CHAPTER 4 RESULTS AND DISCUSSIONS

In the effort to progress the understanding of ADRs, prototype database for ADRs has been developed with the advance of systems pharmacology and knowledge discovery. By integrating various datasets into one, database for ADRs also took the opportunity to build the database of ADR-drug-protein relationships that no database has ever been constructed from these relationships. The created database was checked and discussed for its consistency and usefulness which were described below.

4.1 Database Construction

In order to construct the database for ADRs, six datasets were collected from five data sources. MedDRA, DrugBank, and HGNC were sources for complete reference data of ADR terms, drug information, and protein naming, respectively. Other datasets were gathered from DrugBank, Canada Vigilance Adverse Reaction Online Database, and PubMed that provided association data of drug-target protein, ADR-drug, and ADR-protein, respectively. PubMed was an extraordinary source to recognize ADR-protein relations which were accumulated by literature mining from Modified BioAID_ProteinDiscovery workflow in Taverna Workbench. Because the objective of this project was to construct the prototype database, only PTs in SOC of hepatobiliary disorders were applied for ADR-protein association discoveries. In addition, hepatobiliary disorders were the most common adverse indication leading to drug candidate failure or withdrawal from the market (Watkins, and Seeff, 2006; Uetrecht, 2009) and numbers of them were 163 terms that were suitable to test the prototype database.

After collecting six datasets, data preparation was applied. This process had to be performed for duplicate data removal, field selection, and text editing. While ADR-protein associations were additionally manipulated by python programming in Appendix A, drug information was investigated for the predictive score for ADR class using data mining in Weka. Based on the construction of hypotheses by Swanson (1990), all pairs of association data were finally linked together on a share component and then merged into one. As December 2010, the summary of datasets and the connected associations of ADR-drug-protein that were utilized to assemble the database for ADRs were shown in Table 4.1.

Table 4.1 The summary of datasets and the connected associations of ADR-drug-protein in database for ADRs construction

Data Source	Provided Information	Numbers of Record
MedDRA	Lowest level terms of ADR	68,661
	Preferred terms of ADR	18,919
	System organ classes of ADR	26
HGNC	Information of protein	19,316
DrugBank	Information of drug	4,774
	Drug-target protein association	8,537
Canada Vigilance Adverse Reaction Online Database	ADR-drug association	51,563
PubMed/Taverna	ADR-protein association	16,228
Associations linkage	ADR-drug-protein association	130,446

All datasets in Table 4.1 had been proved for steadiness except ADR-protein associations and the linked ADR-drug-protein associations. These two associations were additionally required to indicate their consistency. Moreover, predictive scores for ADR class of the drug that were originated from data mining were also needed to be justified the correctness. These results were deliberated in further detail below.

4.1.1 ADR-Protein Associations

The associations of ADR and protein were pulled out from literature mining and would be in question about the associations. This controversy would be resolved by focusing on well-studied ADR. Two examples, alcoholic fatty liver and hepatic steatosis, were employed for this task.

Alcoholic fatty liver (steatosis) is an excessive accumulation of triglycerides and other fats inside liver cells which can develop to even high mortal rate disease, hepatocellular carcinoma. There are several molecular mechanisms of alcoholic fatty liver (Bailey, et al., 1999; Kono, et al., 2000; Higuchi, et al., 2001; Pritchard, et al., 2007). Even though cross-talks existed between various factors, Purohit, et al. (2009) summarized major mechanisms of alcoholic fatty liver into three distinct parts: direct ethanol metabolism, fat oxidation by peroxisome, and fat synthesis. The main pathways of alcoholic fatty liver could also occur in nonalcoholic steatohepatitis because common features of fat accumulation and disease progression. The proteins that were found to relate with alcoholic fatty liver and hepatic steatosis from literature mining process were enumerated in Appendix D and were compared to the reviewed pathway by Purohit, et al. (2009). The major depicted pathway was circled for proteins and numbered for pointwise mutual information (PMI) found from literature mining process. Figure 4.1 illustrated proteins and PMI of alcoholic fatty liver and hepatic steatosis in green and purple, respectively. Basically, hepatic steatosis, the broader term, were noticed more proteins than alcoholic fatty liver. Matching proteins of alcoholic fatty liver were five:

ALDH2 (mitochondrial aldehyde dehydrogenase), PGC-1α (peroxisome proliferator activator receptor γ co-activator protein α), PPARα (peroxisome proliferator activator receptor α), TNF-α (tumor necrosis factor alpha), and AMPK (AMP-activated protein kinase). While hepatic steatosis found three matching proteins as same as alcoholic fatty liver, it did not obtain ALDH2, and AMPK. In addition, hepatic steatosis had been observed seven more corresponding proteins: CYP4A (cytochrome P450 4A), SIRT1 (sirtuin1), ACC (acyl-CoA carboxylase), SREBP (sterol regulatory element binding protein), STAT3 (signal transducer and activator of transcription 3), FAS (fatty acid synthase), and CPT-1 (carnitine palmitate transferase-1). Although alcoholic fatty liver and hepatic steatosis found different corresponding proteins, they were specific on their own paths. Alcoholic fatty liver found the matching proteins spatially dispersed throughout the three major parts, while the general term as hepatic steatosis was confined to not be found the matching proteins in the way of direct ethanol metabolism. Moreover, PMI of ADR-protein associations from both alcoholic fatty liver and hepatic steatosis was not up close to zero which indicated about attractive relations between ADR and protein. In conclusion, the process of modified BioAID_ProteinDiscovery workflow in Taverna could be assumed to uncover admirable relations between ADR and protein.

