necessários para o modelo gerar um valor de biomassa determinado manualmente. Neste tipo de análise, o valor de biomassa é fixado, e a ferramenta ajusta as reações existentes do modelo para que o valor de biomassa resultante seja atendido (14).

A simulação do segundo cenário foi motivada pelo questionamento de quais metabólitos seriam necessários para que a simulação gerasse um valor de biomassa aproximado de um experimento *in vitro* de uma *Pseudomonas aeruginosa* (59). O valor aproximado seria de 1.04 mmol/g DCW/h (≈ 0.40 minutos) (12).

O resultado da análise pode ser visualizado pela imagem da tela capturada e exposta pela Figura 54. O *COBRApy* indicou que somente as quatro reações de uptake existentes no modelo seriam necessárias para gerar o valor de biomassa fixado para a simulação.



```

Prompt de Comando
(teste-13.0) D:\Thiago\mvirtual\teste-13.0>python "E:\Google Drive\Mestrado\Biologia Computacional e Sistemas\Dissertação de Mestrado\ccbh4851system\AnaliseMeioMinimoRestricao.py"
CCBH4851
BIOMASSA CCBH
12.4709726918
MEIO MINIMO COM RESTRICAO REDUZA O DE BIOMASSA
EX_cpd00021_LPAREN_e_RPAREN_      0.000010
EX_cpd00060_LPAREN_e_RPAREN_      0.012140
EX_cpd00007_LPAREN_e_RPAREN_      0.008786
EX_cpd00064_LPAREN_e_RPAREN_      10.000000
dtype: float64
(teste-13.0) D:\Thiago\mvirtual\teste-13.0>

```

Figura 54 – Resultado da análise e cálculo do Meio Mínimo do modelo mCCBH4851_v2 para gerar o valor de biomassa 1.04 mmol/g DCW/h

Novamente, prosseguiu-se com a intenção de validar a análise indicada pelo *COBRApy*, portanto, submeteu-se o modelo mCCBH4851_v2 a uma nova simulação de geração de biomassa utilizando somente as reações de uptake indicadas. As reações utilizadas para esta simulação estão listadas com mais detalhes abaixo na Tabela 11.

Id SBML	Substrato	V. Fluxo
EX_cpd00021_LPAREN_e_RPAREN_	Ferro	0.000010
EX_cpd00060_LPAREN_e_RPAREN_	L-Metionina	0.012140
EX_cpd00007_LPAREN_e_RPAREN_	Oxigênio	0.008786
EX_cpd00064_LPAREN_e_RPAREN_	Ortina	10.000000

Tabela 11 – Lista de reações indicadas pelo COBRApy e utilizadas na simulação de geração de biomassa no valor de 1.04 mmol/g DCW/h para meio mínimo com restrições

