

ONTOLOGY-BASED METABOLIC PATHWAY PREDICTION
USING *SACCHAROMYCES CEREVISIAE* DATA FROM
GENBANK, ECOCYC AND KEGG

YUSLINA BINTI ZAKARIA

UNIVERSITI TEKNOLOGI MALAYSIA

ONTOLOGY-BASED METABOLIC PATHWAY PREDICTION
USING *SACCHAROMYCES CEREVISIAE* DATA FROM
GENBANK, ECOCYC AND KEGG

YUSLINA BINTI ZAKARIA

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Science (Computer Science)

Faculty of Computer Science and Information Systems
Universiti Teknologi Malaysia

MAY 2006

ABSTRACT

Nowadays, metabolic pathway prediction is the focus of numerous drug discovery researches and is central to the stage of many biopharmaceutical and genomic companies. The identification and validation of drug targets depends critically on knowledge of the metabolic pathways in which potential target molecules operate within cells. To understand the cellular function, most scientists and biologists study parts of metabolic pathways that contain various types of pathways and large volume of knowledge including genes, enzymes, chemical compounds, and reactions that interlinked with each other. Thus, in order to provide better access to relevant knowledge, the knowledge of metabolic pathway should be conceptualized and formalized using appropriate knowledge representation technique. Currently, there is no ontology which is developed specify in metabolic pathway domain. Therefore, the main objective of this research is to develop an ontology-based representation for metabolic pathway to represent and to describe the concepts in the metabolic pathway domain and the relationships among them. Then, the metabolic pathway ontology is manipulated to predict and analyze metabolic pathways for a target organism using metabolic pathway prediction algorithm. To enhance the efficiency in predicting metabolic pathways, Problem Solving Method approach is proposed to perform the ontology inference by providing the reasoning component to solve the prediction problem of metabolic pathways. This proposed approach is implemented and tested using real data of *Saccharomyces cerevisiae* from GenBank and pathway reference databases from EcoCyc/MetaCyc and KEGG. This research presents the integration of ontology and metabolic pathway prediction algorithm as a possible solution for predicting metabolic pathways. This approach capable to predict metabolic pathway of *S.cerevisiae* with 87 percent accuracy compared to 80 percent accuracy using PathoLogic algorithm.

ABSTRAK

Kini, peramalan laluan metabolik menjadi fokus kepada banyak penyelidikan bagi penemuan ubat-ubatan dan tumpuan utama kepada organisasi biofarmaseutikal dan genomik. Pengenalpastian dan pengesahan bagi sasaran ubat-ubatan secara kritikal bergantung kepada pengetahuan laluan metabolik. Untuk memahami fungsi di dalam sel, kebanyakan ahli sains dan biologi mengkaji bahagian-bahagian laluan metabolik yang mengandungi pelbagai laluan dan saiz pengetahuan yang besar termasuk gen, enzim, bahan kimia dan tindak balas enzim yang berhubungan di antara satu sama lain. Bagi menyediakan capaian yang lebih baik kepada pengetahuan yang relevan, laluan metabolik perlu dikonseptualisasikan dan diformalkan menggunakan teknik perwakilan pengetahuan yang sesuai. Kini, tidak terdapat ontologi yang dibangunkan khusus untuk domain laluan metabolik. Oleh itu, objektif utama penyelidikan ini adalah untuk membangunkan perwakilan pengetahuan bagi laluan metabolik berasaskan ontologi yang boleh menerangkan konsep-konsep dan juga hubungan di antara konsep yang terdapat di dalam domain tersebut. Kemudian, ontologi bagi laluan metabolik tersebut digunakan untuk meramal dan menganalisa laluan serta tindak balas metabolik sasaran menggunakan algoritma peramalan laluan metabolik. Untuk meningkatkan keberkesanan dalam peramalan, Kaedah Penyelesaian Masalah (PSM) dicadangkan untuk melaksanakan penaakulan ke atas ontologi untuk menyelesaikan masalah peramalan. Pendekatan yang dicadangkan ini telah dilaksanakan dan diuji menggunakan data *Saccharomyces cerevisiae* dari GenBank dan pangkalan data rujukan dari EcoCyc/MetaCyc dan KEGG. Penyelidikan ini menunjukkan integrasi ontologi dan algoritma laluan metabolik sebagai salah satu penyelesaian masalah dalam meramal laluan metabolik. Ia berkemampuan untuk meramal laluan metabolik *S.cerevisiae* dengan ketepatan 87 peratus berbanding ketepatan 80 peratus menggunakan algoritma PathoLogic.

