

# CONTENTS

	PAGE
PREFACE	i
ENGLISH ABSTRACT	ii
THAI ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
CONTENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	x
LIST OF TECHNICAL VOCABULARIES AND ABBREVIATIONS	xii
 <b>CHAPTER</b>	
<b>1. INTRODUCTION</b>	<b>1</b>
1.1 Background and Rationale	1
1.2 Objective of Study	2
1.3 Scope of Study	2
1.4 Expected Results	2
1.5 Benefit of Study	3
 <b>2. LITERATURE REVIEWS</b>	<b>4</b>
2.1 Clinical Pharmacology	4
2.1.1 Principles of Clinical Pharmacology	4
2.1.2 Drugs Development	7
2.2 Adverse Drug Reactions (ADRs)	9
2.2.1 Definitions	9
2.2.2 Epidemiology	10
2.2.3 Classification	11
2.2.4 Management	14
2.3 Database in Pharmacology	14
2.3.1 Chemical-Target	14
2.3.2 Chemical-Response	17
2.3.3 Genetic-Response	18
2.4 Biological Network	18
2.4.1 Concepts and Principles	19

	<b>PAGE</b>
2.4.2 Biological Data Integration	20
2.5 Knowledge Discovery (KD)	22
2.5.1 Data Mining	22
2.5.2 Literature Mining	25
<b>3. MATERIALS AND METHODOLOGY</b>	<b>28</b>
3.1 Datasets	28
3.1.1 The Medical Dictionary for Regulatory Activities (MedDRA)	28
3.1.2 HUGO Gene Nomenclature Committee (HGNC)	31
3.1.3 DrugBank	31
3.1.4 Canada Vigilance Adverse Reaction Online Database	31
3.1.5 PubMed	32
3.2 Applications and Tools	32
3.2.1 Weka	32
3.2.2 Literature Mining Tools	32
3.2.3 MySQL Database	36
3.3 Methodology	36
3.3.1 Data Type and Collection	36
3.3.2 Data Preparation	38
3.3.3 Database Construction	41
3.3.4 Interface Design	41
3.3.5 Consistency Check	41
3.3.6 Application Demonstration	43
<b>4. RESULTS AND DISCUSSIONS</b>	<b>44</b>
4.1 Database Construction	44
4.1.1 ADR-Protein Associations	45
4.1.2 Information of Drug	47
4.1.3 ADR-Drug-Protein Associations	48
4.2 Web Interface	50
4.3 Application Demonstration	52
4.3.1 Drug	52
4.3.2 Protein	53
4.3.3 ADR	55

	<b>PAGE</b>
<b>5. CONCLUSIONS AND SUGGESTIONS</b>	<b>58</b>
5.1 Conclusions	58
5.2 Suggestions	58
 <b>REFERENCES</b>	 <b>60</b>
 <b>APPENDIX</b>	
A. Python Source Code for Taverna Outputs Manipulation in Data Preparation Process of ADR-Protein Associations and Example	69
B. The Procedure for Installing AppServ and Applying in Database for ADRs	72
C. Data Dictionary of Database for ADRs (ER diagram was in Figure 3.5)	79
D. List of proteins which were found to connect with alcoholic fatty liver and hepatic steatosis from literature mining process	83
E. List of proteins which were the results of alcoholic fatty liver and hepatic steatosis from database for ADRs	88
 <b>CURRICULUM VITAE</b>	 <b>93</b>

## LIST OF TABLES

TABLE		PAGE
2.1	Total number of emergency admissions which diagnosis of ADRs from 1999 to 2008, modified from Wu, et al. (2010)	11
2.2	Classification of adverse drug reactions, modified from Edwards and Aronson (2000)	13
2.3	Comparison between the data content in DrugBank (release 1.0) versus DrugBank (release 2.0), modified from Wishart, et al. (2008)	16
2.4	Examples of large-scale scientific data integration across domains, modified from Searls (2005)	21
3.1	Twenty-six System Organ Classes in MedDRA version 13.1	30
3.2	The summarized of each dataset	37
3.3	The selected fields of drug and drug-protein information	38
4.1	The summary of datasets and the connected associations of ADR-drug-protein in database for ADRs construction	45
4.2	Results of four prediction models that were built from Weka using ten folds cross-validation technique	48
4.3	Top nine medications that were found to cause the most hepatobiliary disorders	53
4.4	The comparison between DrugBank and database for ADRs to provide the information about medications when searching by protein	54
4.5	The comparison between DrugBank and database for ADRs to provide the information about target proteins when searching by drug	55
4.6	The lists of mapped pathways of hepatic steatosis from Biocarta	56
4.7	The lists of mapped pathways of hepatic steatosis from KEGG	57
D.1	List of proteins which were found to connect with alcoholic fatty liver from literature mining process	84
D.2	List of proteins which were found to connect with hepatic steatosis from literature mining process	85

<b>TABLE</b>	<b>PAGE</b>
E.1 List of proteins which were the results of alcoholic fatty liver from database for ADRs	88
E.2 List of proteins which were the results of hepatic steatosis from database for ADRs	88