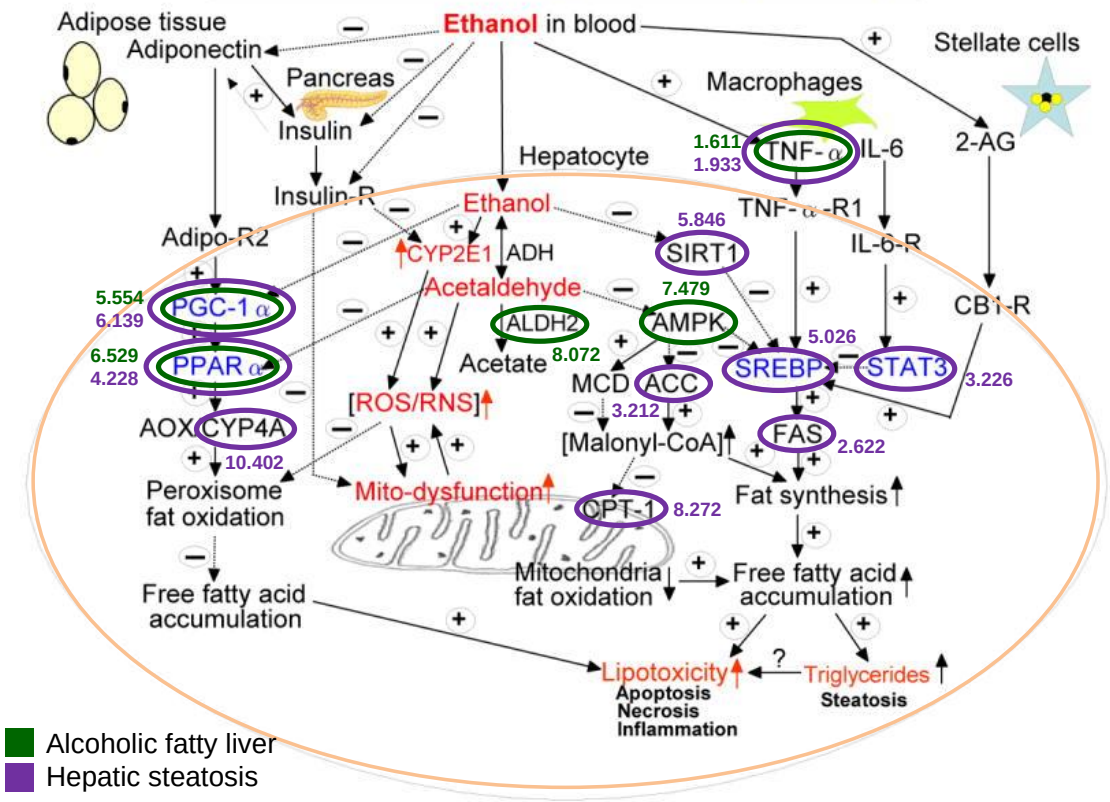

Figure 4.1 The matching proteins of alcoholic fatty liver and hepatic steatosis to the authenticated major pathways, modified from Purohit, et al. (2009)

4.1.2 Information of Drug

In order to evaluate the database performance, a probability score of drugs to ADR class was calculated from nine quantitative structure-properties using data mining. And in the prototype database of this research, only scores of drugs in the ADR class of hepatobiliary disorders had completely examined. Four prediction models were built from Weka using ten folds cross-validation technique and results were shown in Table 4.2. Decision tree gave the highest correctly classified instances, kappa statistic, average true positive rate, average precision, and average recall while it had the smallest value for three error measurements. K-nearest-neighbor formed the second best model by eight metrics and the outstanding mean absolute error. So decision tree and K-nearest-neighbor were two methods for constructing models that were further applied to compute the possible score of drugs to hepatobiliary disorders.

After examining the top two models for predictive score, the results had to be evaluated with experimental data (Ivanciuc, 2008). Unfortunately, the model of decision tree that was the best did not reflect the actuality. It generated drugs that were really mentioned in DrugBank causing hepatobiliary disorders and had predictive score more than fifty percent which was less than that of K-nearest-neighbor model. Furthermore, both decision tree and K-nearest-neighbor models were assessed their predictive scores with acetaminophen, the most common drug that can cause potentially liver damage (Clark, et al., 1973; Yan, 2014). Decision tree model only created predictive score of acetaminophen to hepatobiliary disorders at 9.1 percent. On the other hand, the rival model can generate predictive score at 99.1 percent. K-nearest-neighbor model was finally adopted to make a model for forecasting the possible score of drugs to hepatobiliary disorders because it created satisfied classification model that did not predict the overfit results to experimental data.