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
	DECLARATION	ii
	DEDICATION	iii
	ACKNOWLEDGMENTS	iv
	ABSTRACT	v
	ABSTRAK	vi
	TABLE OF CONTENTS	vii
	LIST OF TABLES	xi
	LIST OF FIGURES	xiii
	LIST OF APPENDICES	xvii
	LIST OF ABBREVIATIONS	xviii
1	INTRODUCTION	1
	1.1 Overview	1
	1.2 Background of Problem	2
	1.3 Problem Statement	7
	1.4 Motivation	8
	1.5 Objectives of the Study	9
	1.6 Scope of the Study	9
	1.7 Thesis Outline	10
2	LITERATURE REVIEW	11
	2.1 Introduction	11
	2.2 Molecular Databases	13
	2.3 Metabolic Pathway	16

2.3.1	Types of Metabolic Pathways	17
2.3.2	Metabolic Pathway Analysis	17
2.3.3	Metabolic Pathway Modeling	19
2.3.4	Metabolic Pathway Prediction	21
2.4	Knowledge Representation Techniques	24
2.4.1	Semantic Networks and Frame Systems	25
2.4.2	First-order Logic	27
2.4.3	Context Logic	28
2.5	Ontology-based Knowledge Representation	29
2.5.1	Type of Ontology	30
2.5.2	Classification of Ontology	31
2.5.3	METHONTOLOGY	33
2.5.4	Advantages and Disadvantages of Ontologies	34
2.6	Problem Solving Methods	35
2.7	Current Approaches	36
2.7.1	Existing Biological Ontologies	37
2.7.2	Comparison of Existing Biological Ontologies	41
2.7.3	Metabolic Pathway Databases	43
2.8	Summary	44
3	RESEARCH METHODOLOGY	45
3.1	Introduction	45
3.2	Research Design	46
3.3	Operational Framework	46
3.4	Data Collection and Preparation	49
3.5	Knowledge Modeling and Representation Using Ontology	50
3.6	Ontology-based Metabolic Pathway Prediction Using Modified PathoLogic	52
3.7	Problem Solving Method Approach for	53

	Predicting Metabolic Pathway	
3.8	Implement and Evaluate the Proposed Approach	53
3.9	Summary	53
4	ONTOLOGY-BASED REPRESENTATION FOR METABOLIC PATHWAY	55
4.1	Introduction	55
4.2	Ontology	56
4.2.1	Ontology Development Process	57
4.2.2	Ontology Building Framework	59
4.3	Metabolic Pathway	64
4.3.1	Metabolic Pathway Information	64
4.3.2	Metabolic Pathway Representation	66
4.4	Implementation and Results	67
4.4.1	Specification Phase	67
4.4.2	Conceptualization Phase	68
4.4.3	Knowledge Acquisition Phase	74
4.4.4	Integration Phase	76
4.4.5	Implementation Phase	77
4.5	Summary	82
5	ONTOLOGY-BASED METABOLIC PATHWAY PREDICTION USING PATHOLOGIC	83
5.1	Metabolic Pathway Prediction	83
5.2	Metabolic Pathway Prediction Using Modified PathoLogic	84
5.2.1	The Procedure for Metabolic Pathway Prediction Using PathoLogic Technique	84
5.2.2	Pathway Score	86
5.3	Rules Development	87

5.4	Implementation and Results	88
5.4.1	Linking Enzymes to Reactions	89
5.4.2	Rules Development for Metabolic Pathway Prediction	90
5.4.3	Prediction over MPO Ontology	92
5.5	Results of Metabolic Pathways Prediction	96
5.6	Summary	100
6	ONTOLOGY-BASED METABOLIC PATHWAY PREDICTION USING PROBLEM SOLVING METHOD (PSM)	101
6.1	Introduction	101
6.2	Problem Solving Method	102
6.2.1	PSM Development	102
6.2.2	Propose-and-Revise Technique	103
6.3	Metabolic Pathway Prediction Using Propose-and-Revise	105
6.4	Results	107
6.5	Summary	108
7	CONCLUSION AND FUTURE WORK	109
7.1	Conclusions	109
7.2	Research Contributions	110
7.3	Future Works	111
7.4	Closing	112
	BIBLIOGRAPHY	113
	APPENDIX A	119

LIST OF TABLES

TABLE NO.	TITLE	PAGE
2.1	A list of biological information sources for metabolic pathway analysis accessible via the Internet.	15
2.2	Metabolic databases concern in enzymatic-reactions	22
2.3	Comparison of available techniques in metabolic pathway reconstruction	23
2.4	Comparison of available ontologies in biological area	42
4.1	Concept dictionary in MPO ontology	73
4.2	Portion of binary-relations table in MPO ontology	73
4.3	Portion of instance table in MPO ontology	74
5.1	Portion of rules for Arginine Degradation VIII, Glycerol Degradation I, Phospholipid Biosynthesis I, Chorismate Biosynthesis inferred from MetaCyc generic reactions.	88

5.2	Portion of 192 pathways predicted using ontology-based representation and modified PathoLogic.	97
------------	--	----

LIST OF FIGURES

FIGURE NO.	TITLE	PAGE
1.1	The exponential growth of DNA sequences in GenBank from 1982 to 2005	3
1.2	Known genes and their functional contribution in different species from KEGG	4
1.3	Target Identification in modern drug discovery process	5
1.4	Target Validation in modern drug discovery process.	6
2.1	Growth of molecular biological databases that are collected by NAR	13
2.2	Typical diagram representation of metabolic pathways (urea cycle) (a) ExPASy (b) KEGG.	20
2.3	TAMBIS Architecture	38
2.4	Pathways Ontology in MBO	40
3.1	Research operational framework	48

