Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biomedical Informatics 39 (2006) 147–159
Journal of Biomedical Informatics
www.elsevier.com/locate/yjbin

A medical bioinformatics approach for metabolic disorders: Biomedical data prediction, modeling, and systematic analysis

Ming Chen [a,b,*], Ralf Hofestädt [b,*,1]

[a] Group of Bioinformatics, College of Life Science, Zhejiang University, Hangzhou 310029, China
[b] Department of Bioinformatics/Medical Informatics, Faculty of Technology, Bielefeld University, Postfach 10 01 31, D-33501 Bielefeld, Germany

Received 24 February 2005
Available online 23 June 2005

Abstract

During the past century, studies of metabolic disorders have focused research efforts to improve clinical diagnosis and management, to illuminate metabolic mechanisms, and to find effective treatments. The availability of human genome sequences and transcriptomic, proteomic, and metabolomic data provides us with a challenging opportunity to develop computational approaches for systematic analysis of metabolic disorders. In this paper, we present a strategy of bioinformatics analysis to exploit the current data available both on genomic and metabolic levels and integrate these at novel levels of understanding of metabolic disorders. PathAligner is applied to predict biomedical data based on a given disorder. A case study on urea cycle disorders is demonstrated. A Petri net model is constructed to estimate the regulation both on genomic and metabolic levels. We also analyze the transcription factors, signaling pathways and associated disorders to interpret the occurrence and regulation of the urea cycle.

Availability. PathAligner's metabolic disorder analyzer is available at http://bibiserv.techfak.uni-bielefeld.de/pathaligner/pathaligner_MDA.html. Supplementary materials are available at http://www.techfak.uni-bielefeld.de/~mchen/metabolic_disorders.

Keywords: Metabolic disorders; Medical bioinformatics; Systems biology; Integrative bioinformatics; Urea cycle disorders; Inborn errors; PathAligner

1. Introduction

Metabolism is defined as the enzyme-catalyzed chemical reactions that breakdown (catabolism) and synthesize (anabolism) the molecules needed for life [1]. Enzymes are important for metabolism because they act as catalysts in making one chemical from another [2]. Metabolic processes result in growth, produce energy, eliminate wastes, and control other body functions which distribute nutrients in the blood after food is digested. Metabolism maintains homeostasis, or a steady state, in the body. Any problem in the body that causes loss of metabolic control of the body's steady state will lead to a metabolic disorder [3]. Most metabolic disorders are inherited as autosomal recessive traits. Autosomal recessive inheritance and enzyme deficiency are features typical for an inborn error of metabolism.

Inborn errors of metabolism are characterized by a block in a metabolic pathway, a deficiency of a transport protein or a defect in a storage mechanism caused by a gene defect. The defective gene leads up to an absent or wrong production of essential proteins, especially enzymes. But these enzymes are important components of the biochemical processes in cells and tissues. They enable, disable or catalyze the biochemical reactions of metabolic pathways. Thus, these disorders of the metab-

* Corresponding authors. Fax: +49 0 521 106 6488.
E-mail address: mchen@techfak.uni-bielefeld.de (M. Chen).
[1] Present address: Bielefeld University, Faculty of Technology, Department of Bioinformatics and Medical Informatics, Universitätsstraße 25 D-33615 Bielefeld, Germany.

1532-0464/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2005.05.005

olism result in a threatening deficiency or accumulation of intermediate metabolites in human.

The field of inherited metabolic disorders has undergone major revolutions in the past century. In earlier days, discoveries in physics and chemistry applied to pathology and clinical medicine led to knowledge of the organic chemistry of dyes, tissues staining, and improved microscopy. Defects that lead to the accumulation of metabolic products were identified in the mid 1960s. If a patient is suspected of having an inborn error of metabolism, specialized biochemical laboratories analyze enzyme activities in specimen of different tissues (skin, liver, etc.) and investigate body liquids as blood, urine, neural tissues, etc., for unusual metabolic pattern. The identification of the metabolites that accumulated in a disease made possible the identification of the enzyme whose activity was deficient. By the mid 1980s, techniques largely had switched from those of the biochemistry of intermediates and enzymes to the identification of mutations in genes. This was done by a large number of techniques that make use of DNA fragments (restriction fragment-length polymorphisms) so as to permit linkage mapping and gene sequencing. As a result, we now know many genetic defects responsible for neurological disease, but frequently we do not know much about the resulting protein product and therefore the pathophysiologic basis for the disease [4].

The past decade witnesses the rapid development of the modern biomedical technology and the information technology. Data from the Human Genome Project surely will be useful in identifying mutations in the thousands of genes that must underlie inherited diseases. Genetic data also will be useful in identifying mutations and polymorphisms that predispose to some of the acquired diseases of the metabolic system. Now we have more ''omics,'' such as transcripomics, proteomics, and metabolomics, which bring together relatively dynamic data (DNA, protein, and other molecules) to study metabolism. The requirements to better understand the complex biomedical systems involve systematically identifying how genes function and how their products interact with other molecules, and how diseases occur, within the context of metabolic processes and their assumed roles. This places, new demands for representing and handling new forms of the complex biomedical information, including enzyme retrieval, gene expression, and disease associated metabolic pathway representation, modeling, and simulation. As a consequence of the current knowledge available in biomedical systems, one would expect that bioinformatics techniques would be ideally suited to tackle these problems.

In this paper, we attempt to predict biomedical data for a given disorder and propose a computational strategy for the systems analysis of metabolic disorders. In Section 2, we explore existing biomedical information sources and address a strategy to integrate genomic and metabolic data, and to perform systematic analysis for the study of metabolic disorders. A case study is presented in Section 3. We integrate urea cycle information, construct a Petri net model of urea cycle network, and analyze the transcription factors of urea cycle genes. We also investigate related signaling pathways and associated diseases of urea cycle disorders. Section 4 is a conclusion of our research findings.

2. Methods

In this section, we explore several major information resources of metabolic disorders. Architecture of biomedical information retrieval system and computational strategy of systems analysis are introduced and presented. We explain how they can be organized to predict biomedical data and to perform a systems analysis.

2.1. Exploration of biomedical data

As a result of the Human Genome Project and related clinical efforts, tremendous amount of useful biomedical information is accumulated. Biological and biomedical data have been exploring and systematically storing in hundreds of public databases. A huge number of genes, enzymes, and metabolic pathways have already been identified, isolated, sequenced, and collected in these databases. For example, EMBL [http://www.ebi.ac.uk/embl/] and GenBank [http://www.ncbi.nlm.nih.gov/Genbank/] contain DNA sequences and BioBase's TransFac/TransPath [http://www.biobase.de/] the knowledge about gene expression. Metabolic pathways and their single biochemical reactions are stored in KEGG [http://www.genome.ad.jp/kegg/] and ExPASy [http://www.expasy.org/]. BRENDA [http://www.brenda.uni-koeln.de/] provides the kinetics of enzymatic driven processes. Most inborn errors of metabolism are also included in OMIM [http://www3.ncbi.nlm.nih.gov/Omim]. OMIM is a catalogue of medically important human traits, genes, and disorders thought to have a genetic basis. MEDLINEplus is a premier source of health information for patients, families, and friends. Developed by the United States National Library of Medicine, part of the National Institutes of Health, MEDLINEplus contains Web links to information on over 600 health topics. Other specific databases on inborn errors are: Metagene [http://www.metagene.de/] that is designed to support the diagnosis of inborn errors of metabolism. Ramedis/MD-Cave [http://mdcave.genophen.de/] is a patient database of rare metabolic diseases. It develops a bioinformatics system for representing, modeling, and simulating genetic effects on gene regulation and metabolic processes in human cells.