With K-nearest-neighbor model, the highest predictive score to hepatobiliary disorders was 99.7 percent and included eight medications: Amodiaquine, Benzyl Benzoate, Diethylcarbamazine, Oxamniquine, Proguanil, Halofantrine, Fumagillin, and Bithionol. Mostly, they were drugs for treatment of blood parasite. Amodiaquine (Kerb, et al., 2009; Srivastava, et al., 2010), Proguanil (Grieshaber, et al., 2005; Jacquerioz and Croft, 2009), and Halofantrine (Watkins and Seeff, 2006; Adaramoye, et al., 2008,) were applied to treat malaria infected liver cells and had the reports about hepatic diseases. While Diethylcarbamazine (Palumbo, 2008) was used for filariasis, Oxamniquine (Utzinger and Keiser, 2004) advanced schistosomiasis. In terms of other medicines, Bithionol (Price, et al., 1993) was a withdrawn anthelmintic medication that used to treat liver flukes and Fumagillin was reported by Zbidah, et al. (2013) to stimulate eryptosis. Benzyl Benzoate (Sparsa, et al., 2006) could also lead to liver disease when it was co-administrated with Ivermectin. According to this information, drugs were checked that they were actually forecasted a possibility to occur hepatobiliary disorders from an appropriate model.

Table 4.2 Results of four prediction models that were built from Weka using ten folds cross-validation technique

Summary	Method			
	Decision Tree	**K-Nearest-Neighbor**	**Neural Network**	**SVM**
Correctly classified instances	64.15%	63.21%	57.55%	50%
Incorrectly classified instances	35.85%	36.79%	42.45%	50%
Kappa statistic	0.283	0.2642	0.1509	0
Mean absolute error	0.4002	0.3705	0.4436	0.5
Root mean squared error	0.544	0.6005	0.6015	0.7071
Average true positive rate	0.642	0.632	0.575	0.5
Average false positive rate	0.358	0.368	0.425	0.5
Average precision	0.643	0.634	0.575	0.5
Average recall	0.642	0.632	0.575	0.5

4.1.3 ADR-Drug-Protein Associations

All pairs of association data were connected on a share element to gain ADR-drug-protein associations and then removed the duplication. These procedures created 130,446 ADR-drug-protein associations which 70,579 were in ADR group of hepatobiliary disorders linked to their relevant biological entities as illustrated the network in Figure 4.2. The consistency of ADR-drug-protein associations was examined with the identical technique of ADR-protein association in topic 4.1.1. Acetaminophen-induced liver necrosis was applied as an example.

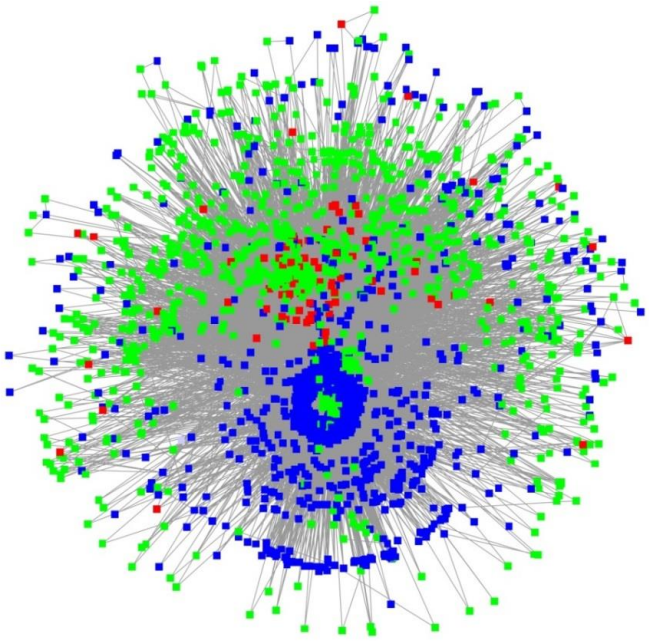

Figure 4.2 Network of ADR-drug-protein associations of hepatobiliary disorders

Acetaminophen (paracetamol or N-acetyl-p-animophenol or APAP) is a commodious-used counter analgesic and antipyretic drug. Although acetaminophen is believed to be safer than other drugs in the same category, it can produce a centrilobular hepatic necrosis that can be lethal. Hinson, et al. (2010) summarized the mechanisms involved in acetaminophen-induced liver necrosis. They mentioned that the mechanistic of acetaminophen-induced hepatic necrosis and the repair system of acetaminophen-induced hepatic necrosis had an important role in development of acetaminophen toxicity. After mapping the generated information of ADR, drug, and proteins to the study of Hinson, et al., the database for ADRs showed that it could produce the appropriate associations of ADR-drug-protein. All of the mechanistic of acetaminophen-induced hepatic necrosis and the repair system of acetaminophen-induced hepatic necrosis were uncovered proteins from database for ADRs, illustrated in Figure 4.3 and Figure 4.4, respectively. In mechanism of acetaminophen-induced hepatic necrosis, three proteins were harmonized. They were iNos (inducible nitric oxide synthase), TNF-α, and IFNγ (interferon gamma). Additionally, the repair mechanism of acetaminophen-induced hepatic necrosis discovered five proteins: TNF-α, STAT3, HGF (hepatocyte growth factor), VEGF (vascular endothelial growth factor), and HMOX (heme oxygenase-1). These findings assumed that database for ADRs could generate the pertinent associations between ADR, drug, and protein.

