Metabolic Organization of *Caenorhabditis elegans*

Mahmoud A. Mahdavi[1,2], Yen-Han Lin[1]

[1]Department of Chemical Engineering, University of Saskatchewan, Saskatoon, SK, Canada
[2]Department of Chemical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract

Metabolic bionetwork of *Caenorhabditis elegans* has been reconstructed using genome information retrieved from KEGG (Kyoto Encyclopedia of Genomes and Genes, Release 32) database. The reconstructed bionetwork contains 103 pathways and 22272 open reading frames comprising 21357 protein genes (coding sequences) and 753 RNA genes. As of November 2004, 990 genes are assigned with pathway association, among which, 561 genes are localized in single pathways whereas 429 genes, so called key enzymes, exhibit multiple pathway association. Key enzymes contribute to the interconnecting behaviour of the reconstructed bionetwork and connect different pathways together, termed as 'functionally linked', to accomplish a biological task. These links are perceived through interception of key enzymes in the visual representation of the reconstructed bionetwork. Key enzymes' participation in pathways ranges from 2 to 15 pathways. In order to visualize the interconnectedness of the reconstructed bionetwork Pajek tool was used. Reconstruction process revealed that, 990 known genes perform 2024 distinct biological functions, mostly metabolic functions and partly signalling and regulation. Metabolic pathways are involved in biosynthesis and biodegradation of key compounds such as ATP to provide energy for cell operations.

Keywords: Bionetwork reconstruction, Genome organization, genome database, *Caenorhabditis elegans*, Key enzymes.

Introduction

The sequencing of complete genomes has created the opportunity not only to analyze individual genes but also to consider the whole genome as a system in which different elements work together to accomplish a biological task. With the completion of *C. elegans* genome [1], this organism has become an outstanding platform for investigating biological organization within this system. It has the largest eukaryotic genome that has ever been completely sequenced, with 22272 genes within 9453 protein families [2]. Reconstructing the metabolic bionetwork of such an organism deciphers biological cooperation of pathways at functional level and reveals the relationship between genes and pathways.

To reconstruct a metabolic bionetwork different strategies have been exploited. With the ignorance of the currency metabolites such as ATP, NADH, and etc., Ma *et al.* [3] reconstructed a global metabolic network for 80 organisms of interest, resulting in the different average path length between any pair of metabolites in 3 domains of life: eukaryotes, archaea, and bacteria. They reconstructed the metabolic network using a revised bioreactions information database in which reversible reactions were represented by undirected connections and directed connections corresponded to irreversible reactions. Famili *et al.* [4] proposed a systemic analysis of genome-scale biochemical conversion properties using singular value decomposition, aiming at comparing overall properties of genome-specific metabolic bionetworks. A three-step procedure was implemented to reconstruct prokaryotic metabolic bionetworks that includes gathering a list of metabolic genes, assigning reactions to the genes, and adding physiological information about the organism to the record related to each gene [5]. In an another report, Sun *et al.* [6] used the same map reconstruction strategy that Ma *et al.* [3] employed along with a modified method to prepare their data set which consists of simultaneous gene finding from genome database and gene annotation. After these two parallel processes, the map reconstruction was performed. Miyake *et al.* [7] proposed a graph analysis method to identify the metabolic sub-networks or building blocks of metabolic bionetworks. They used compound-reaction relations as the dataset. This data set has been searched for highly conserved sequential reactions to identify sub-networks. In the present study the bionetwork of *C. elegans* was reconstructed from genome information retrieved from KEGG [8]. Currently different metabolic databases are available to explore gene expression data in the context of metabolic processes [9-11], however, KEGG is one of the databases that contains pathway association which is essential to this study.

Method

The first step toward the bionetwork reconstruction is to retrieve information relating to *C. elegans* from KEGG and save to a local computer, including pathways (stored at 'cel.html' file), bioreactions (stored at 'reaction' file), and genes (stored at 'c.elegans.ent' file). These three files contain pathway numbers and the descriptions, all reactions carry out in the pathways, and gene entries (ORF names) along with their nucleotide sequences and the amino acids sequences of encoded proteins.