The amount of this electronically available knowledge of genes, enzymes, metabolic pathways, and meta-

bolic diseases increases rapidly. But they are highly heterogeneous both in structure and in semantics and give only highly specialized views of the biological systems. These lead up to the general task of integrating all this knowledge and make it biotechnologically and medically applicable. Researchers are attempting to find ways to deal with all these exponentially accumulating data and trying to better understand the complex biological and biomedical systems. We try to systematically study the relationship between transcription factors, genes, enzymes, and metabolites involved in metabolic disorders.

We collect and integrate metabolic disorder information mainly from OMIM. Basic metabolic reactions involved in the urea cycle are extracted from KEGG, ExPASy, and BRENDA. These databases are rich in metabolic and enzymatic information. Transcription factors and signaling data are obtained from Biobase. The human promoter sequences of urea cycle related enzymes are provided by UCSC [http://genome.ucsc.edu/]. All other sequences were obtained by screening the GenBank.

2.2. Systems analysis strategy

Differs from traditional treatments that mainly based on knowledge of biochemistry, systems treatments require integrative information that describes the components of the metabolic system and how they interact (dot box in Fig. 1).

It is important to take a look into the genome and proteome for the gene(s) and protein(s) responsible for that disease and then that will suggest a way in which the behavior of the gene(s) and protein(s) can be modified to develop a cure. The problem is that most diseases

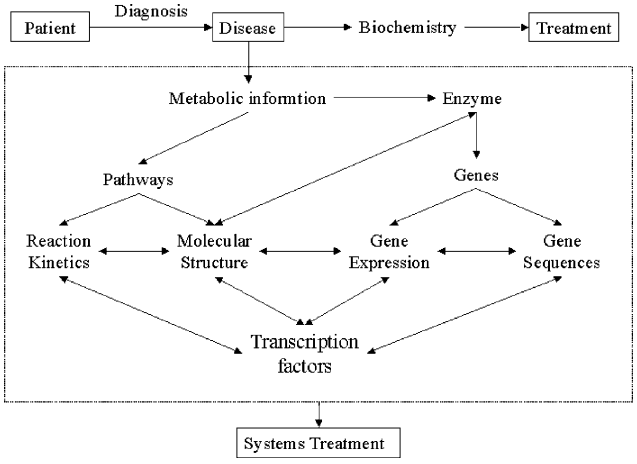

Fig. 1. Biomedical information systems analysis architecture. To gain a precise understanding of these systems, it is important to study the interactions and interrelationships among genomic, metabolic and pharmacological system variables. It differs from the traditional treatment. Systems treatment requires a global knowledge of biomedical system on levels of genome, proteome and metabolome. This paper less addresses diagnosis process and practical treatment, it mainly focuses on the analysis of metabolic disorders.

arise from the interaction between perhaps dozens of genes as well as proteins, in ways that we do not understand yet in many cases. Moreover, metabolic reactions and signal transductions also act critically to many diseases. So that it is not plausible to diagnose disease simply based on clinical appearances, or to predict what the consequences will be of investigating a single gene. Instead, a list of human genes and their protein (or RNA) products and related biological molecules is essential for understanding pathology/pathophysiology/physiology and explaining what goes wrong in disease. Fortunately, more and more of the disease information are stored in widely distributed heterogeneous databases around the world and nowadays are retrievable via the Internet. It requires new methods and approaches for representing and handling the disease information, including enzyme retrieval, gene expression, and disease associated metabolic pathway representation, modeling, and simulation.

Execution of the strategy illustrated in Fig. 1 requires acquisition of the following information: an inventory of the disorders involved and related genes and their SPNs, transcription factors, molecules, metabolic reactions, and signaling events. These values reveal which biological interactions are genetic/cellular relevant and how a cellular process works. These data can provide the key point to model such system and can predict test for genetic/cellular function for each molecule and its role in disease.

2.3. Biomedical information retrieval architecture

The computational architecture proposed in this paper is based on integrating relevant biomedical information sources to provide a systematic analysis of metabolic disorders. This section describes the main characteristics of the proposed framework, besides its general architecture. Frameworks offer an adequate infrastructure to fulfill the requirements of retrieving the biomedical data sources and make them available to the disease associated genomic and metabolic information diagnosis and Petri net modeling. The basic idea of the proposed framework is shown in Fig. 2. The framework may be described through its main functionalities.

As the biomedical data sources are distributed widely in various heterogeneous databases, information retrieval requires data integration and data mining expertise. There are several previous and underlying projects that are trying to address the challenging problem of interoperability among biological databases. They are based on different data integration techniques, e.g., federated database systems (ISYS and DiscoveryLink), multi database systems (TAMBIS), and data warehouses (NCBI/Entrez and SRS). Existing systems have addressed heterogeneous database integration in the realms of molecular biology, hospital information

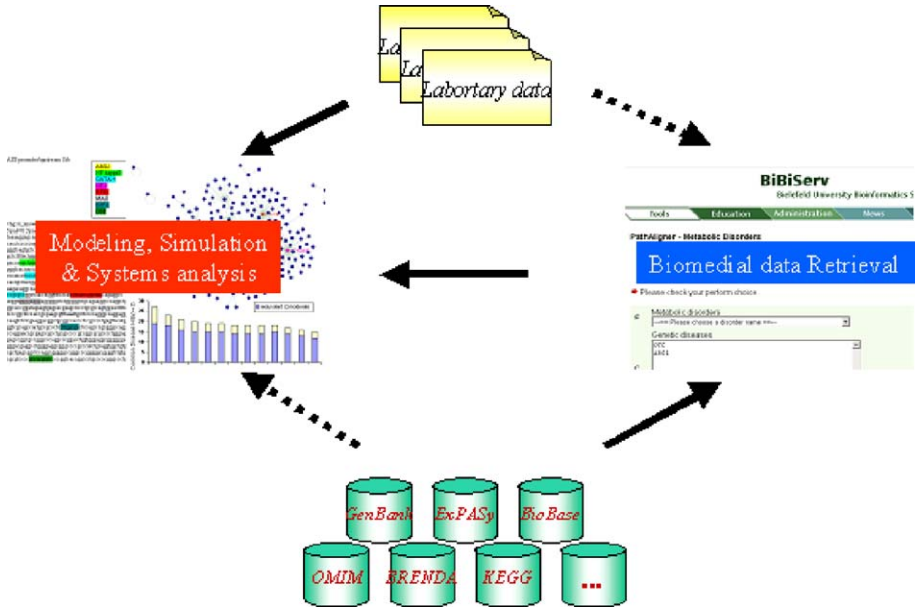

Fig. 2. Framework of biomedical information retrieval and systems analysis. Solid lines indicate real workflow of this research. Dashed lines demand users' extra manual efforts.

systems, and application portability [5]. Mork et al. [6] presented a general model for data integration systems using a mediated schema to represent commonalities in the underlying sources, focused their efforts on modeling online genetic databases, although in principle the approach is flexible enough to accommodate other sources of biomedical data (e.g., multiple medical record systems, multiple indexed reference sources, etc.). We use PathAligner [7] to retrieve/predict biomedical information based on a given disorder. The PathAligner is a web-based biological information retrieval framework designed for the main purpose of functional analysis and representation of the metabolic pathways. It is built on several major biological/biomedical database systems such as BRENDA, KEGG, BioBase, MEDLINEplus, ExPASy, etc., that are located in the Internet. The web-based system allow different distributed biological/biomedical data sources to communicate through the same common interface and enable client clinicians, researchers, and educators to perform efficient and effective disease online functional diagnose and information navigation. It can handle disease-associated enzymes, proteins, metabolites, transfactors as well as genomic and metabolic networks. Users access PathAligner via the web interface, while standard, platform independent Perl applications and modules are used to connect applications to the external databases and retrieve the metabolic information of the diseases.