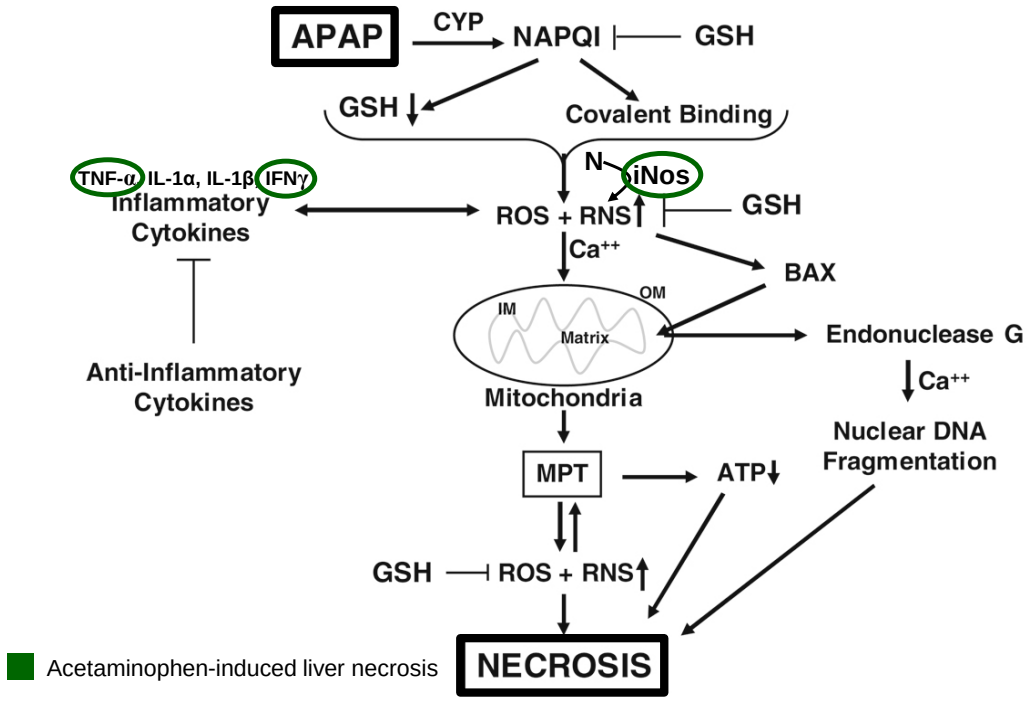

Figure 4.3 The matching proteins of acetaminophen-induced hepatic necrosis to the mechanistic determinants in acetaminophen-induced hepatic necrosis, modified from Hinson, et al. (2010)

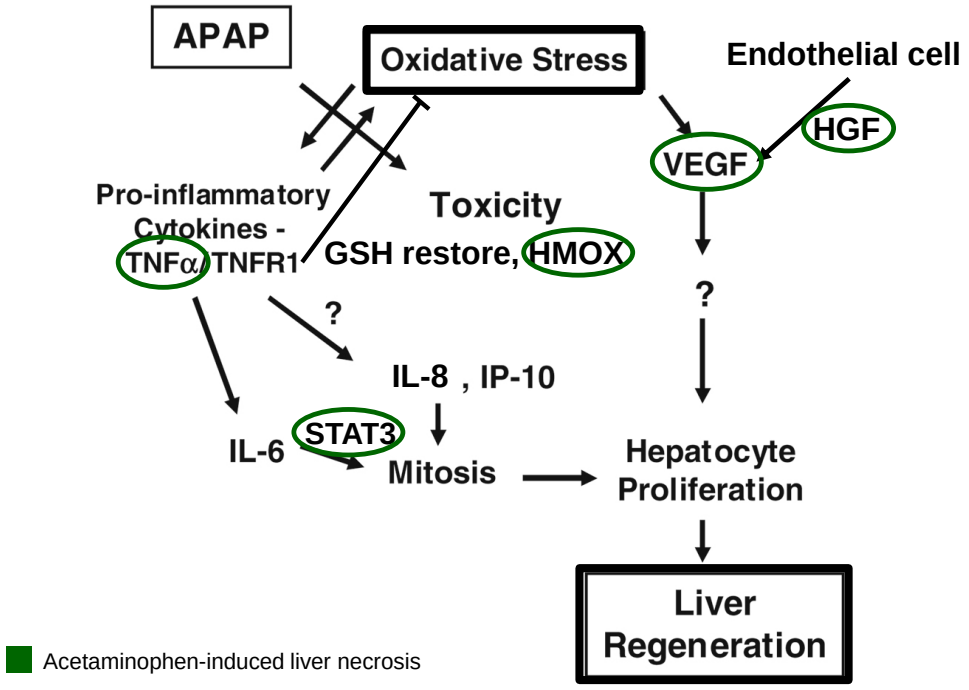

Figure 4.4 The matching proteins of acetaminophen-induced hepatic necrosis to the mechanism in repair of acetaminophen-induced hepatic necrosis, modified from Hinson, et al. (2010)

