基于Django框架的Recon2代谢模型的检索与应用系统开发

作 者 姓 名： 米昊炀

指 导 教 师： 巩天星，Peter A.J Hilbers

单 位 名 称： 中荷生物医学与信息工程学院

专 业 名 称： 生物医学工程

东 北 大 学

2018年6月

**The Development of an Index and Application System upon Metabolic Model Recon2 Based on Django Framework**

by Haoyang Mi

Supervisor: Tianxing Gong, Peter A.J Hilbers

Northeastern University

June 2018

# **毕业设计（论文）任务书Tasks of Final Project**

|  |
| --- |
| **毕业设计（论文）题目：基于Django框架的Recon2代谢模型的检索与应用系统开发**  **Title of Final Project：The Development of an Index and Application System upon Metabolic Model Recon2 Based on Django Framework** |
| **设计(论文)的基本内容：**  **Basic contents of Final Project：**  (1) Import the .json format data to database and connect to Django.  (2) Develop administration and user page to manage metabolic data.  (3) Visualize and analyze complex network for any number of reactions.  (4) Achieve the online simulation and analysis of metabolic model |
| **毕业设计（论文）专题部分：**  **Special topic of Final Project（If have）**  **题目：**  **Title：**  **设计或论文专题的基本内容：**  **Basic contents of Final Project or special topic：** |
| **学生接受毕业设计（论文）题目日期**  **Date of this student got the title of Final Project**  **第　　周 week \_\_\_1\_\_\_**  **指导教师签字：**  **Signature of Supervisor:**  **年　　月　　日**  **Date** |

基于Django框架的Recon2代谢模型的检索与应用系统开发

摘要Abstract（In Chinese）

新陈代谢是机体生命活动的基本特征。系统生物学的发展使代谢网络的研究得到重视，并促进化学反应的分析趋向专门化，而计算机科学为解码人体复杂代谢提供了技术手段。基于此，人类全基因组尺度代谢模型Recon应运而生。Recon描述了人体内近乎所有的遗传代谢信息，为模拟分析提供了必要工具。然而，数据的复杂性使得目前没有针对Recon模型的单一数据库。且已有的模型检索系统功能较为单一，没有应用价值，故有关人类代谢网络的研究十分局限。

在本项目中，一种新型的人体代谢网络检索与应用系统被开发出来。我们选取Recon3D，一个最新的模拟人体全代谢网络模型，作为本系统的数据来源。通过可视化工具将数据库与Django框架相连接，创建数据访问接口。同时，管理与用户界面用来实现反馈、代谢数据的管理以及高级查找等检索功能。由于BiGG数据库提供的模型文件囊括了反应与代谢物的包含关系，以此为参考，在系统内实现了数据间的互联后，便可通过Javascript图形库设计匹配算法实现任意数量反应的可视化并生成力导向图。反应间的相互关系被确定后，则可通过悬停算法实现主要结点的突出与边的高亮。该算法有利于帮助分析代谢物的来源与去向，明确反应间的关系以及在代谢网络中的位置信息。

此外，通过导入COBRApy第三方库以及最优控制器，本系统还实现了代谢模型文件的上传以及在线模拟。用户可根据需要获取模型数据，并进行流平衡分析。

系统生物学的研究与人类健康关系紧密，通过引入人类代谢网络的模拟方法，有助于发现潜在的药物作用位点、预测药物作用并依此提出新型治疗方案。因此，对人类代谢模型数据的严格分类检索、清晰可视化以及在线分析有助于相关领域的进一步探索。

**关键词**：系统生物学；代谢网络；Recon3D；Django框架；Python开发；检索；可视化；COBRApy；在线模拟；流平衡分析

The Development of an Index and Application System upon Metabolic Model Recon2 Based on Django Framework

Abstract

Metabolism is the fundamental of vital movement. As the development of system biology, the research on human metabolism and the compartmentally analysis on biochemical reactions are gaining attentions. Simultaneously, computer science helps to decode the complex metabolic networks. Consequently, the human metabolism model Recon is created. Recon describes majority of metabolism information and contributes to simulation and modeling. However, there is no database solely available for Recon because of the data complexity, besides, the function of current index database is relatively simple and with no application value, so the research on human metabolic network(HMN) is very limited.