2.4. Biomedical data modeling and simulation

Modeling and simulation is important to explore the biomedical information to understand physiology and the molecular mechanisms of disease. As the biomedical system is an information-intensive and mechanism-oriented field, it is very difficult to model and simulate. The availability of high performance computers, coupled with mathematical modeling, however, has contributed to the development of increasingly accurate models of biomedical systems, which makes it possible to identify how genes function and how their products interact with other molecules, how diseases occur, and how molecular function changes medical practice, within the context of metabolic processes.

A number of attempts have been made regarding to the modeling methodology and software. Nebot employed a qualitative modeling and simulation methodology called Fuzzy inductive reasoning (FIR) that would make optimal use of the limited knowledge available to the modeler [8]. Thomaseth developed the tool PANSYM for biomedical system modeling which had been designed primarily for representation of kinetics, transport, and metabolism of biological substances [9]. In this paper, the Petri net methodology is exploited.

Petri nets were developed originally by C.A.Petri in 1962. Because of their good properties in theoretical analysis, practical modeling, and graphical visualization of concurrent systems, Petri nets, especially high-level Petri nets, are widely applied in various fields, and even biological systems. A Petri net is a formal model that is used to model concurrent systems. It is represented by a directed, bipartite graph in which vertices are either places or transitions, signifying conditions and processes, respectively. Tokens are placed on places to indicate that those conditions are true. A Petri net can be executed in the following way. When all the places with arcs to

a transition have a token, the transition is enabled, and may fire, by removing a token from each input place and adding a token to each place pointed to by the transition. High-level Petri Nets include extensions that allow temporal and hierarchical input [10]. Readers with little knowledge about Petri nets are suggested to refer the "Petri Nets World" web page at http://www.daimi.au.dk/PetriNets/. The first application of ordinary Petri net on modeling metabolic pathways was introduced by Reddy et al. [11]. Nevertheless, ordinary Petri nets can only be used to examine discrete systems. They are not qualified for modeling systems with continuously changing state variables. Therefore, they fail on modeling metabolism when real concentrations and kinetic effect are concerned. During the last years some more approaches of extended Petri nets and their application on molecular biological systems appeared [12–17]. These high level Petri nets can support dynamic change, task migration, superimposition of various levels of activities, and the notion of mode of operations. Provided with such powerful Petri nets and computer techniques, data of metabolic pathways, gene regulation, signalling pathways and their kinetic parameters can be converted for Petri net destination application; a Petri net model of the virtual cell can be implemented and the attempting to understand the behaviors of cell activity could be accomplished.

In the medical field, Petri net methodology has been also applied to model neurological networks [18–20], medicine [21], medical information system [22], medical diagnostic system [23], human visual canal [24], and blood coagulation cascade [25]. However, these previous achievements are more or less unrelated to the biomedical information system which not only requires simulations of single metabolic pathways but also the longer time-scale effects of processes such as gene regulation, cell division cycle, and signal transduction, drug effect, etc. Due to the available features of Petri nets such as inhibitor arc, timed transition, and mathematical presentation, more new areas concerning biomedical information may also be possibly modeled and simulated with Petri Nets. The hybrid Petri net methodology is used for mixed quantitative/qualitative modeling and simulation. It allows the modeling of metabolic pathways using actual concentrations, and makes sense to model the biological processes using functions, which allow each transition to simulate and can be realized by using functions for specifying the arrow weight.

3. Case study

The following sections present a case study of systems analysis of metabolic disorders. We choose urea cycle disorders due to the availability of such data and the untouchability of their systems analysis. Urea cycle disorders are typical of many other metabolic diseases. They are estimated to occur in 1 in 30,000 live birth.

3.1. Urea cycle disorders

In human cells, excess nitrogen is removed either by excretion of NH_4^+ (of which only a little happens) or by excretion of urea. Urea is largely made in the liver by the urea cycle, a series of biochemical reactions that are distributed between the mitochondrial matrix and the cytosol. The cycle centers around the formation of carbamoyl phosphate in hepatocyte mitochondria to pick up NH_4^+ and incorporate it into ornithine to make citrulline which is transported to the cytosol where aspartate is added. As urea is removed it is converted back to ornithine that goes back into the mitochondria to start over (Fig. 3).

Urea cycle disorders are metabolic disorders caused by deficiency of one or more enzymes in the cycle that is responsible for removing ammonia from the blood stream. Any of the five enzymes may lead to: carbamoyl phosphate synthetase (CPS) deficiency, ornithine transcarbamylase (OTC) deficiency, citrullinemia, argininosuccinic aciduria, and argininemia. In urea cycle disorders, the nitrogen accumulates in the form of ammonia, a highly toxic substance, and is not removed from the body. Ammonia then reaches the brain through the blood, where it causes irreversible brain damage and/or death.

Although the urea cycle was discovered by Dr. Hans A. Krebs early in 1930s, analysis of the urea cycle so far have not been systematically explored. The following paragraph therefore will focus on the possibility of integrative analysis of the urea cycle within the scope of systems biology. We are going to build a complete network of urea cycle and analyze its basic properties by using the PathAlinger system. We construct an integrative

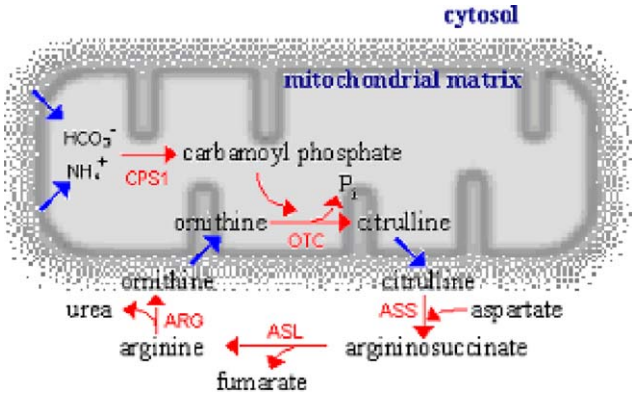

Fig. 3. Key enzymes in regulation of urea cycle in cells. CPS1: Carbamyl phosphate synthetase, EC 6.3.4.16; OTC: Ornithine transcarbamylase, EC 2.1.3.3; ASS: Argininosuccinate synthetase, EC 6.3.4.5; ASL: Argininosuccinate lyase, EC 4.3.2.1; ARG: Arginase, EC 3.5.3.1.

model, analyze the genetic variations and figure out the regulation of signaling pathways. One of the goals of the research is to highlight at large the identification and treatment of urea cycle disorders, so that the timely intervention will result in saving the lives of the disordered children. It also gives some information on the systems analysis of metabolic disorders.

3.2. Urea cycle integrative information

The presence of numerous informational and programming resources on biomedical data described in Section 2, raises an acute problem of data integration. Concerning urea cycle disorders, goal of such integration is to fuse all data available and to create a virtual informational environment enabling an access to the significant information on the basis of simultaneous exploration of many databases available via the Internet.

Based on the retrieved biomedical data and a protocol graph of urea cycle pathways and by manual searching OMIM and other drug sources, a graphical representation of the urea cycle metabolic network using the objects presented above for describing entities and interactions is drawn (Fig. 4). It shows an intricate network that links entities and interactions. This network includes not only the succession of biochemical reactions that lead to the transformation of CO_2 and NH_4^+ to urea, but also the regulation of gene expression and enzymatic activities. It furthermore displays (e.g., asparate, fumarate) the links to other pathways, which are not detailed on the graph to preserve clarity. In comparison with other existing urea cycle networks, such as the one that is drawn by KEGG, our network presents a more detailed information. The integrative network provides us a more real view on the urea cycle. There is more information that can effect the expression of urea cycle genes, which will give clinician more information about the mechanism of urea cycle disorders.