4.2 Web Interface

The constructed database was displayed via web browser which applied PHP for interacting with the database. The interfaces were designed and written in HTML to the most friendly and understandable to users. The four main pages were home, search, browse, and help. The home page gave the general information of the database. The search page, as illustrated in Figure 4.5, was designed for supporting user query as record number, ADR name, drug name, or protein name. Database for ADRs allowed combinatory search and gave the suggestion to prevent question of users from no result. After applying keyword, the next page repeated user inquiry and showed possible terms that keyword could also be harmonized. This enabled users to select what the closest was according to their requirement and assisted in case users questioned an ambiguous name, as shown in Figure 4.6. After choosing the particular query, the result page was then shown the outcome of user query and interesting related entity. Each result could be expanded to look at the table that contained user question with interesting related entity and remained substance from ADR-drug-protein relationships. Users could perceive more details of each entity by clicking on number or fetched the details of ADR, drug, protein, and reference in tab delimited text file from download function. Furthermore, the result page of database for ADRs provided the link to DAVID for annotating pathway of proteins. This was beneficial to acknowledge users about mechanisms involved in their questions. The result page was illustrated in Figure 4.7. Users also had the ability to examine by the group of each term in browse page, as shown in Figure 4.8, which ADRs were grouped by SOC, drugs were categorized by the

first level of ATC classification, and proteins were systematized as their enzyme reaction. Finally, the help page served the introduction of the project and user guidelines for using the database.

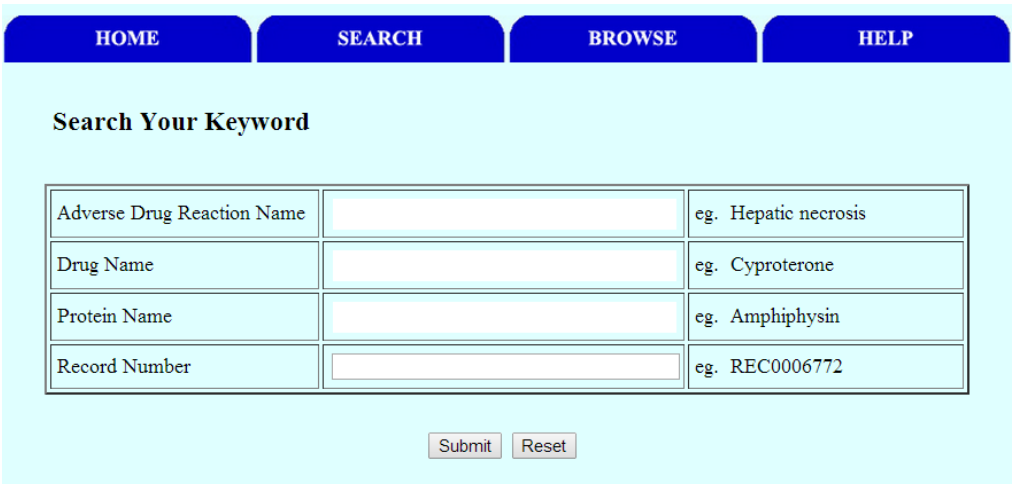

Figure 4.5 Web-based design of search page

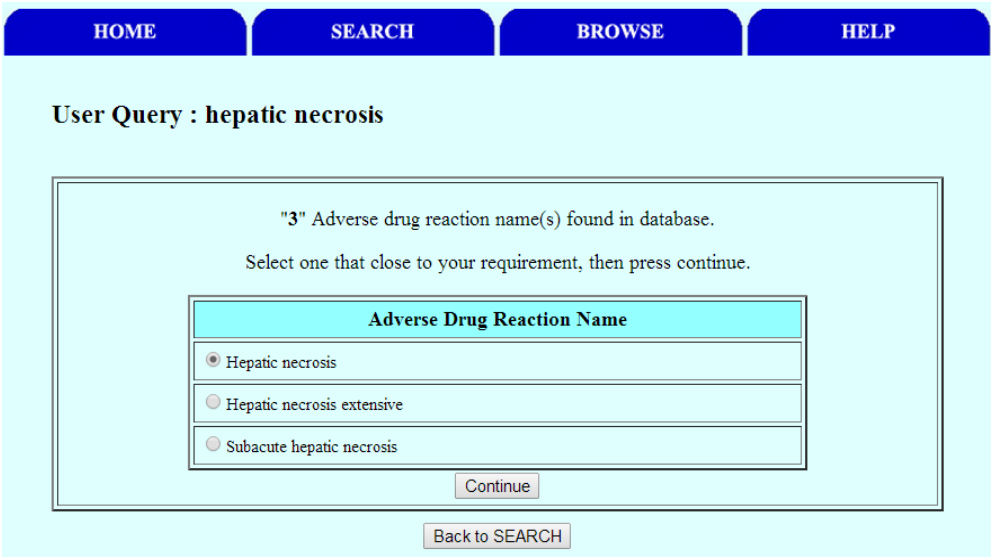

Figure 4.6 Web-based design of page that repeat query from users and show possible word of it in the database

