BIOSILICO: A Biochemical Database Retrieval System

Jin Sik Kim\textsuperscript{1} \quad Ji Hoon Jun\textsuperscript{4} \quad Yong Wook Kim\textsuperscript{4}
\texttt{jinsikkim@kaist.ac.kr} \quad \texttt{gene2@bioinfomatix.com} \quad \texttt{yongari@bioinfomatix.com}

Sujin Chae\textsuperscript{4} \quad Mira Roh\textsuperscript{4}
\texttt{sujin@bioinfomatix.com} \quad \texttt{mrroh@bioinfomatix.com}

Yong Ho In\textsuperscript{4} \quad Sang Yup Lee\textsuperscript{1,2,3}
\texttt{inyh66@bioinfomatix.com} \quad \texttt{leesy@kaist.ac.kr}

\textsuperscript{1} Deptartment of Chemical & Biomolecular Engineering, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea
\textsuperscript{2} BioProcess Engineering Research Center, Center for Ultramicrochemical Process Systems and Bioinformatics Research Center, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea
\textsuperscript{3} Department of BioSystems, Korea Advanced Institute of Science and Technology, 373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea
\textsuperscript{4} Bioinformatix Inc., 5 floor, Nam Chang bldg, 748-16 Yeoksam-dong, Gangnam-gu, Seoul 135-925, Korea

Keywords: BIOSILICO, integrated database system, metabolism

1 Introduction

In the post genomic era, the number of completely sequenced genomes is increasing continuously and large amount of data are accumulated from various sources. Molecular and comparative analysis of the sequenced genomes enables it convenient to identify enzymes and biochemical reactions involved in the metabolic pathways of organisms. Metabolic pathway is defined as a linked set of biochemical reactions which involves different substrates, products, activators, inhibitors and enzymes. Analyses of metabolic pathways are important for the elucidation of the characteristics of organisms. We have gathered information involved in the metabolic characteristics from source databases. In this paper, we report an integrated database retrieval system with a querying system and visual representation system.

2 Method and Results

Metabolic data in the retrieval system are gathered and continuously updated from the public source databases, ENZYME \cite{1}, LIGAND \cite{2}, MetaCyc \cite{3, 6}, and EcoCyc \cite{4, 5}. At this time, the integrated system contains biological information on 12,200 biochemical reactions, 13,500 compounds and 10,000 enzymes including duplications.

The relations between entities are mainly focused on the information of EC number, biochemical reaction, compound and enzyme. For example, The EcoCyc::reaction, MetaCyc::reaction, LIGAND and ENZYME generate relations among them by using the information of EC number. Additionally, the 3-dimensional images and list of compounds in the Kloth of Biochemical Compound Declarative Database \cite{7} are used in the integration step to generate the structure information of the queried compounds.

The database retrieval system is developed to allow various users to access a lot of information available on the biochemical reactions, compounds, enzymes gathered from heterogeneous sources for the analysis of metabolism. Therefore we have designed the retrieval interfaces to show a well-designed view pages by determining the characteristics of users' queries.
3 System Environments

The Database Retrieval System is installed on the IBM RS/6000 server running the IBM AIX 5.0L as an operating system. BIOSILICO uses two DBMS, the MySQL (version 4.0.12) and the IBM DB2 Universal Database System (version 7.0).

4 Conclusions

BIOSILICO database retrieval system can be utilized as an integrated tool for the analysis of metabolism including various enzymes and biochemical reactions. The system can be more useful as the amount of contents stored in the database increases.

Acknowledgments

This work was supported by the Advanced Backbone IT Technology Development Project (IMT2000-C3-1) of the Ministry of Information and Communication (MIC) and Korean Ministry of Science and Technology (MOST), by the National Research Laboratory Program of the MOST and by a grant from the Korean Ministry of Science and Technology (Korean Systems Biology Research Grant, M10309020000-03B5002-00000). Hardware support by the IBM-SUR program is greatly appreciated.

References