While today the molecular knowledge is still rudimentary, in some cases if the complete interrelationships of the network is unclear, or only a rudimentary pathway is provided, a prediction solution is needed. Meanwhile, methods of modeling and simulation will help to understand many important scientific questions. It can be interpreted as the basic step for implementing virtual worlds that allow virtual experiments. We use PathAligner [7] to retrieve metabolic information and reconstruct/predict the complete network.

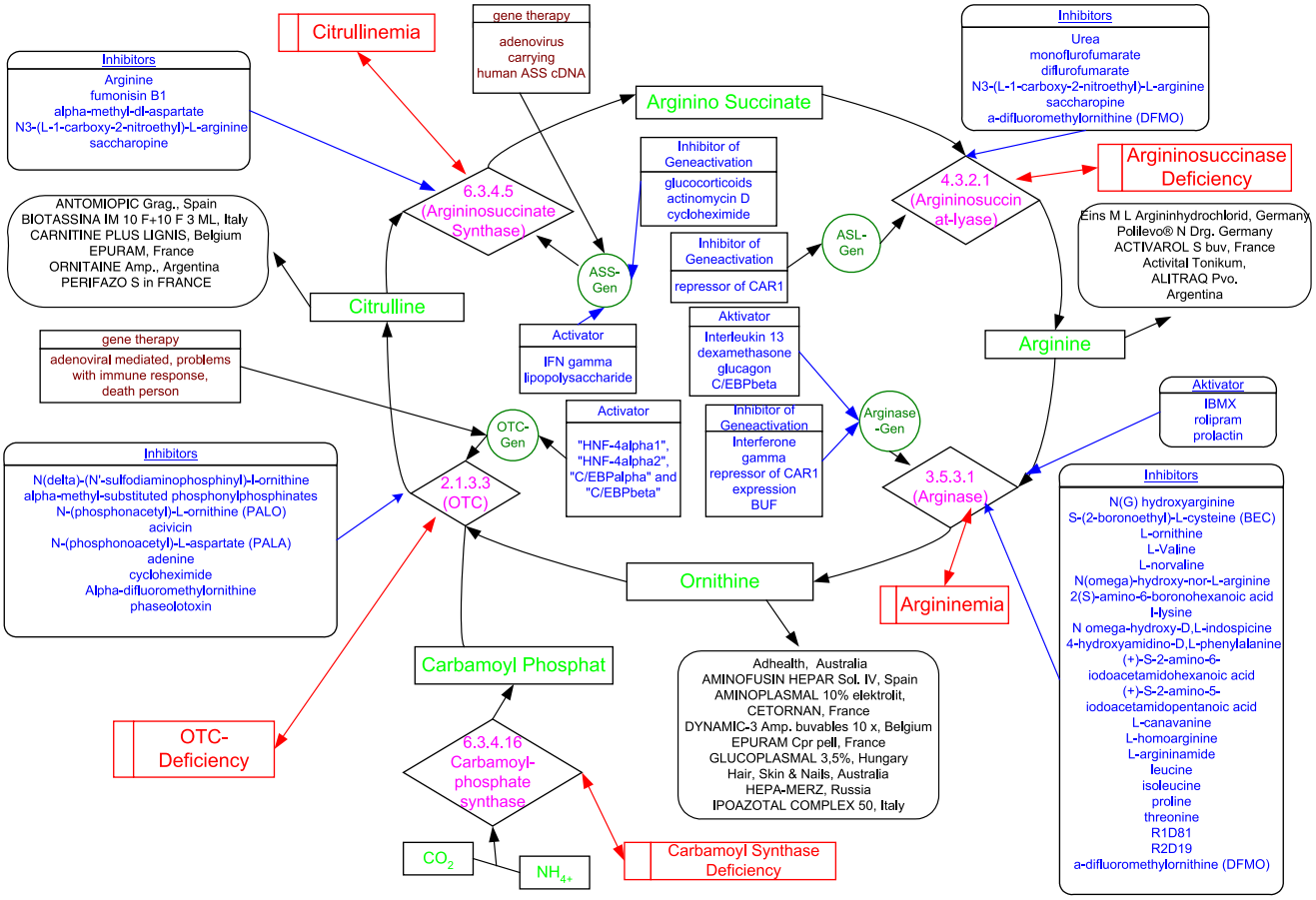

Fig. 4. Integrative diagram of urea cycle. The circle shows metabolic reactions of urea cycle. Enzymes are encoded by responding genes which are regulated by activator and inhibitors. Effective drugs (ingredients) show effects to these targets (enzymes). Drawn by Dr. Ralf Kauert.

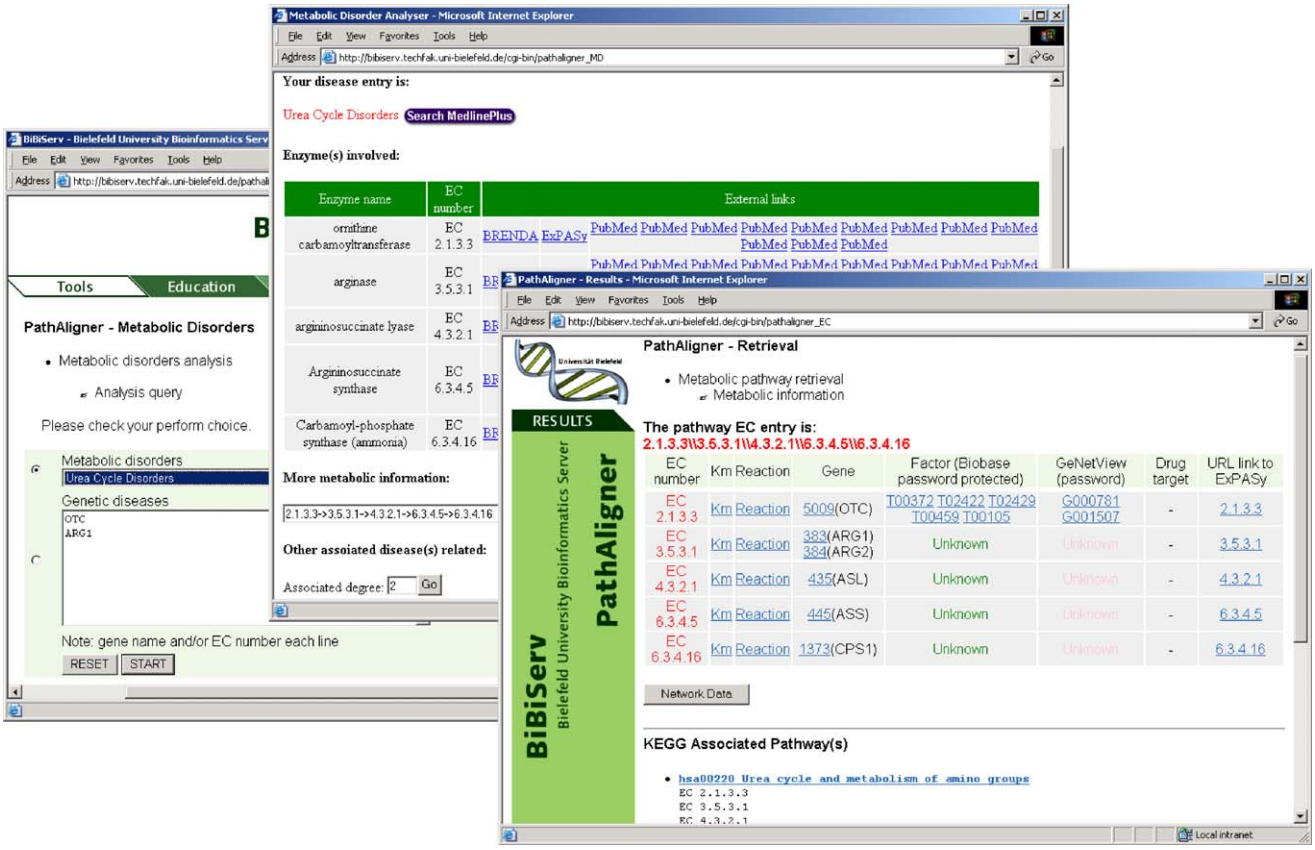

Fig. 5. Screenshots of the example query with PathAligner.