Three perl scripts were developed and used to extract relevant information from three files. Then we automatically integrated the informa-

tion and categorized enzymes into each particular pathway. Also for each enzyme, only the reaction which is catalyzed in the pathway of interest is reported and other reactions which may be catalyzed by this enzyme are excluded form this record of results. Therefore, different levels of information including pathway, genes, proteins (enzymes), and reactions are integrated to reconstruct the metabolic bionetwork of the genome of interest. Figure 1 illustrates the flowchart of reconstruction process.

Results and discussion

Analysis of the reconstructed bionetwork shows, there are currently 103 pathways specified in *C. elegans* genome, and are generally categorized into three major pathway groups: metabolism (94), regulation (8), and signalling transduction (1). There are 22272 open reading frames in the genome of *C. elegans*, 990 of them have been associated with pathways and the rest of them are waiting for further investigation. These 990 genes perform 2024 biological functions across genome. Of 990 pathway associated genes, 561 genes are found in single pathways, whereas 429 genes have multiple pathways association, ranging from 2 pathways to 15 pathways. These genes, so-called key enzymes, contribute to the interconnecting behaviour of the bionetwork.

Pajek [12] tool was used to visualize the bionetwork in terms of connections between pathways and the role of key enzymes in pathway connectedness. Figure 2 shows the whole reconstructed metabolic bionetwork of *C. elegans*. In this visualized bionetwork there are 103 pathways (red circles for metabolic pathways and blue circles for regulatory pathways) and 990 genes (green circles) in two sets of vertices. Each line between a pathway and a gene represents the relationship between those two elements.

Key enzymes are elements by which two or more pathways are connected together. These pathways are called functionally linked pathways and create paths in the bionetwork. Pathways participating in each path work co-ordinately to accomplish a biological process. Figure 3 illustrates a typical path in the bionetwork. In this path, six pathways are involved and several key enzymes connect two consecutive pathways together. The six pathways are as follows: DNA polymerase, purine metabolism, pyruvate metabolism, TCA cycle, oxidative phosphorylation, and ATP synthesis. These pathways work together to provide sufficient ATP to replicate DNA molecules.

Conclusion

Metabolic bionetwork reconstruction of *C. elegans* genome as a multi-cellular eukaryotic organism reveals that 990 genes out of 22272 ORFs are localized in 103 metabolic, regulatory, and signalling pathways. The visualized bionetwork depicts the interconnecting nature of this bionetwork which is caused by the participation of several enzymes in more than one pathway (key enzymes). Key enzymes create paths in the bionetwork with physical meaning. Each path consists of several pathways, whose cooperation base on key enzymes establishes a biological process within a living cell.

References

[1] The *C. elegans* sequencing consortium, “Genome sequence of the nematode C. elegans: A platform for investigating biology”, Science, 282, 2012-2018, 1998

[2] Robin G.M., *et al.*, "Comparative genomics of the eukaryotes", Science, 287, 2204-2215, 2000

[3] Ma H., Zeng A.P., "Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms", Bioinformatics, 19 (2003) 270-277.

[4] Famili I., Palsson B.O., "Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices", Journal of Theoretical Biology, 224 (2003) 87-96

[5] Covert M.W., Schilling C.H., Famili I., Edwards J.S., Goryanin I.I., Selkov E., palsson B.O., "Metabolic modeling of microbial strains *in silico*" Trends Biochem. Sci. 26 (2001), 179-186.

[6] Sun J., Zeng A.P., "IdentiCS- Identification of coding sequence and *in silico* reconstruction of the metabolic network directly from unannotated low-coverage bacterial genome sequence", BMC Bioinformatics, 5 (2004) 112-124.

[7] Miyake S., Takenaka Y., Matsuda H., "A graph analysis method to detect metabolic sub-networks based on phylogenetic profile" Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference, August 16-19 (2004) Stanford, CA, pp. 634-635.

[8] Kanehisa M., Goto S., "KEGG: Kyoto encyclopedia of genes and genomes" Nucleic Acid Res. 28 (2000) 27-30.