3.2	The main phases of this research	49
3.3	The development of ontology-based representation	51
4.1	Steps in the ontology development process	57
4.2	Ontology development life cycle	59
4.3	Set of Intermediate Representation (IRs) in the conceptualization phase	61
4.4	Seven reactions from the Gluconeogenesis pathway.	65
4.5	Ontology requirement specification document for MPO	68
4.6	Conceptualization according to METHONTOLOGY	69
4.7	Portion of concept-classification-trees in MPO.	70
4.8	Diagram of binary-relations in MPO ontology.	71
4.9	Screenshot of defining classes for MPO ontology using Protégé	77
4.10	Screenshot of defining slot for each class in MPO ontology using Protégé	78
4.11	Screenshot of Chorismate Biosynthesis instance in MPO Ontology using Protégé	79
4.12	Main classes for MPO ontology	80

4.13	Expanded classes for MPO ontology	81
4.14	Instances for MPO ontology	82
5.1	Procedure for metabolic pathway prediction using PathoLogic algorithm	85
5.2	Linking enzymes to reaction algorithm	90
5.3	Portion of rules for inferring pathways using ESS	91
5.4	Procedure for metabolic pathway prediction using MPO ontology and PathoLogic algorithm	93
5.5	Linking enzymes to reaction algorithm by using MPO ontology	94
5.6	Pathway scoring algorithm to assess the evidence for the presence of metabolic pathways in organism	95
5.7	Comparison of EcoCyc/MetaCyc, KEGG and the proposed technique based on metabolic information	98
5.8	No of reactions predicted by EcoCyc/MetaCyc, KEGG and the proposed technique	99
6.1	The pseudo-code notation of the Propose-and-Revise algorithm of PSM	104
6.2	Procedure for predicting metabolic pathways based on the Propose-and-Revise technique	106

6.3	Comparison of ontology-based metabolic pathway prediction using modified PathoLogic and PSM	108
------------	---	-----

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	The Data of <i>Saccharomyces cerevisiae</i>	117

LIST OF ABBREVIATIONS

CLIPS	C Language Integrated Production System
DL	Description Logic
EC	Enzyme Commission
EMP	Enzyme and Metabolic Pathways
ER	Entity-Relation
FIPA	Foundation for Intelligent Physical Agents
FOL	First-order Logic
FS	Frame-Slot
DNA	Deoxyribonucleic Acid
GO	Gene Ontology
GT	Glossary of Terms
IR	Intermediate Representation
ESS	Java Expert System Shell
KBS	Knowledge-Based System
KEGG	Kyoto Encyclopedia of Genes and Genomes
KIF	Knowledge Interchange Format
KRL	Knowledge Representation Language
MBO	Molecular Biology Ontology
MPO	Metabolic Pathway Ontology
NAR	Nucleic Acid Research
PSM	Problem Solving Method

CHAPTER 1

INTRODUCTION

1.1 Overview

Metabolism is the chemical engine that drives the living process. Through the utilization of a large repertoire of enzymatic reactions and transport processes, unicellular and multicellular organisms can process and convert thousands of organic compounds into the various biomolecules necessary to support their existence. The definition of a metabolic pathway lies in the basis of metabolic bioinformatics. In accordance with the Enzyme and Metabolic Pathways database (EMP) nomenclature (www.biobase.com/EMP), a metabolic pathway is a set of oriented reactions interacting under given physiological conditions via simple or apparently simple intermediates. This definition is based on the definitions of metabolic intermediates. The term 'pathway' refers to a sequence of related reactions that start with a given compound and terminate in intermediary metabolism, metabolic cycles, or when no further microbial catabolism is presently known.

An understanding of the structural design and capabilities of the cellular metabolic network clearly places the biochemical engineer in an advantageous position to manipulate the cell for various purposes. Like many areas of biological study today, the future of metabolic pathway analysis may depend greatly upon its ability to capitalize on the wealth of genetic and biochemical information that currently being generated from the fields of genomics, and similarly proteomics.

This research focuses on developing metabolic pathway representation to model metabolic pathway knowledge based on metabolic information. The knowledge representation of metabolic pathway is then will be manipulated to predict and analyze metabolic reactions for target organism using metabolic pathway prediction techniques. The next section will discuss some background of problem and in Section 1.3, the problem statements. Section 1.4 will discuss the motivation of the research, that is, why metabolic pathway analysis is needed prior to prediction technique is presented. The goal and objectives of the research will be presented in Section 1.5 and the scope of the research will be presented in Section 1.6. The thesis outline for the following chapters will close the chapter in the final section.

1.2 Background of Problem

In the past, much of biological research has focused on data collection. The main reason for this is a much work needed to gather data and information. However biology is changing, especially because of the availability of large amounts of molecular data that is easily accessible via the Internet (Collado-Vides and Hofestadt, 2002 and Covert *et al.*, 2001). Genome projects generate enormous amounts of information. The amount of sequence data is increasing exponentially over time (Figure 1.1), and this growth will likely continue for the next future.