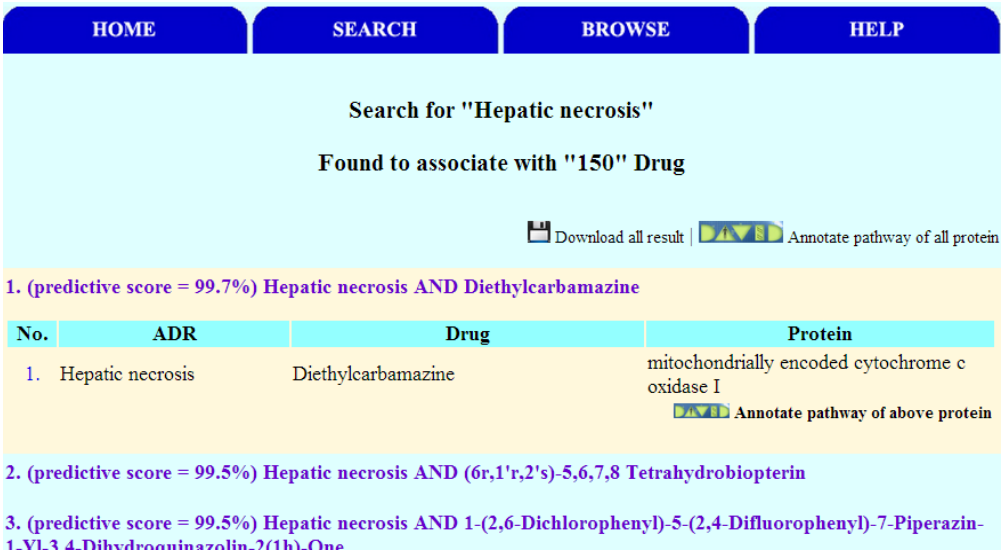

Figure 4.7 Web-based design of result page

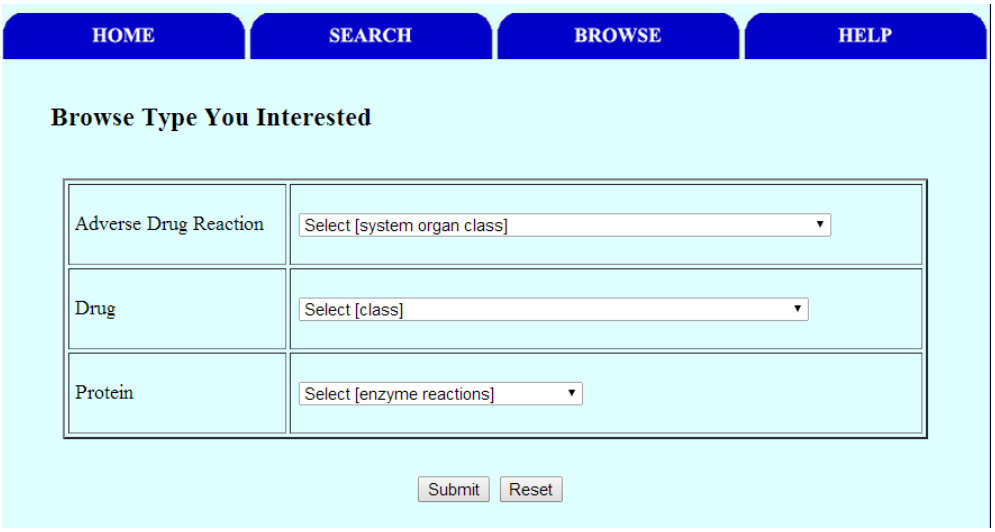

Figure 4.8 Web-based design of browse page

4.3 Application Demonstration

Following the web interface designs of the database for ADRs, there mainly are three groups of user who benefited from utilizing the database on their own purposes. The points of database usage, as previously displayed in Figure 3.8, were examined and described in further detail below.

4.3.1 Drug

From the point of drug, doctors and patients desired to have the ability for realizing what ADR could even possibly cause from their medication. Even though doctors and patients might have their knowledge or had the ability to read from medical package insert, there were not necessary complete or up-to-date. As example when user browsed

in hepatobiliary disorders, Table 4.3 listed the top nine medications that had the highest degree to ADR in this group. These nine drugs were implied that they were found to cause the most hepatobiliary disorders. Simvastatin, a lipid-lowering agent, was discovered that it was the top to cause the substantial hepatobiliary disorders. Simvastatin could induce fifty hepatobiliary disorders such as hepatic fibrosis and hyperbilirubinaemia. Even though Simvastatin was found out to originate more hepatobiliary disorders than other medicines, it had low predictive score at 0.9 percent. This was corrected according to the precautions in WebMD (Boots UK Limited and WebMD UK Limited, 2013) that Simvastatin was informed about rare liver problems. The second to eighth top medications that can cause hepatobiliary disorders were nutritional supplements and had not been stated for hepatobiliary disorders before. This information might beneficial to make the warning for patients who were unwanted persons from an excessive supplementation and avoidance of unintended reactions. Lastly, Drotrecogin alfa (Lai, 2013), the criticized medicine that was used in severe sepsis, was ranked at the ninth drug to cause a lot of hepatobiliary disorders. It activated protein C combines with protein S on platelet surfaces and then degraded factor Va and factor VIIIa, thereby reducing blood coagulability. These might make Drotrecogin alfa to be implicated in, for example, peliosis hepatis which was randomly distributed multiple blood-filled cavities throughout the liver. Although some ADRs that related to drug had not been previously informed, they would be a new knowledge or an infrequent event, such as idiosyncratic reaction, which could be formed drug precaution and interested to be further researched.