[9] Hodges P.E., McKee A.H., Davis B.P., Payne W.E., Garrels J.I., "The yeast proteome database (YPD): a model for the organization and presentation of genome-wide functional data" Nucleic Acids Res. 27 (1999) 69-73.

[10] Karp P.D., Riley M., Saier M., Paulsen I.T., Paley S.M., Pellegrini-Toole A., "The EcoCyc and MetaCyc databases" Nucleic Acids Res. 28 (2000) 56-59.

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

[12] Batagelj V., Mrvar A., "Pajek-program for large network analysis" Connections 21 (1998) 47-57.

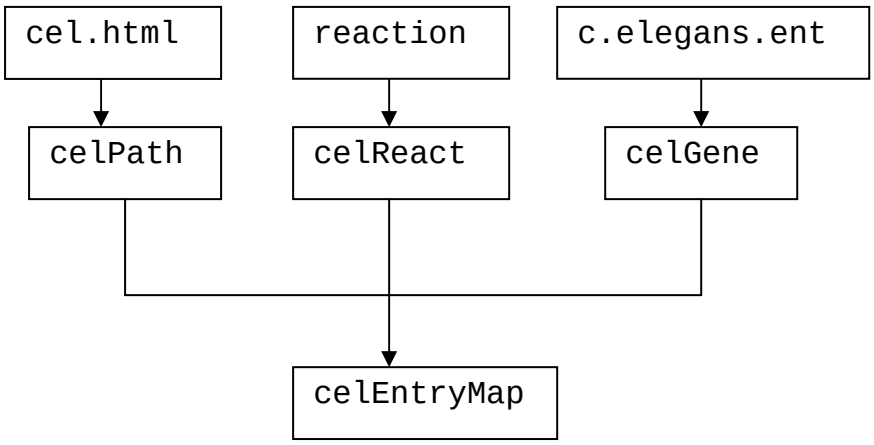

Figure 1 *The flowchart for the reconstruction of metabolic bionetwork of C. elegans. In this flowchart three KEGG reference files: cel.html, reaction, and c.elegans.ent, were used as input data for three perl scripts. The output from these scripts were used as input data for perl script celEntryMap to reconstruct the bionetwork.*

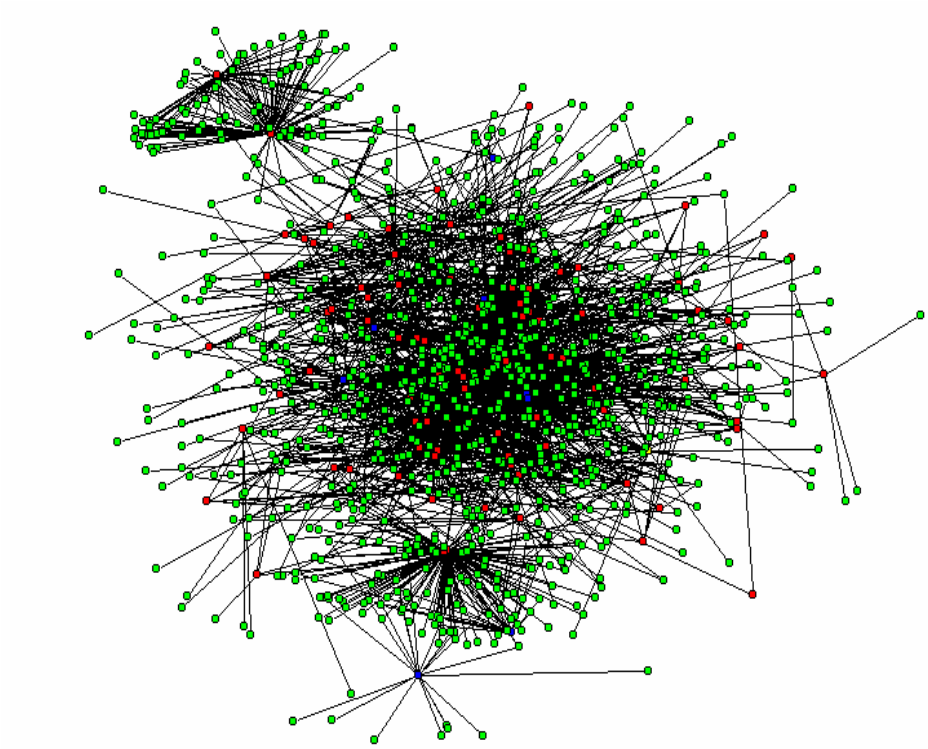

Figure 2. *The visualized reconstructed bionetwork. Red, blue, and yellow circles represent metabolic, regulatory, and signalling pathways, respectively, and green circles represent genes.*