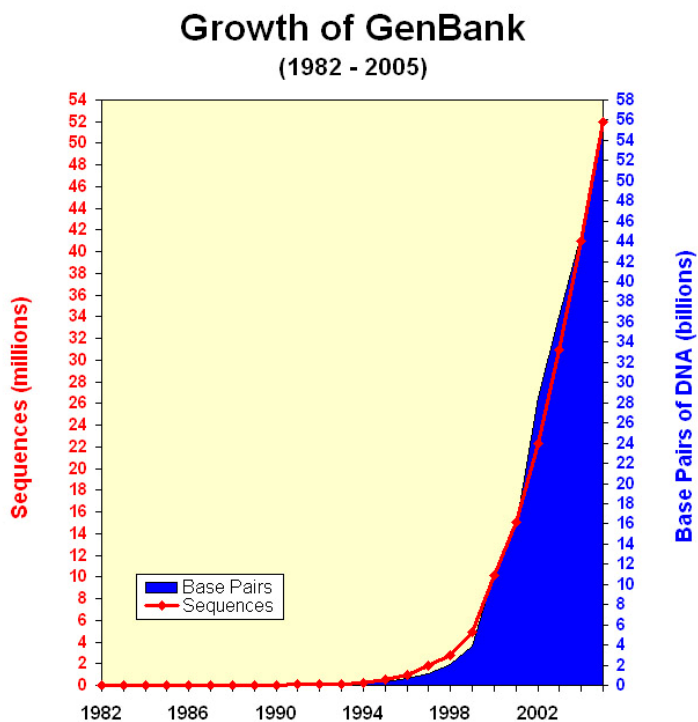


Figure 1.1 : The exponential growth of DNA sequences in GenBank from 1982 to 2005. (Revised: March 7, 2006).

The diversity and accumulation of biological data both on genomic and metabolic levels from different species (Figure 1.2) bring a new challenge for revealing what life really is. Extraordinary successes of the genome projects push the need for the development of more sophisticated and powerful computational techniques.

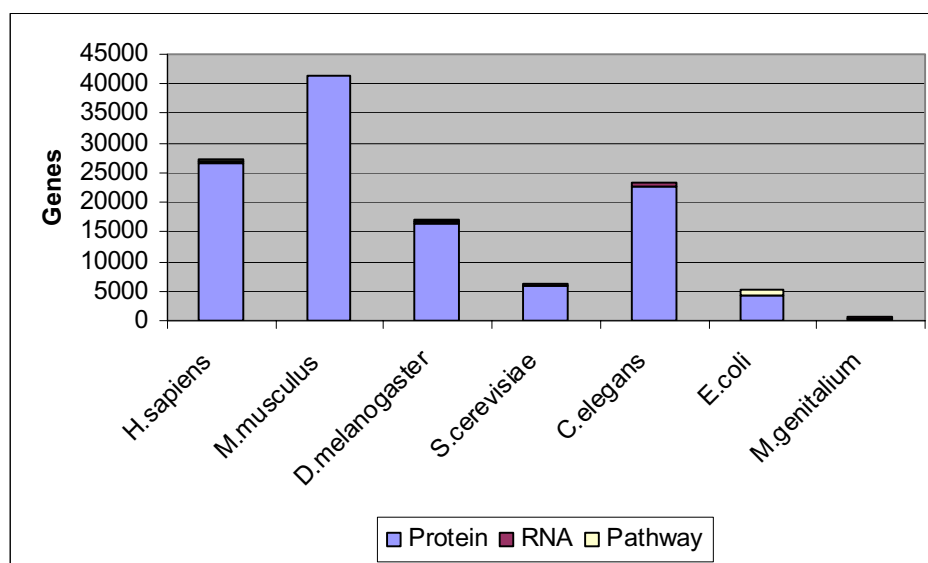


Figure 1.2: Known genes and their functional contribution in different species from KEGG. (Revised: 8 May 2006)

Moving from sequence to structure to function to application, bioinformatics developments are occurring in genome modeling and annotation, comparative protein modeling and folding assignment, *in silico* drug design, and modeling of cellular processes as well (Goesmann *et al.*, 2002). For example, metabolic pathways are used in two important steps in modern drug discovery process that are target identification (Figure 1.3) and target validation (Figure 1.4) process.

In Figure 1.3, the circle in metabolic pathway diagram (right panel) shows the particular area of metabolism that is affected in this particular disease model. Once identified, the targets, proteins or enzymes involved in creating the metabolic change can be deduced. While in this particular example of Figure 1.4, two different genes encoding candidate targets are eliminated in knockout mouse models. The tissue is then evaluated through metabolic profiling and compared with the disease. If the sample shows a similar or identical metabolic fingerprint, as in the panel on the left, then the

target is validated metabolically. If it is not similar (right panel), then the target is predicted not a good candidate.

Biological data functional analysis is a major topic beyond genome research. Computational metabolic analysis focuses on the computational interpretation of cellular phenomena that involve not only in nucleic acid and protein sequences, but also metabolic pathways. In this sense, systems analysis of metabolic network is becoming a promising field.