A PathAligner query of metabolic disorder "urea cycle disorders" is to predict metabolic information (Fig. 5). PathAligner helps to retrieve all the related EC (Enzyme Commission) numbers and involved metabolic pathways and more information on gene expressions. The output of query, the screen shot, is shown in the middle window of Fig. 4. The table lists the enzymes that are involved in the queried disease. Clicking the corresponding hyperlinks can retrieve additional information about the enzymatic reactions and K_m values (Michaelis constants). Further information about these enzymes can be determined according to the retrieved table. For example, the EC 2.1.3.3, its encoding gene is OTC and there are a number of transcription factors are shown in the column Factor. Moreover, the interactions between the genes and the transcription factors are also available by clicking the hyperlinks in the GeNetView. However, not all these data are available due to the incomplete source data. A list of the disease associated metabolic pathways is provided as a blueprint for further modeling and simulation with Petri net tools.

3.3. Modeling and simulation

PathAligner retrieves most desired data for a further Petri net model construction. We present a Petri net model of the integrative urea cycle network by using a hybrid Petri net. A hybrid Petri net supports both discrete model and continuous model. That is, the dynamic behavior of the model system, such as the metabolite fluxes, NH_4^+ input and urea output are well described with continuous elements; while control of gene expression are outlined with discrete ones due to the insufficiency of explicit expression data. Nevertheless, when explicit knowledge about expression levels of the enzymes are available; it is possible to exploit our model of gene regulatory network to handle realistic gene expression data with the state equations. The initial values of variables were assigned and tuned so that the model system behavior would comply maximally with available experimental data on the dynamic characteristics of the system's behavior.

Based on the retrieved data and some complementary data from literatures, a hybrid Petri net model of the gene regulated urea cycle metabolic network is presented and the dynamics of the main components on the model are shown in Fig. 6 and in Supplementary materials provided in the Web. In this paper, we would like to focus only on the model construction. Detailed description, testing/simulation and discussions of the model have been published in our previous work [15]. Petri net allows easy incorporation of qualitative insights into a

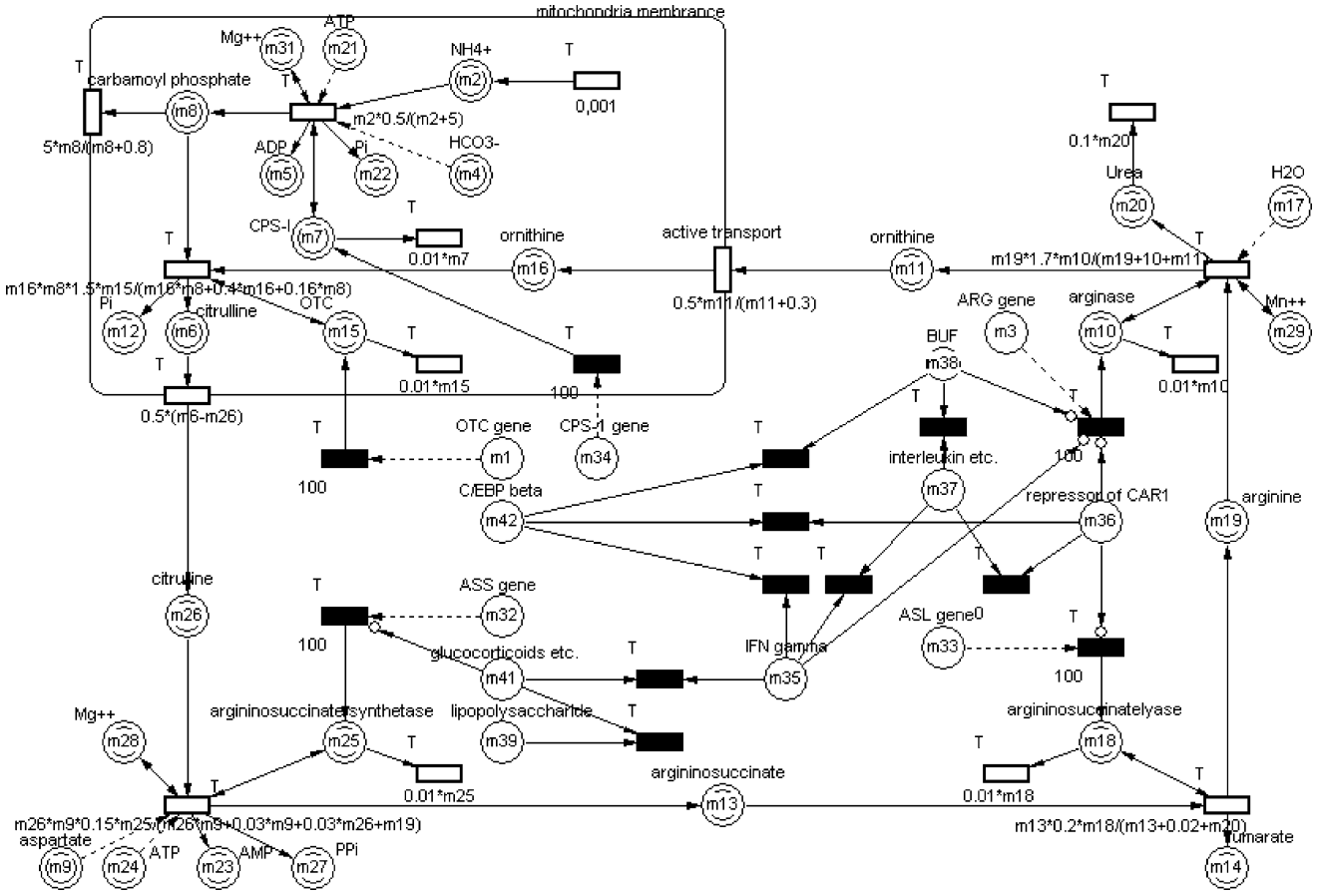

Fig. 6. Petri net model of urea cycle gene regulated metabolic network and dynamic simulation results.

pure mathematical model and adaptive identification and optimization of key parameters to fit system behaviors observed in gene regulated metabolic networks. The study of modeling and simulation can help detect genetic/metabolic defects. It is useful for diagnosis and therapy of genetic/metabolic defects as well as drug research.