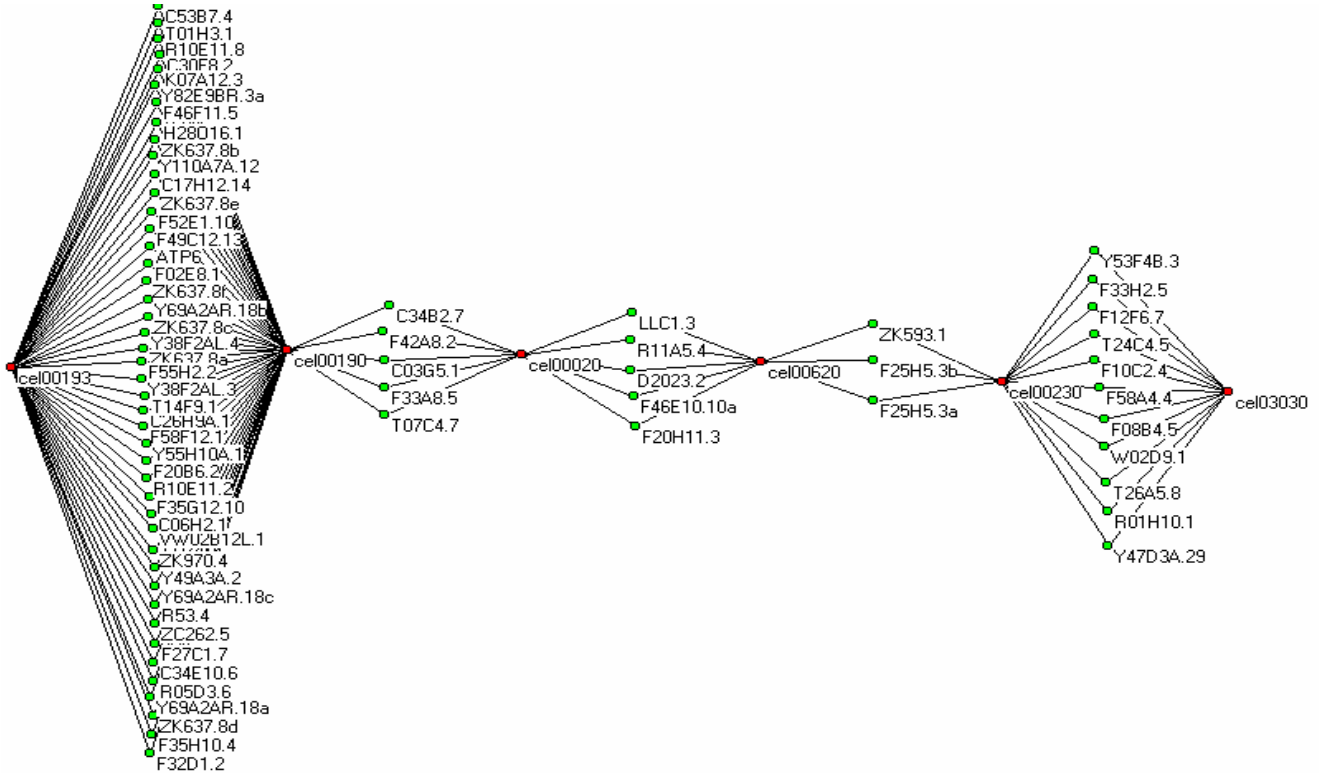

Figure 3. *A typical path, taken from the whole reconstructed bionetwork in figure 2. In this path 6 pathways are involved and several key enzymes between each pair of pathways connect them together. These 6 pathways work co-ordinately to provide energy to replicate DNA molecules.*