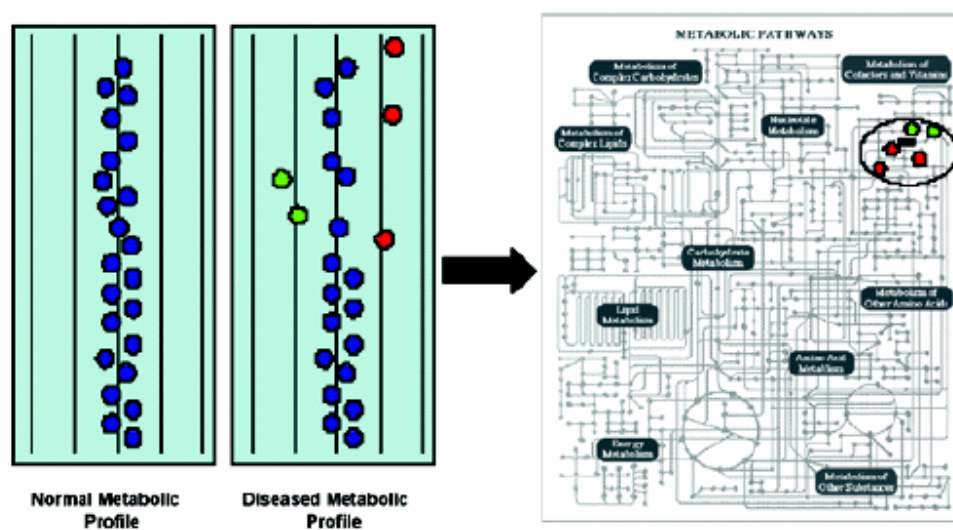


Figure 1.3: Target Identification in modern drug discovery process

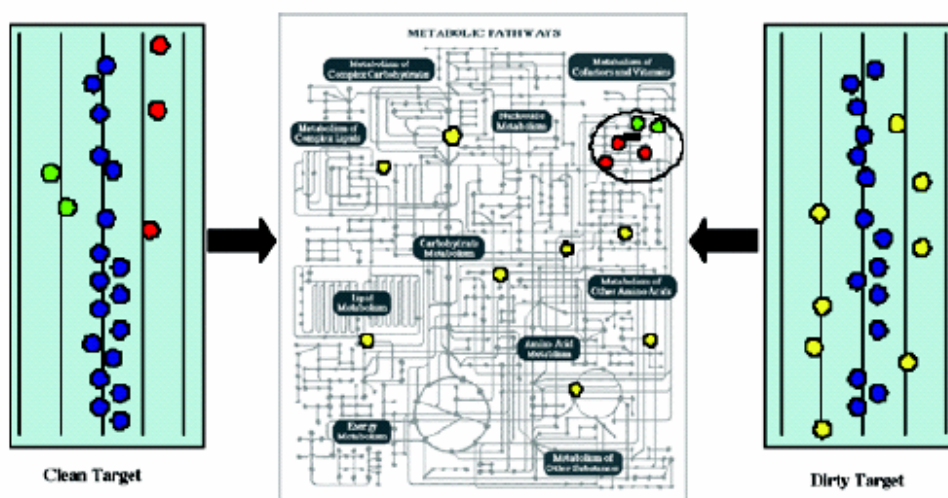


Figure 1.4: Target Validation in modern drug discovery process.

The development of computer science makes it possible to represent the complex metabolic network of physical and functional interactions, which take place in living cells, which enable us to manipulate, analyze and understand of how cells function.

In order to understand the logic of cells, methods of systems modeling and simulation are needed to find the interrelationships among different molecules and reactions. A major current challenge in biology is to clarify the relationship between structure, function and regulation in complex cellular networks. Metabolic pathway analysis offers the possibility to discover and analyze meaningful routes in metabolic networks. Fortunately the data and knowledge of genes, proteins and pathways are available, and various biology database systems are also accessible.

1.3 Problem Statement

In the analysis of metabolic pathway, the basic questions are:

Problem 1: *Given a model of metabolism and a set of enzymes, is it possible to synthesize each of a defined set of compounds?*

Problem 2: *Given a model of metabolism and a particular compound, what enzymes are necessary to synthesize that compound?*

Problem 3: *Given an incomplete/incorrect model of metabolism, a set of enzymes, and a set of compounds that can be synthesized, discover missing/incorrect reactions.*

Now, suppose a patient gets a diagnosis of metabolic disease (a disorder caused by malfunction of normal enzyme reactions), what is the metabolic mechanism of it? Most diseases are related to some kind of enzyme insufficiency and the malfunction of signal transduction pathways which regulate the expression of the genes that encode the desired enzymes. A good model of the metabolic reactions is appreciated to see the detail information about the essential proteins or enzymes and their regulations to the disease. With such a model, we might easily figure out the real causes, further development of the disorder, and possible alternative pathways to overcome the blockades.

Based on literature review that has been done, there is a need for computational techniques which can precisely answer all of the questions above to make the tasks for biologists easier. There are a few questions that have to be described:

1. How to represent all metabolic pathway knowledge into a conceptual form so that it can be easy to understand and formulate?
2. How to infer the metabolic pathway representation so that it can predict metabolic pathways precisely?

1.4 Motivation

Although in the post-genomic era sequence analysis have been and still are the most common topics in the bioinformatics studies, bioinformaticists and biologists are now looking for computational methods and tools to predict functional details. This takes bioinformatics beyond its original boundaries. It is certainly not data acquisition for molecular biology, but it is about the application of computer techniques, such as data abstraction, data manipulation, modeling, simulation, and functional analysis. The data generated by the experimental scientists requires annotation and detailed analysis in order to turn it into knowledge that can then be applied to, for example, healthcare, agriculture, industry and environment, to improve health care via gene prediction, drug design, gene therapy, and much more.