As the urea cycle operates only to eliminate excess nitrogen. High concentration level of ammonia in the cell results in hyperammonemia which is a typical, coma and death ever been reported. Laboratory studies reveal elevated arginine levels, mild hyperammonemia, and a mild increase in urine orotic acid. The diagnosis now can be confirmed by enzymatic analysis in the model. On high-protein diets or under starvation state, proteins are degraded and amino acid carbon skeletons are used to provide energy, thus increasing the quantity of amino nitrogen that must be excreted. To facilitate this process, enzymes of the urea cycle are controlled at the gene level to enhance the concentration of enzymes. As urea cycle takes place both in mitochondria and cytoplasm, the effects involved also come from the membrane transportation. Some mitochondrial membrane diseases, e.g., ornithine transporter deficiency, surely effect the transportation of ornithine into matrix and results in high concentration of ornithine accumulation in plasma, which get a feedback to the transition of arginine into urea and finally hyperammonemia. From the model we know the treatment for defects in urea cycle enzymes could be either limit input of ammonia (limit protein intake) or replace missing intermediates from cycle (supplement with arginine or citrulline). Patients with OTC deficiency benefit from citrulline supplementation because citrulline can accept ammonia to form arginine.

3.4. SNP and transcription factors

Because the basis of the urea cycle disorders is a defect in a gene, research have been working on ways of getting a working gene into cells. Researchers have established that in the animal model, sparse fur (spf/Y) mouse, partial correction with gene therapy may be sufficient to normalize urea synthesis [26,27]. However, the current therapy is unsatisfactory for human. Optionally, we would like to target the working gene to the right cells and have it regulated and expressed just as well as the normal gene would be. Single nucleotide polymorphism (SNP) as well as transcription factor binding sites are the two aspects that have to be considered.

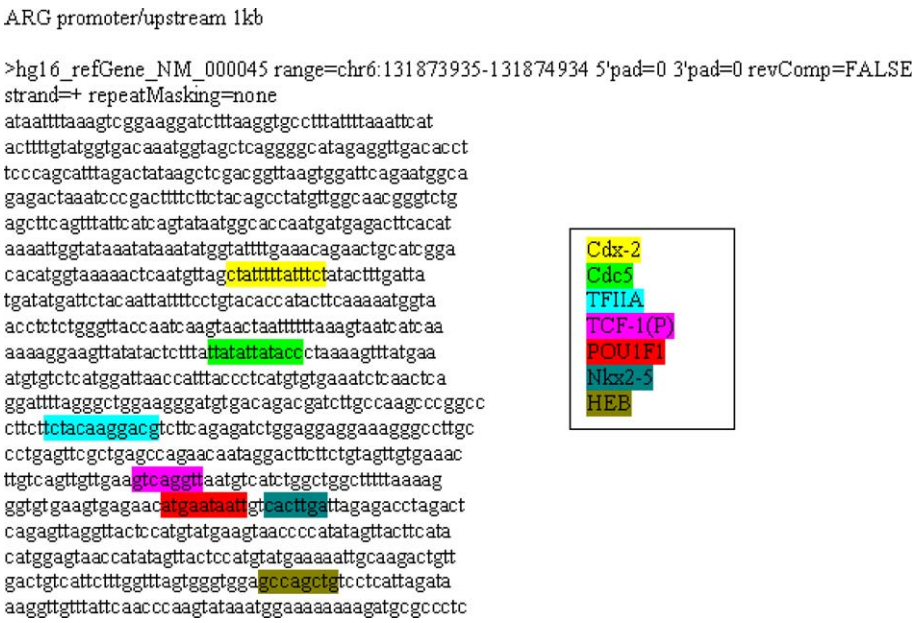

Fig. 7. Computational prediction of transcription factor binding sites of the human ARG genes. Binding sites of other involved genes in urea cycle are shown in the supplementary material.

In the progress of the Human Genome Project, scientists recognized that the existence of SNP in genome is helpful to explain the rich diversity of individuals, and the difference of susceptibility to diseases. A single base variation may cause gene function abnormalities. Therefore, searching and studying SNPs has become an important objective of biomedical informatics. The supplementary material shows the computationally annotated mutations of genes in the urea cycle. Further information can be obtained by browsing the related web-pages at http://mutdb.org/cgi-bin/Search.py?GOCODE=0000050. However, these mutations are meant to be used for basic research and not to make clinical decisions. In this section, we focus on the discovery of transcription factor binding sites by computational searching the 1 kb upstream promoter region sequences.

A TransFac database search for the transcription factor binding sites, using the human promoter sequences that are provided by UCSC [http://genome.ucsc.edu/], are shown in Fig. 7 and in the supplementary material. All potential binding sites of the urea cycle genes are summarized in Table 1 The search found 23 functional binding sites. ARG and OTC share four binding sites, which means that the expression of ARG and OTC might be simultaneously regulated by Cdx-2, Cdc5, Nkx2-5, and POU1F1. Nkx2-5 also affects ASL regulation. More functional events of transfactor can be retrieved and analyzed with the Biobase database.

3.5. Related signaling pathways

Based on the transcription factors, we obtain a list of signaling events that effect the gene expression of urea cycle. A graphical layout of the signaling pathways is constructed by using the Biolayout tool [28] (Fig. 8).

The average connectivity of the network is low (2.18). It is worthy to remark, that the out degrees are of lower grade when compared to the in degrees. In addition, the phenomenon of few key molecules such as NF-1 and POU2F1 with extraordinary high connection which work as hubs in the network is a general feature of biological molecular networks [29–35].

3.6. Associated diseases

We also investigated the associated diseases of these signaling pathways. A list of 227 transregulable genes related to the urea cycle signaling network were retrieved from Transpath. By querying against biological databases, such as Swiss-Prot and KEGG, all related enzymes can be retrieved. Then, searching the BRENDA database helps to determine the involved diseases. All diseases that are regulated by these signaling pathways are listed in the supplementary material. A hit is the total number of a disease retrieved association. On the left column, we do not consider the redundancy of enzymes that are encoded by different genes. For example, there are 10 hits of the enzyme protein kinase (EC 2.7.1.37) that is involved in various diseases, such as "acromegaly," "adhesions," "amyotrophic lateral sclerosis," "anemia, sickle cell," and so on. While on the right column, these 10 hits are regarded as 1 hit. Under this treatment, some diseases with high hits on the left column may show low hit score on the right column. On both lists, there are some already known diseases related to the urea cycle diseases, including "chronic liver disease," "ornithine carbamoyltrans-

Table 1
List of potential transcript factor binding site of the urea cycle genes

Transcription factor	Biobase Ac. No.	ARG	ASL	ASS	CAD	OTC
AML1	T01067		+	+		
Cdx-2	T02002	+				+
Cdc5	T04728	+				+
E2F-1	T01542 / T01543			+		
GATA-4	T02532 / T02687 ...			+		+
HEB	T01503	+				
HNF-3alpha	T02343 / T02512 ...					+
HNF-4alpha2	T02343 /T00373 ...					+
Lentiviral Poly A						+
MAZ	T00490 / T02303 ...			+		
NF-1	T01298 / T00539 ...			+		
NF-kappaB	T00587 / T00588 ...			+		
Nkx2-5	T01675 / T04323 ...	+	+			+
Pax-2	T01823 / T03013					+
Pax-4a	T02983 / T03001		+			+
POU1F1	T00325 / T00707 ...	+				+
POU2F1	T00641 / T00642 ...					+
RFX1	T00229 / T01666			+		
SREBP-1	T01559 / T04556 ...				+	
TCF-1(P)	T01109	+				
TFIIA	T00815 / T00816 ...	+				
USF	T00870 / T00871 ...			+		
Xvent-1	T04665				+	