In this project, a brand-new index and application system of HMN is developed. Recon 3D, an updated model for metabolic simulation, is chosen as the data source. visualization tool connects database with Django framework and creates interface for data access. Meanwhile, administration and user page is designed for realize the feedback, data modify and advanced search function. Because the model file from BiGG database reflects the reactions-metabolites relationship, after achieving the interconnection, force directed graphs for any numbers of reactions visualization would be derived based on Javascript graph library and matching algorithm. When confirmed the relationship, the mouseover algorithm is processed to highlight major nodes and links. This helps to analyze the source and destination of metabolites, explicit reactions interrelation and their locations within networks.

Additionally，COBRApy package and optimizer solver are imported to handle with the metabolic model uploading and online analysis. Users can obtain metabolic data on request and conduct flux balance analysis.

System biology tightly associated with human healthcare. Introducing the HMN modeling methods would contribute to identify potential drug targets and test drug effects, upon which novel treatment is proposed. Therefore, the strict index, clear visualization and online analysis is conducive to explore relative research areas.

**Key words**：System Biology; Metabolic Network; Recon3D; Django Framework; Python Development; Index; Visualization; COBRApy; Online Analysis; Flux Balance Analysis

目录Index

**毕业设计（论文）任务书Tasks of Final Project** I

摘要Abstract（In Chinese） II

Abstract III

目录Index IV

第一章 ：绪论 Chapter I: Introduction 1

1.1 Computational modeling and simulation in system biology 1

第二章 ：工具与方法 Chapter II: Tools and Methodology 6

第三章 Chapter III 7

第四章 Chapter IV 8

第五章 Chapter V 9

参考文献References 10

致谢Acknowledgements 11

注：目录中一级标题顶格，二级标题缩进两字符，三级标题缩进四字符。

# 

# ：绪论 Chapter I: Introduction

The study mainly orients to the development of web-based system framework and related functions of model data application. With the reference of master thesis “Recon 2.2: from reconstruction to model of human metabolism”, I mainly worked with analyzing the structure and content of Recon model in order to design a system better fit the characteristic of Recon and utilize it. Following the guidance of Neil’s paper, I figured out the layout of metabolic model by evaluating Recon2.2 in SBML format. By comparison, the principle of how .json format file describe metabolic information is clarified. JSON is then considered acceptable for my project and thus an update has been made and the data source is switched to Recon3D.

## Computational modeling and simulation in system biology

## System biology is an interdisciplinary area of study that focuses on complex interactions within biological systems. The major goal is to ensembles different scales of information, upon which to understand the how biological networks function as an integrity. Obviously, unlike molecular biology, which only concerns the behavior of individual gene and proteins, system biology concentrate on emergent properties(mainly on genetic level) and how organisms functioning as a system. Just like experimental methods to traditional biology, computational

## Figure 1.1: The overview of signal transduction pathways

## and mathematical modeling are core towards system biology research. In 1952, one of first numerical simulation in cell biology was published[1]. In 1960, Denis Noble developed the first computer model of the heart pacemaker[2].

## Subsequently, several approaches have been used to study complex molecular systems, the computing efficiency exploded making large volume of biological data accessible and manageable. In 1997, a Japanese research group published the first quantitative model of the metabolism of a hypothetical cell. Around 2000, the well-known Institutes of System Biology have been founded both in Seattle and Japan. Meanwhile, several genomics projects have been put forwards, among which the Human Genome Project[3], led to novel, collaborative ways of working on problems in the biological fields of genetics. So far, system biology has gradually emerged as a relatively independent interdiscipline that combines computational and bioinformatics methods to explore cellular signaling networks, metabolic networks and genome-scale network domains.