Today's users and IT professionals have high expectations towards software applications which are motivation for an ontology-based approach:

- i. They want to access the content they need
- ii. This content must be accurate and free of redundancy
- iii. The application must be intuitive and easy to use
- iv. The application must be reusable and extendable
- v. The application must be implemented in a short and inexpensive way and within
- vi. The current IT legacy environment

1.5 Objectives of the Study

The aim of this research is to develop an ontology-based metabolic pathway prediction using Problem Solving Method (PSM) approach. In order to achieve this aim, the following objectives must be fulfilled:

1. To analyze and study metabolic pathways information and prediction.
2. To construct an ontology to describe and conceptualize metabolic pathway information.
3. To develop the metabolic pathway prediction algorithm using PSM approach.

1.6 Scope of the Study

This research focuses on metabolic pathway prediction for *Saccharomyces cerevisiae* (*S.cerevisiae*). The data is already background-corrected and based on *in vitro* experiments, and these problems are not the focus of this research. Full dataset is needed for the experiment and no missing data is allowed.

Metabolic pathway prediction problems are the core in this research where the goal is to develop metabolic pathway prediction techniques for predicting metabolic pathways existing in an organism. Several prediction techniques have been studied. Moreover, several software tools have been developed to assist reconstruction of pathways. For instance, PathoLogic (Paley and Karp, 2002) is used by Sophia *et al.* (2003) and PathMiner by McShan *et al.* (2003). However, these approaches to predict

each gene function based on sequence similarity searches often fail to reconstruct cellular functions with all the necessary components.

The results are compared to other cited literature based on the classification performance from their research experiments.

1.7 Thesis Outline

This remainder of this thesis is subdivided into the following chapters:

Chapter 1, Introduction, provides a brief overview of research background, brief description of defined keywords, motivation and objective of this research.

Chapter 2, Literature Review, reviews briefly the available literature dealing with the defined keywords, the contributions of other researchers in the areas and provides a description of issues and problem modeling.

Chapter 3, Research Methodology, reviews and describes the research methodology, research design, operational framework that will be performed during the development.

Chapter 4, Ontology-based Representation for Metabolic Pathway, describes the process of ontology modeling, ontology development, and implementation of metabolic pathway ontology.

Chapter 5, Metabolic Pathway Prediction Using Modified PathoLogic Algorithm, describes the prediction of metabolic pathway for target organism, *Saccharomyces cerevisiae* using ontology and modified PathoLogic algorithm.

Chapter 6, Metabolic Pathway Prediction using Problem Solving Method (PSM), describes the prediction of metabolic pathway using PSM technique, Propose-and-Revise.

Chapter 7, Conclusions and Future Work concludes the thesis by summarizing the project including the research contributions and provides suggestions for future research.

BIBLIOGRAPHY

- Aguado, G. , Bañón, A., Bateman, J.A, Bernardos, S., Fernández, M., Gómez-Pérez, A., Nieto, A., Olalla, A., Plaza, R., and Sánchez, A.(1998). Ontogeneration: Reusing Domain and Linguistic Ontologies for Spanish Text Generation, *Workshop on Application of Ontologies and Problem Solving Methods* (part of ECAI '98: 1996 European Conf. AI), European Coordinating Committee for Artificial Intelligence, 1-10.
- Angele, J., Fensel D., Studer R. (1998). Developing Knowledge-based Systems with MIKE. *Journal of Automated Software Engineering*. **5(4)**:389-418.
- Barcza, S. , Kelly, L. A. and Lenz, C. D. (1990). Computerized Retrieval of Information on Biosynthesis and Metabolic Pathways. *Journal of Chemical Information and Computer Science*. **30**: 243–251.
- Benson, A. D. , Karsch-Mizrachi, I. and Lipman, J. D. (2003). GenBank. *Nucleic Acids Res.*, **31(1)**:23–27.
- Bobrow, D. and Winograd, T. (1977). An Overview of KRL, A Knowledge Representation Language, *Cognitive Science*, **1(1)**.
- Boeckmann, B., Bairoch, A. and Apweiler, R. (2003). The Swissprot Protein Knowledgebase and Its Supplement TrEMBL in 2003, *Nucleic Acids Res.*, **31(1)**: 365–370.
- Bono, H., Ogata, H. and Goto , S. (1998). Reconstruction of Amino Acid Biosynthesis Pathways from the Complete Genome Sequence, *Genome Research*. **8**: 203–210.
- Bower, J.M., and Bolouri, H. (2001). Computational Modeling of Genetic and Biochemical Networks, Massachusetts Institute of Technology.
- Bylander, T. and Chandrasekaran ,B. eds (1988). Generic Tasks in Knowledge-based Reasoning: The Right Level of Abstraction for Knowledge Acquisition. In: B.