A simulação afirmou a lista de reações de *uptake* indicadas pelo *COBRAPy*, o valor de biomassa gerado foi igual ao valor dado como entrada como parâmetro para cálculo do **meio mínimo estrito**. A saída do resultado do *script* pode ser confirmada pela Figura 55.

```
ca. Prompt de Comando
(teste-13.0) D:\Thiago\mvirtual\teste-13.0>python "E:\Google Drive\Mestrado\Biologia Comp
utacional e Sistemas\Dissertação de Mestrado\ccbh4851system\CalculoBiomassaMM.py"

MODELO: CCBH4851

SOMENTE BIOMASSA CCBH
12.4709726918
0.0555808434267

MEIO DE CRESCIMENTO ATUAL
{'EX_cpd00001_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00067_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00039
_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00322_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00048_LPAREN_e_RPA
REN_': 10.0, 'EX_cpd00007_LPAREN_e_RPAREN_': 20.0, 'EX_cpd00254_LPAREN_e_RPAREN_': 10.0,
'EX_cpd00021_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00051_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00107_
_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00084_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00011_LPAREN_e_RPAR
EN_': 10.0, 'EX_cpd00205_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00209_LPAREN_e_RPAREN_': 10.0, '
EX_cpd00064_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00035_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00060_L
PAREN_e_RPAREN_': 10.0, 'EX_cpd00023_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00161_LPAREN_e_RPARE
N_': 10.0, 'EX_cpd00013_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00066_LPAREN_e_RPAREN_': 10.0, 'E
X_cpd00027_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00041_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00009_LP
AREN_e_RPAREN_': 10.0, 'EX_cpd00054_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00071_LPAREN_e_RPAREN
_': 10.0, 'EX_cpd00069_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00033_LPAREN_e_RPAREN_': 10.0, 'EX
_cpd00156_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00119_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00129_LPA
REN_e_RPAREN_': 10.0, 'EX_cpd00065_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00221_LPAREN_e_RPAREN
_': 10.0}

MEIO DE CRESCIMENTO COM REDUCAO PARA MEIO MINIMO
{'EX_cpd00060_LPAREN_e_RPAREN_': 0.126926, 'EX_cpd00007_LPAREN_e_RPAREN_': 0.977488, 'EX
_cpd00064_LPAREN_e_RPAREN_': 10.0, 'EX_cpd00021_LPAREN_e_RPAREN_': 0.000105}

BIOMASSA MEIO MINIMO
1.0455890769
0.66294416839 hours

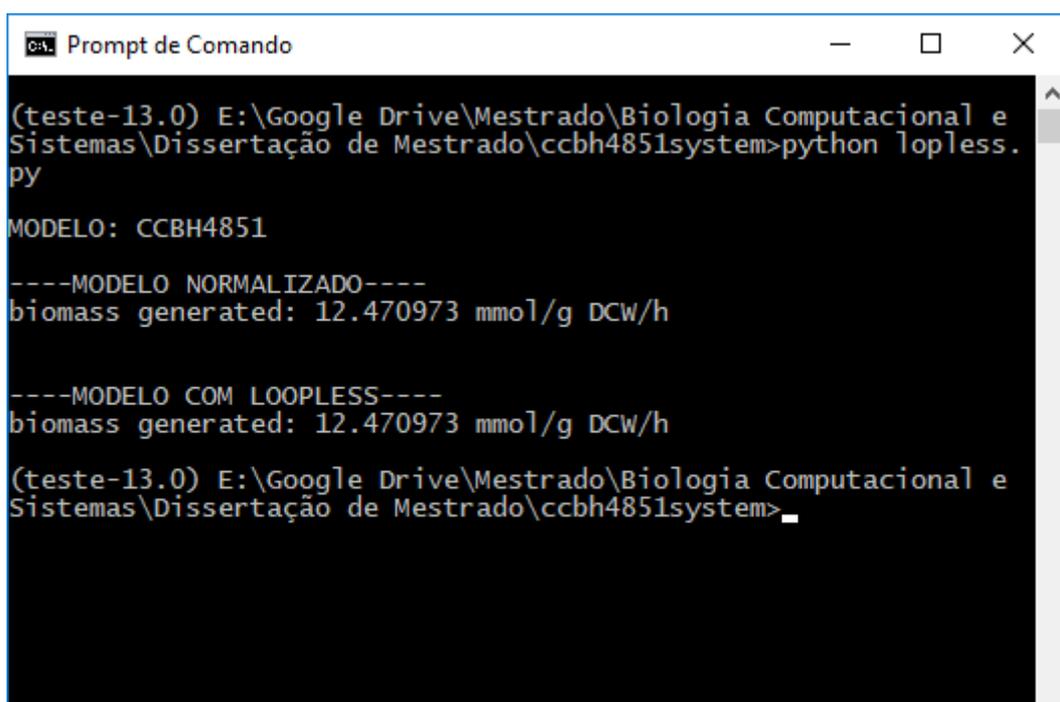
(teste-13.0) D:\Thiago\mvirtual\teste-13.0>_
```

Figura 55 – Simulação de geração de biomassa no meio mínimo estrito utilizando somente as reações indicadas pelo *COBRAPy*.

A terceira e última análise feita para o modelo mCCBH4851_v2 foi com o objetivo de verificar possíveis loops na rede. E ainda retirar fluxos cíclicos de metabólicos que podem estar causando alguma interferência na geração de

biomassa (14). Para isso, utilizou-se uma análise chamada de *loopless* do *COBRAPy*. Este método faz uma varredura e elimina os loops encontrados, em seguida, executa a simulação de geração de biomassa (14).

A análise obteve como resultado o valor de biomassa gerado idêntico ao valor original, se pode observar o resultado na Figura 56. Apesar da eliminação de alguns ciclos no fluxo de geração de biomassa, demonstrado no gráfico da Figura 57, os valores de biomassa são equivalentes, confirmando que o modelo não tem nenhuma interferência cíclica no resultado final no valor de biomassa.



```
(teste-13.0) E:\Google Drive\Mestrado\Biologia Computacional e
Sistemas\Dissertação de Mestrado\ccbh4851system>python loopless.
py

MODELO: CCBH4851

----MODELO NORMALIZADO----
biomass generated: 12.470973 mmol/g DCW/h

----MODELO COM LOOPLESS----
biomass generated: 12.470973 mmol/g DCW/h

(teste-13.0) E:\Google Drive\Mestrado\Biologia Computacional e
Sistemas\Dissertação de Mestrado\ccbh4851system>
```

Figura 56 - Resultado da biomassa gerada pela análise *loopless* da Rede da *P. aeruginosa* CCBH4851 – mCCBH4851_v2