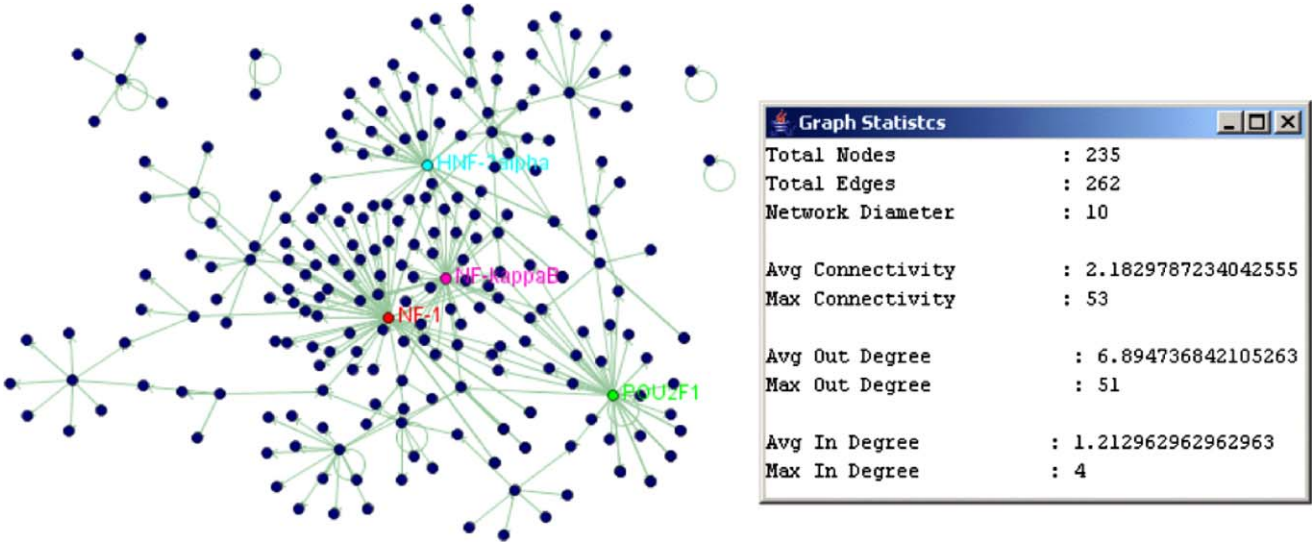

Fig. 8. A graphical layout of the signaling pathways involved in the urea cycle. Molecules with high degree of convergences are calculated. Colored nodes are those molecules with intensive divergence interconnection with others. They are NF-1 (51), POU2F1 (41), NF-κB (37) and, HNF-3α (27). Intensive convergences: Cdx2 (4), SREBP-1 (4), NF-1 (2), POU1F1(2).

ferase deficiency," "citrullinemia," etc. We are more interested in those with high hit scores. Common diseases with association degree (hits $\geqslant 3$) are shown in Fig. 9.

We surprisingly find that Rheumatoid arthritis is highly related. This is consistent with a recent research by Nissinen et al. [36]. They studied whether the en-

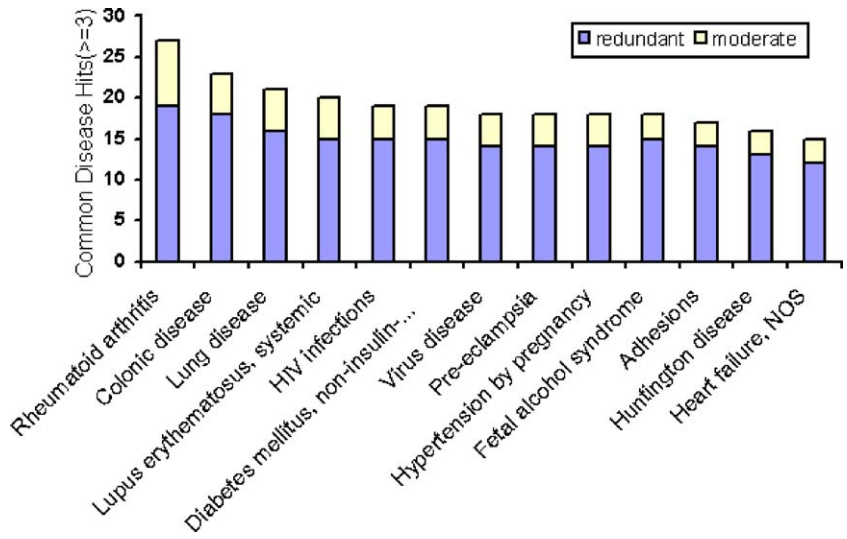

Fig. 9. Diseases related to the list of signaling pathways. Redundant hits are all counts of a retrieved association; a moderate hit is defined as the presence of one or more hits in a single enzyme query.

zyme peptidylarginine deiminase (PAD; EC 3.5.3.15), responsible for the post-translational modification of peptide-bound arginine residues to citrulline, constitutes an antigen for patients with rheumatoid arthritis (RA). The study shows that the arginine–citrulline converting enzyme PAD was recognized as a new antigen against which patients with inflammatory rheumatic diseases frequently show IgG class antibodies. From Fig. 9, we can see that systemic lupus erythematosus (SLE) also shows a significant involvement. Both RA and SLE are due to disorders of the musculoskeletal system and connective tissue, which is intensively related to immune systems. It is interesting that three decades ago researches have observed the altered immunoglobulin metabolism between SLE and RA [37]. Later, the prevalence clinical and laboratory associations of SLE and RA were determined by many researches [38–40]. Other latest observations of the association between RA and urea cycle relevance were achieved by Yonekura et al. [41] and Iwashige et al. [42].

Currently, the main urea cycle disorders' management is dietary manipulation by reducing in protein intake. It is possible to increase residual enzyme activity by supplying cofactor. The alternative pathway therapy [43] by intaking chemicals to remove NH_3 via other pathways are practiced, but having limited effectiveness in preventing hyperammonemia and must be combined with effective dietary management [44]. The future therapy will focus on gene repair, or genetic counseling. This needs more knowledge about cellular function. We hope our approach can give a highlight on this direction. The bioinformatics analysis approach will also represent the backbone of the concept of disorders management in the post-genomic era.

4. Conclusion

Genetic/metabolic defects often lead to metabolic blockades and result in metabolic disorders. Regarding the development of methods and concepts of bioinformatics to analyze metabolic disorders, it is necessary to first understand the reaction pathways that are affected by the encoded genes, directly and indirectly and to know the effect of modification of reaction steps and depletion of metabolites on overall reaction networks. Biomedical information retrieval and systems modeling and simulation are some major tasks in this field. We presented a general framework of systems biomedical information analysis. The web-based PathAligner was implemented to enable users easily navigate disease related genomic and metabolic information. The Petri net methodology was exploited to model and simulate the biomedical system. We have presented an analysis of urea cycle in a systematic way. By exploiting the existing huge amounts of data available in the various databases, we described metabolic mechanisms and pathways, patterns of regulatory regions; transcriptomics, and metabolomics data of urea cycle. The construction and analysis of signaling networks in general allowed us to examine the integrity of the accumulated data. This basic network is useful for further modeling, prediction and comparison of different cell systems. The graphical representation and dynamic simulation results give us an intuitional understanding about the logics of the cellular metabolism. The bioinformatics analysis can be a safe method to look insight and simulate the function of a gene network. We suggest the urea cycle signaling network related list of genes be considered for further investigations related with Rheumatoid arthritis and the other mentioned diseases.

By integrating multiple levels of metabolic data, we can gain a global perspective of how a biomedical system works, the mechanism by which the variant gene(s) or an error in a metabolic pathway produces clinical manifestation, and the role of gene therapy, nutrition, and life style modification as strategies to treat or prevent disease development. Implementation of it will enable us to apply the advanced data integration and Petri net modeling procedures to (a) integration of biological and biomedical data; (b) systems functional structure model construction; (c) predictions of genetic predispositions, disease discovery, diagnostics, drug development, and toxicology; (d) alternative metabolic pathways identification by altering the disrupted metabolic pathway(s) and/or compensating for them; and (e) systematic experimental design to provide the right set of facts that permit accurate analysis; and so on. As a result, for example, scientists can use the systems model to predict metabolic diseases and test drugs comprehensively before testing them in the laboratory or clinic.