- Gaines and J. Boose, *Knowledge Acquisition for Knowledge Based Systems*, London: Academic Press, **1**: 65-77.
- Chandrasekaran, B., Johnson, T. R. and Smith, J. W.(1992). Task-structure Analysis for Knowledge Modeling. *Communications of the ACM*, **35(9)**: 124-137.
- Chen, M. and Hofestädt, R. (2004). Web-based Information Retrieval System for the Prediction of Metabolic Pathways. *IEEE Transactions on Nanobioscience*, **3(3)**: 192-199.
- Chen, M. and Hofestädt, R. (2004). PathAligner: Metabolic Pathway Retrieval and Alignment. *Applied Bioinformatics*, **3(4)**: 241-252.
- Collado-Vides, J., Hofestädt, R. (2002). Gene Regulation and Metabolism, *Post Genomic Computational Approaches*, MIT Press, Cambridge, MA.
- Covert, M., Schilling, W. and Famili, C.H. (2001). Metabolic Modeling of Microbial Strains *in silico*. *Trends in Biochemical Sciences*, **27**: 179–186.
- Danchin, A. (1997). Comparison Between *Escherichia Coli* and *Bacillus Subtilis* Genomes Suggests that a Major Function of Polynucleotide Phosphorylase is to Synthesize CDP, *DNA Resources*, **28**: 9-18.
- Darvas, F., Predicting Metabolic Pathways by Logic Programming. *Journal of Molecular Graphics*, **6**: 80–86.
- Fensel, D., Benjamins, V.R., Motta, E., and Wielinga, B.(1999). UPML: A Framework for Knowledge System Reuse. *Proceedings of the 11th International Flairs Conference (FLAIRS-98)*, Sanibal Island, Florida.
- Fensel, D., Decker, S., Erdmann, and Studer, R. (1998). Ontobroker: How to Enable Intelligent Access to The WWW. Technical Report WS-98-14. AAAI Press.
- Fensel, D., Motta, E. (2001). Structured Development of Problem Solving Methods, *IEEE Transactions on Knowledge and Data Engineering*, **13(6)**:913-932.
- Fernandez, M., Gomez-Perez, A. and Juristo, N. (1997). METHONTOLOGY: From Ontological Art to Ontological Engineering. *Workshop on Knowledge Engineering: Spring Symposium Series (AAAI'97)*, 33-40, Menlow Park, Ca: AAI Press.

- Fernandez, M., Gomez-Perez, A. and Vicente, A.D. (1996). Towards a Method to Conceptualize Domain Ontologies. *Proceedings of ECAI96's Workshop on Ontological Engineering*, 41–52.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.-F., Dougherty, B.A., Merrick, J.M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J.D., Scott, J., Shirley, R., Liu, L.-I., Glodek, A., Kelley, J.M., Weidman, J.F., Phillips, C.A., Spriggs, T., Hedblom, E., Cotton, M.D., Utterback, T.R., Hanna, M.C., Nguyen, D.T., Saudek, D.M., Brandon, R.C., Fine, L.D., Fritchman, J.L., Fuhrmann, J.L., Geoghagen, N.S.M, Gnehm, C.L., McDonald, L.A., Small, K.V., Fraser, C.M., Smith, H.O, and Venter, J.C. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**: 496-512.
- Forst, C.V. and Schulten, K. (1999). Evolution of Metabolisms: A New Method for the Comparison of Metabolic Pathways Using Genomics Information, *Journal of Computational Biology*. **6**: 343-360.
- Gaasterland, T. and Selkov, E. (1995). Reconstruction of Metabolic Networks using Incomplete Information. *Proceedings of 3rd International Conference of Intelligent Systems for Molecular Biology*, 10268-10273.
- Gaasterland, T., Sensen, C.W. (1996). MAGPIE: Automated Genome Interpretation. *Trends Genetic*. **12**: 76-78.
- Galperin, M.Y. (2004). The Molecular Biology Database Collection: 2004 update, *Nucleic Acids Res.*, **32(1)**:D3-D22.
- Gennari, J.H, Tu, S.W, Rotenfluh, T. E, and Musen, M. A. (1998). Mapping Domains to Methods in Support of Reuse. *International Journal of Human-Computer Studies*, **41**:399-424.
- Goesmann, M., Haubrock, M., Meyer, F., Kalinowski J., and Giegerich, R. (2002). PathFinder: Reconstruction and Dynamic Visualization of Metabolic Pathways. *Bioinformatics*, **18(1)**:124-129.
- Goto, S., Nishioka, T. and Kanehisa, M. (1998). LIGAND: Chemical Database for Enzyme Reactions, *Bioinformatics*, **14**: 591-599.

