

## Chapter 53

# Chemoinformatics on Metabolic Pathways: Attaching Biochemical Information on Putative Enzymatic Reactions

**Masahiro Hattori**

*Tokyo University of Technology, Japan*

**Masaaki Kotera**

*Kyoto University, Japan*

### ABSTRACT

*Chemical genomics is one of the cutting-edge research areas in the post-genomic era, which requires a sophisticated integration of heterogeneous information, i.e., genomic and chemical information. Enzymes play key roles for dynamic behavior of living organisms, linking information in the chemical space and genomic space. In this chapter, the authors report our recent efforts in this area, including the development of a similarity measure between two chemical compounds, a prediction system of a plausible enzyme for a given substrate and product pair, and two different approaches to predict the fate of a given compound in a metabolic pathway. General problems and possible future directions are also discussed, in hope to attract more activities from many researchers in this research area.*

### INTRODUCTION

A cellular system is composed of a considerable number of elements including genes, proteins or enzymes, and metabolites. They are dynamically connected and interdependent to construct a network diagram, which is referred to as a metabolic

pathway. The significance of a systems approach has been recognized for understanding the activities of living organisms (Eisenberg *et al.*, 2000; Kanehisa, 2001; Kanehisa & Bork, 2003).

Among the elements mentioned above, genetic information has been studied from a holistic viewpoint since the late 1990s. Whole genome sequencing projects have uncovered the genomic aspects of biological substances, *i.e.*, the reper-

DOI: 10.4018/978-1-4666-3604-0.ch053

toire of genes and their products encoded on each genome, generating an exhaustive encyclopedia containing macromolecules such as DNA, RNA and proteins. The information of molecular interactions like metabolic pathways and regulatory systems has also been compiled from text-based literature and stored in several computer databases including KEGG (Kanehisa & Goto, 2000; Kanehisa *et al.*, 2006 & 2008), EcoCyc (Keseler *et al.*, 2009), MetaCyc (Caspi *et al.*, 2008), and BioCarta (BioCarta LLC). Motivation of the early genome analyses was mainly to obtain the whole set of fundamental molecules in biology, hoping that the interaction network of them would be depicted by collecting the text information on pathways. Initially, pathway information was utilized to improve genome annotation in order to add in-depth information for macromolecules (Kanehisa, 1997).

As we gain more omic-scale information, *e.g.*, proteome (Wilkins *et al.*, 1996), transcriptome (Velculescu *et al.*, 1997), metabolome (Tweeddale *et al.*, 1998), and interactome (Sanchez *et al.*, 1999), studies based on those holistic viewpoints has become applicable to the other substances in living organisms. As a result, the emerging main stream has become what is referred to as chemical genomics, integrating heterogeneous information including genomic, chemical and interaction data (Yamanishi *et al.*, 2004). It has been shown that a sophisticated method combined with both genomic and chemical knowledge is also applicable to computationally predict missing enzymes in particular species in a given metabolic pathway (Yamanishi *et al.*, 2007).

Attaching biological information on a given genome sequence is referred to as genome annotation, which is mostly based on the assumption that if the sequences of genes are similar enough, then their functions would be identical. In the case of annotating enzyme genes, the function is generally inferred by a previously known enzyme reaction. The missing-enzyme finding problem mentioned above suggests putative genes on pre-defined

enzyme reactions, which can also be described as one of the genome annotation approaches.

On the contrary, chemoinformatics has the potential to provide a reverse method of annotation, which we refer to as a “chemical annotation” approach. This is based on the assumption that when an enzyme reaction is found to be similar to a previously known reaction, then we can gain a clue what the enzyme or enzyme gene could be. The topic can be divided into two parts: how to find pairs of metabolites that are connected in metabolic pathways, and what kind of intermediate enzymes would catalyze those reactions. These are the main themes of this chapter.

Here we report our recent efforts concerning this research area. First, we have defined a similarity measure of chemical substances for metabolic pathway analysis, and elucidated the strong correlation among chemical and genomic structures *i.e.*, metabolic pathways and operons (Hattori *et al.*, 2003). As an application of this, we have developed a prediction system of enzymes for a given partial reaction equation (Kotera *et al.*, 2004). The next achievement was the prediction scheme of biodegradation pathways for a given set of xenobiotics of which molecular fate is unknown (Oh *et al.*, 2007). This type of research will be fundamental in predicting the bioremediation capability of environment chemicals by a bacterial community. The most recent work is the GREP system, where possible enzyme reaction equations are generated from the given set of compounds (Kotera *et al.*, 2008).

## EC NUMBER CLASSIFICATION

The topics presented are currently performed through the Enzyme Committee (EC) number classification, with which reaction mechanisms and, more importantly, the corresponding enzyme genes in the genome could be deduced. The EC number classification has been developed by the Enzyme Committee in IUBMB to classify all

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/chemoinformatics-metabolic-pathways/76106](http://www.igi-global.com/chapter/chemoinformatics-metabolic-pathways/76106)

## Related Content

---

### Key Issues and Challenges for Managing and Evaluating B2B E-Commerce Projects within the Australian Pharmaceutical Supply Chain

Chad Linand Geoffrey Jalleh (2012). *Pharmacoinformatics and Drug Discovery Technologies: Theories and Applications* (pp. 54-71).

[www.irma-international.org/chapter/key-issues-challenges-managing-evaluating/64066](http://www.irma-international.org/chapter/key-issues-challenges-managing-evaluating/64066)

### Classifier Ensemble Methods for Diagnosing COPD from Volatile Organic Compounds in Exhaled Air

Ludmila Ilieva Kuncheva, Juan Jose Rodríguez, Yasir Iftikhar Syed, Christopher O. Phillips and Keir Edward Lewis (2012). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 1-15).

[www.irma-international.org/article/classifier-ensemble-methods-diagnosing-copd/77927](http://www.irma-international.org/article/classifier-ensemble-methods-diagnosing-copd/77927)

### Identification of Distinguishing Motifs

Wangsen Feng and Lusheng Wang (2010). *International Journal of Knowledge Discovery in Bioinformatics* (pp. 53-67).

[www.irma-international.org/article/identification-distinguishing-motifs/47096](http://www.irma-international.org/article/identification-distinguishing-motifs/47096)

### Indonesian Legal Perspectives on Biotechnology and Intellectual Property Rights

Theofransus Litaay, Dyah Hapsari Prananingrum and Yakub Adi Krisanto (2011). *Genomics and Bioethics: Interdisciplinary Perspectives, Technologies and Advancements* (pp. 171-183).

[www.irma-international.org/chapter/indonesian-legal-perspectives-biotechnology-intellectual/47300](http://www.irma-international.org/chapter/indonesian-legal-perspectives-biotechnology-intellectual/47300)

### Structure of Hydrogenase in Biohydrogen Production Anaerobic Bacteria

Ming Du and Lu Zhang (2011). *Interdisciplinary Research and Applications in Bioinformatics, Computational Biology, and Environmental Sciences* (pp. 251-258).

[www.irma-international.org/chapter/structure-hydrogenase-biohydrogen-production-anaerobic/48382](http://www.irma-international.org/chapter/structure-hydrogenase-biohydrogen-production-anaerobic/48382)