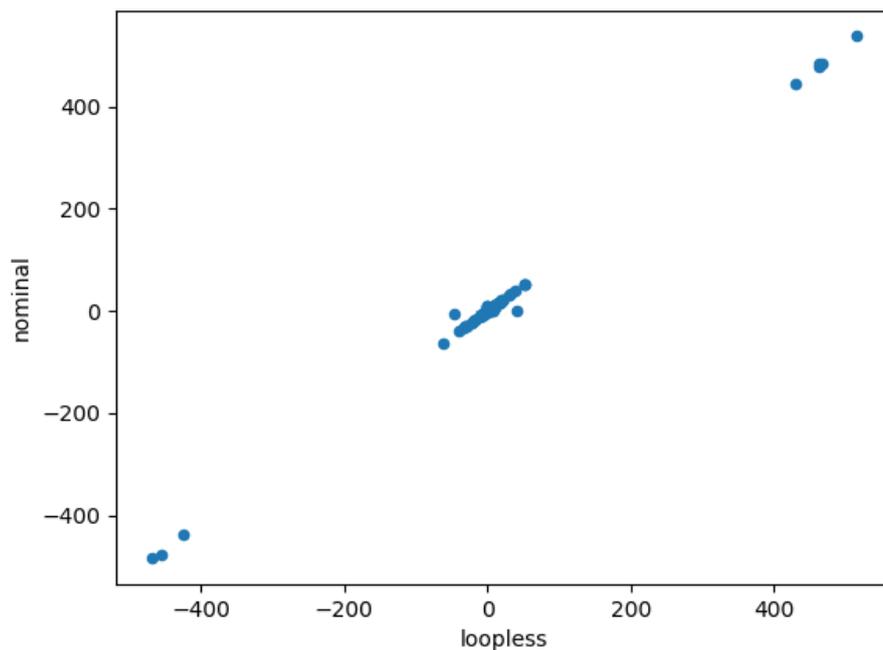


Figura 57 - Gráfico de Dispersão entre os dois fluxos de biomassa do modelo da *P. aeruginosa* CCBH4851

Conforme descrito nas publicações dos modelos da *Pseudomonas aeruginosa* em 2008 e 2017 (mPAO1_v2008, mPAO1_v2017 e mPA14), um modelo de rede metabólica traz inúmeras possibilidades de simulações, com diferentes resultados e análises (6,12). No caso deste trabalho buscou-se somente uma reconstrução de modelo metabólico inicial da *P. aeruginosa* CCBH4851, o principal objetivo era reconstruir um modelo que simule a geração de biomassa. Esse modelo existe, é o modelo mCCBH4851_v2. Com a evolução da curadoria espera-se que se consiga executar várias outras análises, como por exemplo a busca de alvos terapêuticos descritos por Meriguetti (9).

7.5 Apêndice E – Saída da análise do modelo mCCBH4851_v2 feita no *FindTargetsWeb*

	UNIPROTID	DRUGBANKID	DRUGBANKDRUGID	DRUGNAME	DRUGGROUP	PHARMAACTION	ACTIONS
15	P08506	BE0003685	DB00303	Ertapenem	approved, investigational	yes	inhibitor
14	P08506	BE0003685	DB00274	Cefmetazole	approved, investigational	yes	inhibitor
13	P08506	BE0003685	DB00430	Cefpiramide	approved	yes	inhibitor
12	P08506	BE0003685	DB01331	Cefoxitin	approved	yes	inhibitor
11	P08506	BE0003685	DB01329	Cefoperazone	approved, investigational	yes	inhibitor
10	P0A749	BE0000468	DB00828	Fosfomicin	approved	yes	inhibitor
0	P0A953	BE0000801	DB01034	Cerulenin	approved	yes	inhibitor
8	P0AC13	BE0000803	DB01582	Sulfamethazine	approved, investigational, vet_approved	yes	inhibitor
9	P0AC13	BE0000803	DB06729	Sulfaphenazole	approved	yes	inhibitor
6	P0AC13	BE0000803	DB01015	Sulfamethoxazole	approved	yes	inhibitor
5	P0AC13	BE0000803	DB00259	Sulfanilamide	approved	yes	inhibitor
4	P0AC13	BE0000803	DB00634	Sulfacetamide	approved	yes	inhibitor
3	P0AC13	BE0000803	DB00263	Sulfisoxazole	approved, vet_approved	yes	inhibitor
2	P0AC13	BE0000803	DB01298	Sulfacytine	approved	yes	inhibitor
1	P0AC13	BE0000803	DB00576	Sulfamethizole	approved, investigational, vet_approved	yes	inhibitor
7	P0AC13	BE0000803	DB01581	Sulfamerazine	approved, vet_approved	yes	inhibitor
17	P0AEB2	BE0002063	DB01331	Cefoxitin	approved	yes	inhibitor
16	P0AEB2	BE0002063	DB01329	Cefoperazone	approved, investigational	yes	inhibitor
18	P0AEB2	BE0002063	DB00274	Cefmetazole	approved, investigational	yes	inhibitor

Figura 58 - Lista de inibidores indicados pela análise do modelo da rede metabólica da *P. aeruginosa* CCBH4851 feita pelo FindTargetsWeb (9).