## Genome-scale metabolic model

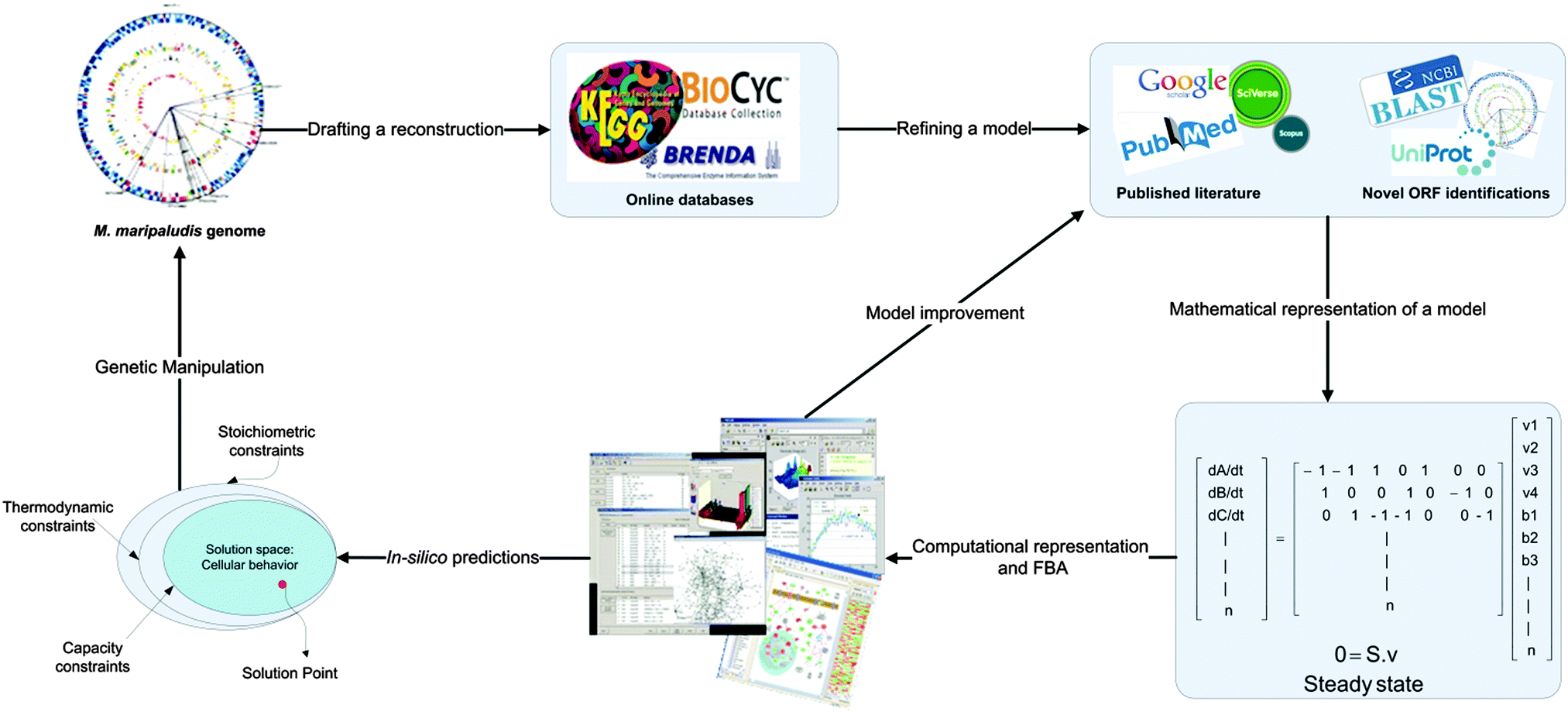
Metabolism can be defined as a comprehensive cluster of biochemical reactions to satisfy the fundamental need of living organisms. In order to quantitative analyze metabolic process, bioengineers introduced mathematical and modeling thoughts to simulate biological networks. To describe metabolic pathways as accurate as possible, researchers reconstructed genome-scale metabolic netork of chemical reactions from enzyme-reaction interrelations. Basically,