Table 4.3 Top nine medications that were found to cause the most hepatobiliary disorders

Generic Name	Number of Hepatobiliary Disorders	Predictive Score (%)
Simvastatin	50	0.9
L-Aspartic Acid	49	99.1
L-Alanine	48	99.5
L-Glutamic Acid	48	99.1
Pyridoxal Phosphate	48	0.9
Glycine	47	99.1
L-Cysteine	47	0.9
Pyruvic acid	47	99.5
Drotrecogin alfa	40	99.1

4.3.2 Protein

Many of databases in pharmacology and the knowledge from literatures provided the information about drugs and target proteins. Neither the study of drug developers nor researchers necessarily completed this information. The linkage between association

data that was applied in this research might expand the knowledge about drugs and target proteins. As the comparison in Table 4.4 and Table 4.5, database for ADRs increased the information about drug and target protein, respectively. In Table 4.4, DrugBank declared that fatty acid desaturase 1 was the target protein of Alpha-linolenic acid and Icosapent. Surprisingly, database for ADRs found Fluconazole, Acetaminophen, Omeprazole, Clozapine, Fluxetine, Methotrexate, Tamoxifen, and Oxybutynin to also be drugs that had fatty acid desaturase 1 as their target protein. These indicated that database for ADRs had the power to suggest eight more medications to have the same target protein than DrugBank. Even though Fludarabine was the only one medication that was found in database for ADRs as a drug for apoptosis regulator Bcl-2, it also expanded the knowledge of four medications from DrugBank. On the other hand, drug promiscuity was recognized by many target proteins (Brown and Okuno, 2012; Hu and Bajorath, 2013). The comparison between DrugBank and database for ADRs when searching for target proteins was displayed in Table 4.5 and also exhibited the increment of knowledge. Both Orlistat and Calcitriol were found an addition target protein from DrugBank. Lipoprotein lipase was remarked to be a target protein of Orlistat, while Cytochrome P450 27B1 was noticed for Calcitriol. Overall, this represented the capability of database for ADRs to advise the possible drugs and target proteins.

Table 4.4 The comparison between DrugBank and database for ADRs to provide the information about medications when searching by protein

Protein	**DrugBank**	**Database for ADRs**	
Fatty acid desaturase 1	Alpha-Linolenic Acid Icosapent	Alpha-Linolenic Acid Fluconazole Omeprazole Fluxetine Tamoxifen	Icosapent Acetaminophen Clozapine Methotrexate Oxybutynin
Apoptosis regulator Bcl-2	Docetaxel Ibuprofen Paclitaxel Rasagiline	Fludarabine	

Table 4.5 The comparison between DrugBank and database for ADRs to provide the information about target proteins when searching by drug

Drug	DrugBank	Database for ADRs
Orlistat	Pancreatic triacylglycerol lipase Gastric triacylglycerol lipase Fatty acid synthase Cytochrome P450 3A4 Cytosolic phospholipase A2	Lipoprotein lipase Fatty acid synthase
Calcitriol	Vitamin D3 receptor (1,25-dihydroxyvitamin D3 receptor) Cytochrome P450 24A1, mitochondrial Cytochrome P450 3A4	Vitamin D3 receptor (1,25-dihydroxyvitamin D3 receptor) Cytochrome P450 27B1

4.3.3 ADR

In the database for ADRs, searching ADR allowed researchers to acknowledge drugs and relevant proteins of ADR which assisted in mechanism studies. These proteins were utilized to map for underlying mechanism using functional annotation tools in Database for Annotation, Visualization and Integrated Discovery (DAVID). This procedure was beneficial to those who examined unknown mechanisms or needed to increase the knowledge of ADRs. To test the efficiency, the comparisons to known system, alcoholic fatty liver and hepatic steatosis, were analyzed in further description.

The related proteins of alcoholic fatty liver and hepatic steatosis from database for ADRs, as listed in Appendix E, were annotated for the pathway using functional annotation tools in DAVID. Unfortunately, proteins that related to alcoholic fatty liver were straggly annotation and did not found general agreement of pathway on each protein. On the other hand, given proteins of hepatic steatosis were accompanied to be mapped on nine pathways from Biocarta and seventeen pathways from KEGG. The lists of mapped pathways from Biocarta and KEGG and annotated proteins were shown in Table 4.6 and Table 4.7, respectively. Because of DAVID might overestimate on some pathways, the annotated pathways from KEGG and Biocarta had to be reconsidered by reading their descriptions even they related to the interested subject (Huang, et al., 2009). After revising all pathways, KEGG remained six pathways while five pathways persisted in Biocarta. The names of related pathway from KEGG were PPAR signaling pathway, Adipocytokine signaling pathway, Arachidonic acid metabolism, Fatty acid metabolism, Steroid hormone biosynthesis, and Linoleic acid metabolism. The names of related pathway from Biocarta were Nuclear receptors in lipid metabolism and toxicity, Mechanism of gene regulation by peroxisome proliferators via PPARα, Reversal of insulin resistance by leptin, Visceral fat deposits and the metabolic syndrome, and Role of PPAR-gamma coactivators in obesity and thermogenesis. All of revised pathways from KEGG and Biocarta could involve in hepatic steatosis mechanism from their responsibilities in lipid metabolism, storage, or elimination. Some pathways were consensus amongst Biocarta and KEGG. First couple was Reversal of insulin resistance

