Building Integrated Pathway Genome Database in Cell Type Specific Manner -Encyclopedia of Human Liver

Taejin Ahn tj@bric.postech.ac.kr
YangSuk Kim biostone@bric.postech.ac.kr
SungHwan Kim shkim@bric.postech.ac.kr
DongSun Park dspark@bric.postech.ac.kr
HongGil Nam hgn@bric.postech.ac.kr

Biological Research Information Center, Pohang University, San 37, Hyoja-Dong, Nam-Gu, Pohang, Kyungpook, Korea

Keywords: pathway, semantic network, human liver, knowledge presentation

1 Introduction

The storage and the expression of complex biological information in intelligent manner are important issue of bioinformatics for knowledge-discovery in databases [2]. EcoCYC [3], WIT [4] and KEGG [1] are pioneer databases for intelligent expression of pathway information and the integration of genome data. However, these databases revealed a limitation that they cannot express cell specific information in eukaryote. This may be due to the complexity of the 'gene to protein network' in higher eukaryote. We have developed the system to represent 'gene to protein network' information in a cell specific manner for a higher eukaryote organism using semantic network. A Developed system was applied to the metabolic pathway of glycolysis or gluconeogenesis in the human liver.

2 Method and Results

To represent the 'gene to protein network' information, we define the semantic object and network. The semantic objects are consistent with cell, functional object, isozyme, protein complex, polypeptide, transcript, and gene. Information related to liver glycolysis and gluconeogenesis was gathered from external databases and references. Gathered data were allocated to the proper objects and the relationship between objects were assigned manually. The constructed semantic network could represent protein complexes that are involved in enzymatic reactions and represent their origins (their genes and transcript) whatever those polypeptides were made by alternative splicing or different promoter (Fig. 1). This semantic network has been implemented in JAVA with JDBC and mySQL.

References

[1] Kanehisa, M. and Goto, S., KEGG, *Nucleic Acids Res.*, 28(1):27–30, 2000.

[2] Karp, P.D., Krummenacker, M., Paley, S. and Wagg, J., Integrated pathway-genome databases and their role in drug discovery, *Trends Biotechnol.*, 17(7):275–281, 1999.

[3] Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Paley, S.M. and Pellegrini-Toole, A., The EcoCyc and MetaCyc databases [In Process Citation], *Nucleic Acids Res.*, 28(1):56–59, 2000.

[4] Overbeek, R., Larsen, N., Pusch, G.D., D'Souza, M., Jr, E.S., Kyrpides, N., Fonstein, M., Maltsev, N. and Selkov, E., WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction, *Nucleic Acids Res.*, 28(1):123–125, 2000.

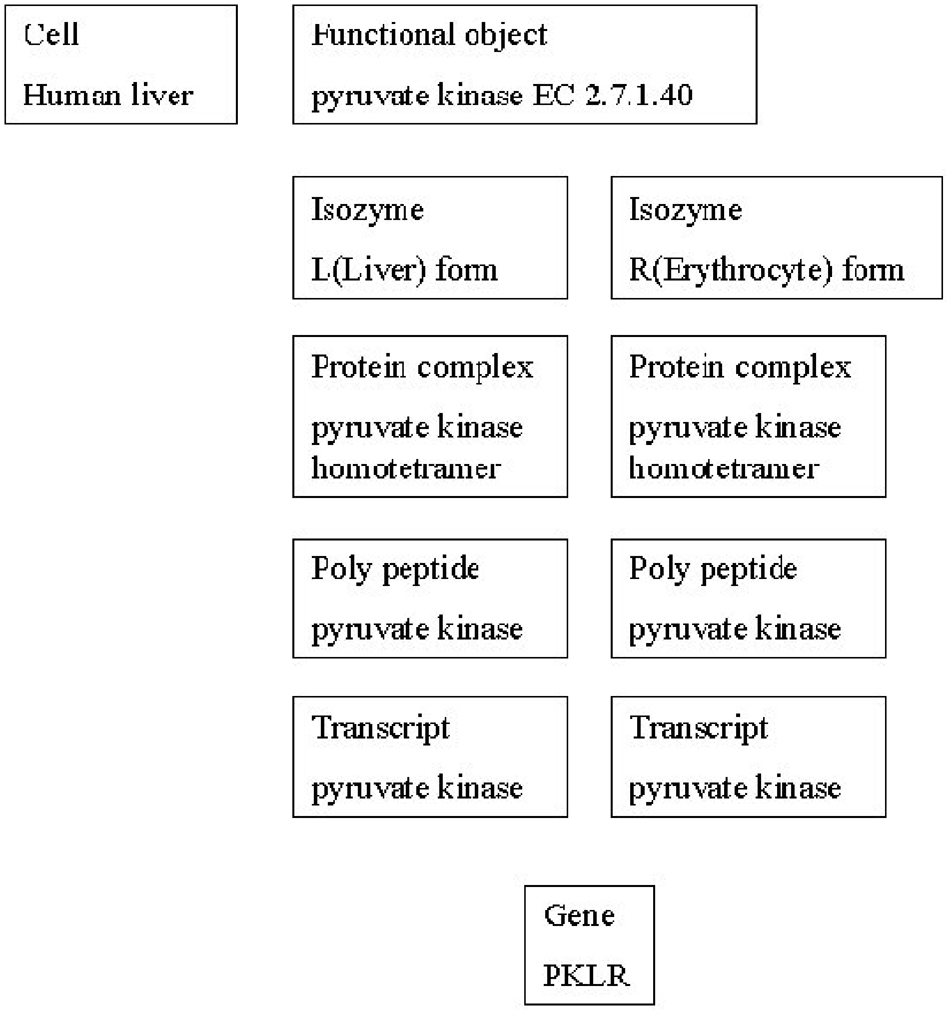

Figure 1: An example of semantic network of pyruvate kinase. There are 4 isozymes of pyruvate kinase in mammals l, r, m1 and m2. L type is major isozyme in the liver, r is found,in red cells, m1 is the main form in muscle, heart and brain, and m2 is found in early fetal tissues. The l- and r-type isozymes are produced from a single gene by use of different promoters (SWISS-PROT).