Figure 1.2: An example process of model reconstruction for a specific organism

a well-designed model should contains three compartments: genes, metabolites and reactions. The highly mutual related property becomes the basis for network reconstruction, and together with the curation/refinement procedure a model for simulation purpose is derived.

In 1999, Palsson’s group from University of California, San Diego published the first genome-scale metabolic network model that can be used to simulate the phenotype of metabolism[4]. Afterwards, a bunch of models have been constructed and the research on computational biology is increasing sharply. Many classical models have been identified, among which the E.coli model is gaining the most attentions. Escherichia coli K-12 MG1655 bacterial strain has 5 types of models, which are iJE660a GSM(1999)[5]、iJR904 GSM/GPR(2003)[6]、iAF1260(2007)[7]、iAF1260b(2010)[8] and iJO1366(2011)[9]. Multiple databases are established to store biological data, like KEGG, BiGG and BioModels. All of these advancements reflect the prosperous of computational and system biology.

## The overall research status on human metabolic network pathways

The construction of human metabolic model is based on the former experience of developing micro-organisms models. Plenty of useful databases, which are mutually supplementary, have been built for search in genome, metabolites or reactions. In 2007, the first model of human genome-scale metabolic networks models-Recon1 was finished and published based on KEGG database, and later was thermodynamically curated by LE Quek[10]. Almost at the same time, another high-quality human metabolic network was reconstructed by a research group from the University of Edinburgh. This model, as referred to as Edinburgh Human Metabolic Networks(EHMN), though comparable to Recon1 regarding the scale, is innovative since it reorganized 70 human-specific metabolic pathways according to their functional relationships[11].

In 2016, the ever-most predictive model update Recon2.2 have been published. With extensive manual curation, this model reaches the size of 5324 metabolites, 7785 reactions and 1675 genes[12]. Written in SBML format, this model is thus can be manipulated in multiple software like MATLAB and Pycharm. Combined with COBRA methods and visualization tools, the research on Recon2.2 has attracted increasingly number of investigators.

Recently, the model for human metabolic pathways has been further perfected. With more genes, metabolites and reactions been depicted in the model for the first time, the Recon3D, a most updated model for homo sapiens which was published in 2018, provides a more powerful source for human metabolic pathways, even the structure of gene variation.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Human Metabolic Model | | | | | | |
| Recon3D | Recon2.2 | Recon2.04 | Recon2 | EHMN | Quek | Recon1 |
| Metabolites | 5835 | 5324 | 5063 | 5063 | —— | 4962 | 2766 |
| Reactions | 10600 | 7785 | 7440 | 7440 | 2823 | 7785 | 3741 |
| Genes | 2248 | 1675 | 2140 | 2191 | 2322 | 1675 | 1905 |

## Table 1.1: Reconstruction scales for multiple human metabolic models

## The significance of developing index and application system for human metabolic model

## This project is initially inspired by the development history of human metabolic models. Within few years, multiple models and updates have been made upon human metabolic pathways research. However, the models were designed oriented from different perspectives, the simulation results may also vary to some degrees. To unify the further study on related areas, a standard model should be confirmed.

## The project is also by part inspired by my personal learning experience on genome-scale models. The prerequisite work for metabolic pathways is quite complicated. First the developing environment should be configured, and MATLAB or Python should be installed as the primary tool. Analysis package is ought to be imported in order to use analytical functions. Furthermore, optimizer solver should be set to perform mathematical simulation. Besides, the function of current existing databases for metabolism study are imperfect. For example, though BiGG database possesses the updated model for human metabolic networks, the visualization for core reactions is not available. Though KEGG database provides complete pathway maps for biochemical reactions and are accessible for download as well as clear reference for thousands of reactions, genes and pathways, it only functions for index not for any form of calculation. And even though BioModels database includes ever-most comprehensive models from *E.coli* to human beings and provide extremely detailed relevant information, SBML is the only format available for metabolic modeling, and consequently visualization is not possible.