by leptin from Biocarta and Adipocytokine signaling pathway from KEGG which were consistent with insulin response. Another harmonious pathway was PPAR signaling pathway from KEGG that was compatible with Mechanism of gene regulation by peroxisome proliferators via PPARα and Role of PPAR-gamma coactivators in obesity and thermogenesis from Biocarta. All of reconsidered annotated pathways can be relocated to previous described pathways by Purohit, et al. (2009), except Arachidonic acid metabolism, Linoleic acid metabolism, and Steroid hormone biosynthesis. These three pathways from KEGG were so advantages to be researched which could obtain the additional explanations of hepatic steatosis.

Table 4.6 The lists of mapped pathways of hepatic steatosis from Biocarta

Pathway Term	Gene Count	Gene Symbol
Nuclear Receptors in Lipid Metabolism and Toxicity	8	CYP3A4, CYP4A11, PPARA, NR1I2, CYP2C9, PPARG, ABCC3, ABCB1
Hypoxia and p53 in the Cardiovascular system	5	AKT1, HIF1A, BAX, ABCB1, MAPK8
Toll-Like Receptor Pathway	5	PPARA, MYD88, TLR2, MAPK8, TLR4
Mechanism of Gene Regulation by Peroxisome Proliferators via PPARα	5	LPL, PPARA, TNF, FABP1, PPARGC1A
Reversal of Insulin Resistance by Leptin	3	LEP, ACACA, CPT1A
Visceral Fat Deposits and the Metabolic Syndrome	3	LPL, TNF, PPARG
Role of PPAR-gamma Coactivators in Obesity and Thermogenesis	3	LPL, PPARG, PPARGC1A
Antigen Dependent B Cell Activation	3	HLA-DRB1, FAS, IL10
Regulation of transcriptional activity by PML	3	TNF, FAS, SIRT1

Table 4.7 The lists of mapped pathways of hepatic steatosis from KEGG

Pathway Term	Gene Count	Gene Symbol
Neuroactive ligand-receptor interaction	15	DRD1, TACR3, DRD2, DRD4, LEP, HRH1, HTR1A, CHRM3, CHRM2, CHRM1, HRH4, F2, HTR2C, F2R, HTR2A
Pathways in cancer	11	AKT1, HIF1A, PTGS2, BAX, SOS2, PPARG, NKX3-1, IGF1, MAPK8, FAS, STAT3
PPAR signaling pathway	11	LPL, CPT1B, CYP4A11, PPARA, CPT2, CD36, SCD, PPARG, FABP1, CPT1A, ANGPTL4
Adipocytokine signaling pathway	10	LEP, AKT1, CPT1B, PPARA, CD36, TNF, MAPK8, PPARGC1A, STAT3, CPT1A
Toll-like receptor signaling pathway	9	AKT1, TNF, MYD88, IRF7, TLR2, IRF3, MAPK8, TLR4, SPP1
Calcium signaling pathway	9	DRD1, HRH1, TACR3, CHRM3, CHRM2, CHRM1, HTR2C, F2R, HTR2A
Retinol metabolism	8	CYP3A4, CYP4A11, DGAT1, CYP1A1, CYP2C19, CYP2C18, CYP2C9, UGT2B4
Metabolism of xenobiotics by cytochrome P450	8	CYP3A4, GSTA3, CYP1A1, CYP2C19, CYP2C18, CYP2C9, GSTK1, UGT2B4
Arachidonic acid metabolism	8	CYP4A11, PTGS2, CYP2C19, CYP2C18, CYP2C9, PTGS1, EPHX2, GGT1
Drug metabolism	7	CYP3A4, GSTA3, CYP2C19, CYP2C18, CYP2C9, GSTK1, UGT2B4
Jak-STAT signaling pathway	7	LEP, AKT1, IL6ST, SOS2, IL11RA, STAT3, IL10
Apoptosis	7	TNFRSF10A, AKT1, TNF, TNFRSF10B, MYD88, BAX, FAS
Gap junction	5	DRD1, DRD2, SOS2, HTR2C, HTR2A
Fatty acid metabolism	4	CPT1B, CYP4A11, CPT2, CPT1A
Steroid hormone biosynthesis	4	CYP3A4, CYP17A1, CYP1A1, UGT2B4
Linoleic acid metabolism	4	CYP3A4, CYP2C19, CYP2C18, CYP2C9
Allograft rejection	4	TNF, HLA-DRB1, FAS, IL10