- Gruber, T.R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2):199-220.
- Guarino, N. and Giaretta, P. (1995). Ontologies and Knowledge Bases: Towards a Terminological Clarification. In: N. Mars, ed., *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing*, Amsterdam: IOS Press, 25-32.
- Guarino, N. (1998). Formal Ontology and Information Systems. *Proceedings of Formal Ontology and Information Systems*, Trento, Italy: IOS Press.
- Guha, R.V. (1991). *Contexts: A Formalization and Some Applications*. Stanford University: Ph.D. Thesis.
- Kanehisa, M. (1997). A Database for Post-Genome Analysis, *Trends Genet.*, **13**: 375-376.
- Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes, *Nucleic Acids Research*, **28(1)**: 27-30.
- Kanehisa, M., Goto S., and Kawashima, S. (2002). The KEGG Databases at GenomeNet. *Nucleic Acids Res.*, **30(1)**:42-46.
- Karp, P. D. and Paley, S. (1994). Representations of Metabolic Knowledge: Pathways. In: Altman, R., Brutlag, D., Karp, P., Lathrop, R., and Searls, D., eds., *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*, 203-211.
- Karp, P.D. (1998). Metabolic databases, *Trends in Biochemical Sciences*, **23(3)**: 114-116.
- Karp, P.D., Riley, M., Paley, S.M. (1999). EcoCyc: Encyclopedia of Escherichia coli Genes and Metabolism, *Nucl. Acids Res.*, **27(1)**: 55, 59–61.
- Karp, P. D., Riley M. and Paley S.M. (2002). The MetaCyc Database, *Nucleic Acids Res.*, **30(1)**:59–61.
- Lenat, D., Guha, R.V., Pittman, K., Pratt, D., Shepherd, M. (1990). Cyc: Toward Programs with Common Sense. *Communication of the ACM*, **33(8)**:30-49.
- Marcus, S. (1998). *Automating Knowledge Acquisition for Experts Systems*. Kluwer Academic Publisher, Boston.
- McCarthy, J. (1993). Notes on Formalizing Context. *International Conference on Artificial Intelligence (IJCAI-93)*, 555-560.

- McShan, D. C., Rao, S., and Shah, I. (2003). PathMiner: Predicting Metabolic Pathways by Heuristic Search. *Bioinformatics*, **19(13)**: 1692–1698.
- Michal, G. (1982). *Biochemical Pathways Wall Chart*, Boehringer Mannheim GmbH Biochemical.
- Michal, G. (1999) *Biochemical Pathways: An atlas of Biochemistry and Molecular Biology*, Wiley, Heidelberg.
- Minsky, M. ed.(1975). A Framework for Representing Knowledge. In: Winston, P. *The Psychology of Computer Vision*, New York: McGraw Hill, 211-277.
- Minsky, M. (1975). A Framework for Representing Knowledge. In Patrick-Henry Winston, editor, *The Psychology of Computer Vision*. New York: McGraw-Hill.
- Mushegian, R.A. and Koonin, V.E. (1996). A Minimal Gene Set for Cellular Life Derived by Comparison of Complete Bacterial Genomes, *Proceedings of the National Academy of Sciences USA*, 93: 10268-10273.
- Musen, M. (1993). An Overview of Knowledge Acquisition. In: J. M. David, J.P. Krivine, and Simmons R., eds., *Second Generation Expert Systems*. Springer Verlag.
- Overbeek, R., Larsen, N. and Pusch, G. D. (2000). WIT: Integrated System for High-throughput Genome Sequence Analysis and Metabolic Reconstruction. *Nucleic Acids Res.* **28(1)**:123-125.
- Paley, S. and Karp, P.D. (2002). *Evaluation Of Computational Metabolic Pathway Predictions for Helicobacter pylori*. *Bioinformatics*. **18(5)**: 715–724.
- Pireddu L., Poulin B., Szafron D., Lu P. and Wishart D.S. (2005). Pathway Analyst – Automated Metabolic Pathway Prediction. *Proceedings of the IEEE 2005 Symposium on Computational Intelligence in Bioinformatics and Computational Intelligence*, 243-250.
- Schomburg, I., Chang, A. and Schomburg, D. (2002). BRENDA, Enzyme Data and Metabolic Information. *Nucleic Acids Res.*, **30(1)**:47-49.
- Schreiber, A. T., Wielinga, B. J., Jansweijer, de Hoog, R. and Akkermans, J.M. (1994). CommonKADS: A Comprehensive Methodology for KBS Development. *IEEE Expert*, **9(6)**:28-37.

- Spann, M. L., Chou, K. C., Wipke, W. T. and Ouchi, G. J. (1978). Use of Computerized Methods to Predict Metabolic Pathways and Metabolites. *Journal of Environmental Pathology and Toxicology*. **2**:123–131.
- Kanehisa, M., Goto, S. and Kawashima, S. (2002). The KEGG Databases at Genome Net, *Nucleic Acids Res.*, **30(1)**: 42–46.
- Sophia, T., David, S., and Christos, A.Q. (2003) Automated metabolic reconstruction for *Methanococcus jannaschii*. *Arachaea*, **1**, Canada: Heron Publishing.
- Teije, A.T, Harmelen, F. V, Schreiber, A. T, and Wielinga, B. (1998). Construction of Problem-Solving Methods as Parametric Design. *International Journal of Human-Computer Studies*, Special Issue on Problem-Solving Methods. **49(4)**:363-389.
- Uschold, M., and Gruninger, M. (1996). Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review*, **11(2)**:93-113.
- Wipke, W. T., Ouchi, G. J. and Chou, J. T. (1983). Computer-Assisted Prediction of Metabolism. In: L. Golberg, ed. *Structure-Activity Correlations as Predictive Tools in Toxicology*, Washington, DC: Hemisphere, 151–166.