## LIST OF FIGURES

FIGURES	PAGE
2.1 Possible routes for metabolism of drugs (Thorp, 2008)	5
2.2 Changing view of drug actions (Berger and Iyengar, 2009)	7
2.3 Preclinical phase of pharmaceutical pipeline, modified from Roses (2004)	8
2.4 Clinical phase of pharmaceutical pipeline, modified from Roses (2004)	8
2.5 The success rate of new molecular entities (NMEs) through preclinical research and development (R&D) and clinical development phase (Csermely, et al., 2013)	9
2.6 Type A and Type B classification of adverse drug reactions, modified from Thien, (2006)	12
2.7 Statics of SIDER (Kuhn, et al., 2010a)	17
2.8 Degree: the number of links connected to node $i$	19
2.9 Process diagram of Cross Industry Standard Process for Data Mining (CRISP-DM), modified from Wirth (2000)	23
3.1 Five data sources and six purposes that were utilized in the database for ADRs	28
3.2 Structural Hierarchy of the MedDRA Terminology	29
3.3 The original BioAID_ProteinDiscovery workflow	34
3.4 The Modified BioAID_ProteinDiscovery workflow	35
3.5 Work process diagram representing the overall methodology	37
3.6 Illustration of manipulating outputs from Modified BioAID_ProteinDiscovery workflow in Taverna by program written in python language (example was clarified in Appendix A)	40
3.7 ER diagram of constructed database (type and explanation of attributes were in Appendix C)	42
3.8 Different users of the database for ADRs	43

<b>FIGURES</b>	<b>PAGE</b>
4.1 The matching proteins of alcoholic fatty liver and hepatic steatosis to the authenticated major pathways, modified from Purohit, et al. (2009)	46
4.2 Network of ADR-drug-protein associations of hepatobiliary disorders	48
4.3 The matching proteins of acetaminophen-induced hepatic necrosis to the mechanistic determinants in acetaminophen-induced hepatic necrosis, modified from Hinson, et al. (2010)	49
4.4 The matching proteins of acetaminophen-induced hepatic necrosis to the mechanism in repair of acetaminophen-induced hepatic necrosis, modified from Hinson, et al. (2010)	50
4.5 Web-based design of search page	51
4.6 Web-based design of page that repeat query from users and show possible word of it in the database	51
4.7 Web-based design of result page	52
4.8 Web-based design of browse page	52
A.1 Example of input and output from Python manipulating extracted ADR-Protein association using Modified BioAID_ProteinDiscovery workflow in Taverna	71
B.1 GNU/GPL license agreement screen	73
B.2 Choose install location screen	73
B.3 Choose package components screen	74
B.4 Apache web server configuration screen	74
B.5 MySQL Database configuration screen	75
B.6 Ready to use AppServ in web browser screen	75
B.7 Create new database named db4adr screen	76
B.8 Import .sql file into db4adr database screen	76
B.9 Add information and privilege for a new user screen	77
B.10 The appearance of proper Database for ADRs	77
B.11 StartConnect.inc source code	78

## **LIST OF TECHNICAL VOCABULARY AND ABBREVIATIONS**

ADME	=	Administration or Absorption, Distribution, Metabolism, and Elimination/Excretion
ADRs	=	Adverse Drug Reactions
AER	=	The Adverse Event Reposting
APAP	=	Acetaminophen
ATC	=	Anatomical Therapeutic Chemical
Bio-NLPdb	=	Bio-NLP resources database
Caco-2	=	A continuous line of heterogenous human epithelial colorectal adenocarcinoma cells
COSTART	=	Coding Symbols for a Thesaurus of Adverse Reaction Terms
CRISP-DM	=	Cross Industry Standard Process for Data Mining
CSV	=	Comma-Separated Values or Character-Separated Values
CYP	=	Cytochrome P450 monooxygenases
DAVID	=	Database for Annotation, Visualization and Integrated Discovery
ER	=	Entity Recognition
ER diagram	=	Entity Relationship Diagram
FDA	=	U.S. Food and Drug Administration, Department of Health and Services
G6PD	=	Glucose-6-Phosphate Dehydrogenase
HGNC	=	HUGO Gene Nomenclature Committee
HPRD	=	Human Protein Reference Database
HT	=	High-Throughput
HUGO	=	Human Gene Organization
ICH	=	International Conference on Harmonization
IE	=	Information Extraction
IR	=	Information Retrieval



KD	=	Knowledge Discovery
KEEG	=	Kyoto Encyclopedia of Genes and Genomes
LLT	=	Lowest Level Term
LogP	=	Water/octanol partition coefficient
LogS	=	Water solubility
MedDRA	=	The Medical Dictionary for Regulatory Activities
MSSO	=	The Maintenance and Support Services Organization
NCBI	=	The National Center for Biotechnology Information
NLP	=	Natural-Language Processing
NME	=	New Molecular Entity
PD	=	Pharmacodynamics
PDTD	=	Potential Drug Target Database
PGt	=	Pharmacogenetics
PGx	=	Pharmacogenomics
PharmGED	=	Pharmacogenetic Effect Database
PharmGKB	=	The Pharmacogenetics Knowledge Base
PK	=	Pharmacokinetics
PMI	=	Pointwise Mutual Information
PMID	=	PubMed Identifier
PT	=	Preferred Term
R&D	=	Research and Development
SIDER	=	Side Effect Resource
SOC	=	System-Organ Class
STITCH	=	Search Tool for Interactions of Chemicals
SVM	=	Support Vector Machine
TCM	=	Traditional Chinese Medicine
TRMP	=	A Database of Therapeutically Relevant Multiple Pathways

TTD	=	Therapeutic Target Database
WHO	=	World Health Organization
WSDL	=	Web Service Description Language
WWW	=	World Wide Web
XML	=	Extensible Markup Language