## To significantly reduced the preparing work for researchers, in this study, I designed a web-based index and application system(IAS) for Recon model. Recon3D is considered as the ‘standard’ model, since it is the most updated version of Recon. Indexation is fully achieved to provide information look up function for bioengineers. Besides, constrained-analysis construction and analysis package in Python and optimizer solver are imported to project through backend. In that case bioengineers can complete their work without referencing multiple databases and wasting their time on environment configuration, instead, they could only use one system and finish simulating online.

Figure 1.3: Layout for IAS structure

# ：工具与方法 Chapter II: Tools and Methodology

## Operating system and development environment

## In ou

### Operating system selection

### Development environment

### 2.1.2.1 Introduction to Python Language

### 2.1.2.2 Introduction to Pycharm IDE

## Database selection and visualization tool

### Database selection

### Introduction to Navicat

## Webpage development technology

### Introduction to HTML5

### Introduction to CSS

### Introduction to Javascript graph library

## Analytic tool

### System Biology Markup Language(SBML)

### Constrained-Based Reconstruction and Analysis for Python(COBRApy)

# Chapter III

# Chapter IV

# Chapter V

[参考文献](file:////Users/mihaoyang/Downloads/20080512515147/模版/自动化本科毕业设计格式说明、模板及示例v2.0/桌面/论文模板/文献格式(篇数大于30篇且含近三年五篇文献)%201专著格式:序号.编著者.书名【M】,出版地:出版社,年代,起止页码%202期刊论文格式:序号.作者.论文名称【J】,期刊名称,年度,卷（期）:起止页码%203学位论文格式:序号.作者.学位论文名称【D】,发表地:学位授予单位,年度%20例:1.张艺.铸造工艺【M】,北京:机械工业出版社,1994,4-5%202.张颖伟.一类大组合系统的容错控制【J】,东北大学学报,2000,52(4):351-355%203.周丽.挖掘机的优化【D】:沈阳:沈阳东北大学:2000)References

1. Hodgkin A L, Huxley A F. A quantitative description of membrane current and its application to conduction and excitation in nerve[J]. The Journal of physiology, 1952, 117(4): 500-544.
2. NOBLE D. Cardiac action and pacemaker potentials based on the Hodgkin-Huxley equations [J]. Nature, 1960, 188(4749): 495.
3. Watson J D. The human genome project: past, present, and future[J]. Science, 1990, 248(4951): 44-49.
4. Schilling C H, Schuster S, Palsson B O, et al. Metabolic pathway analysis: basic concepts and scientific applications in the post‐genomic era[J]. Biotechnology progress, 1999, 15(3): 296-303.
5. Edwards J S, Palsson B O. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities[J]. Proceedings of the National Academy of Sciences, 2000, 97(10): 5528-5533.
6. Reed J L, Vo T D, Schilling C H, et al. Escherichia coli iJR904: an expanded genome-scale model of E. coli K-12[J]. Genome Biol, 2003, 4: R54.
7. Feist A M. iAF1260: A genome-scale metabolic model of E. coli as a platform for metabolic engineering[C]//The Annual Meeting and Exhibition 2007. 2007.
8. Feist A M, Zielinski D C, Orth J D, et al. Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli[J]. Metabolic engineering, 2010, 12(3): 173-186.
9. Orth J D, Conrad T M, Na J, et al. A comprehensive genome‐scale reconstruction of Escherichia coli metabolism—2011[J]. Molecular systems biology, 2011, 7(1): 535.

致谢Acknowledgements