Acknowledgments

The work was supported by the German Research Foundation (DFG). Thanks to Mr. Sridhar Hariharaputran for his useful discussion.

References

[1] Lehninger AL, Nelson DL, Cox MM. Principles of biochemistry, 2nd ed., New York, 1993, p. 12.
[2] Clarke TR. A clinical guide to inherited metabolic diseases. Cambridge: Cambridge University Press; 1996.
[3] Glanze WD, Anderson KN, Anderson LE, et al. Mosby's medical and nursing dictionary. 2nd ed. Toronto: The C.V. Mosby Company; 1986.
[4] Gomez CM. Polyglutamine aggregates in SCA6 Purkinje cells: a tail of two toxicities. Neurology 2001;56(12):1618–9.
[5] Sujansky W. Heterogeneous database integration in biomedicine. J Biomed Inform 2001;34(4):285–98.
[6] Mork P, Halevy A, Tarczy-Hornoch P. A model for data integration systems of biomedical data applied to online genetic databases. J Biomed Inform 2001;35(Suppl.):S473–7.
[7] Chen M, Hofestaedt R. Web-based Information Retrieval System for the Prediction of Metabolic Pathways. IEEE Trans Nanobiosci 2004;3(3):192–9.
[8] Nebot A. Qualitative modeling and simulation of biomedical systems using fuzzy inductive reasoning. PhD dissertation, Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, Barcelona, Spain. 1994.
[9] Thomaseth K. A modeling tool for biomedical systems. Simulation Pract Theory 2001;9(1–2):1–19.
[10] Peterson JL. Petri Net Theory and the modelling of systems. Englewood Cliffs, NJ: Prentice Hall; 1981.
[11] Reddy VN, Mavrovouniotis ML, Liebman MN. Petri net representation in metabolic pathways. In: Hunter, L. et al. editors. Proceedings first international conference on intelligent systems for molecular biology. Menlo Park: AAAI Press; 1993; p. 328–36.
[12] Hofestaedt R, Thelen S. Quantitative modeling of biochemical networks. In silico Biol 1998;1:980006.
[13] Matsuno H, Doi A, Nagasaki M, et al. Hybrid Petri net representation of gene regulatory network. In: Pacific symposium on biocomputing 2000;5:338–49.
[14] Genrich H, Kueffner R, Voss K. Executable Petri net models for the analysis of metabolic pathways. Intl. J. STTT 2001;3:394–404.
[15] Chen M, Hofestaedt R. Quantitative Petri net model of gene regulated metabolic networks in the cell. In Silico Biol 2003;3:347–65.
[16] Heiner M, Koch I, Will J. Model validation of biological pathways using Petri nets—demonstrated for apoptosis. Biosystems 2004;75(1–3):15–28.
[17] Doi A, Fujita S, Matsuno H, et al. Constructing biological pathway models with hybrid functional Petri nets. In Silico Biol 2004;4(2):271–91.
[18] Atanassov KT, Bustince H, Daskalov M, et al. Generalized net models in neurology. In: Proceedings of the international workshop bioprocess engineering'95, Sofia, Oct. 2–5, 1995.
[19] Daskalov M, Nikolov N, Bustince H. Gerneralized net models in neuroplogy. In: Proceedings of the seventh national conference on biomedical physics and engineering, Sofia, 17–19 Oct. 1996, 185–8.
[20] Bustince H, Kim S, Kim Y, et al. Gerneralized nets in neurology. In: Proc. Of the Int. Symp. Bioprocess Systems'97, Sofia, Vol. VI, 1997, 20–7.
[21] Sorsich J, Georgiev P, Kim S. Application of the generalized net in medicine. In: Proceedings of the seventh national conference on biomedical physics and engineering, Sofia, 1996, 195–7.
[22] Ermel C, Padberg J, Ehrig H. Requirements engineering of a medical information system. In: Proceedings of the 2nd world conference on integrated design and process technology, 1: 1996, 186–93.
[23] Ouchi Y, Tazaki E. Medical diagnostic system using fuzzy coloured Petri nets under uncertainty. Medinfo 1998;9(1):675–9.
[24] Gerhard E, Wippich K. Structural description of the human eye using Petri net. Biomed Tech (Berl) 1991;36:66–9.
[25] Mounts WM, Liebman MN. Qualitative modeling of normal blood coagulation and its pathological states using stochastic activity networks. Int J Biol Macromol 1997;20(4):265–81.
[26] Ye X, Robinson MB, Batshaw ML, et al. Prolonged metabolic correction in adult ornithine transcarbamylase-deficient mice with adenoviral vectors. J Biol Chem 1996;271(7):3639–46.
[27] Ye X, Robinson MB, Pabin C, et al. Transient depletion of CD4 lymphocyte improves efficacy of repeated administration of recombinant adenovirus in the ornithine transcarbamylase deficient sparse fur mouse. Gene Ther 2000;7(20):1761–7.
[28] Enright AJ, Ouzounis CA. BioLayout—an automatic graph layout algorithm for similarity visualization. Bioinformatics 2001;17(9):853–4.
[29] Fox JJ, Hill CC. From topology to dynamics in biochemical networks. Chaos 2001;11(4):809–15.
[30] Wagner A, Fell DA. The small world inside large metabolic networks. In: Proc R Soc Lond B Biol Sci 2001;268(1478): 1803–10.
[31] Agrawal H. Extreme self-organization in networks constructed from gene expression data. Phys Rev Lett 2002;89(26):268702.
[32] Bahn A, Galas DJ, Dewey TG. A duplication growth model of gene expression networks. Bioinformatics 2002;18(11):1486–93.
[33] Karev GP, Wolf YI, Rzhetsky AY, et al. Birth and death of protein domains: a simple model of evolution explains power law behavior. BMC Evol Biol 2002;2(1):18.
[34] Ravasz E, Somera AL, Mongru DA, et al. Hierarchical organization of modularity in metabolic networks. Science 2002;297(5586):1551–5.

[35] Ma H, Zeng AP. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 2003;19(2):270–7.
[36] Nissinen R, Paimela L, Julkunen H, et al. Peptidylarginine deiminase, the arginine to citrulline converting enzyme, is frequently recognized by sera of patients with rheumatoid arthritis, systemic lupus erythematosus and primary Sjogren syndrome. Scand J Rheumatol 2003;32(6):337–42.
[37] Levy J, Barnett EV, MacDonald NS, et al. Altered immunoglobulin metabolism in systemic lupus erythematosus and rheumatoid arthritis. J Clin Invest 1970;49(4):708–-715.
[38] Newkirk MM, Rauch J, Mageed RA, et al. Restricted immunoglobulin variable region gene usage by hybridoma rheumatoid factors from patients with systemic lupus erythematosus and rheumatoid arthritis. Mol Immunol 1993;30(3):255–63.
[39] Witte T, Hartung K, Sachse C, et al. Rheumatoid factors in systemic lupus erythematosus: association with clinical and laboratory parameters. Rheumatol Int 2000;19(3):107–11.
[40] Carlson E, Rothfield N. Etanercept-induced lupus-like syndrome in a patient with rheumatoid arthritis. Arthritis Rheum 2003;48(4):1165–6.
[41] Yonekura Y, Koshiishi I, Yamada K, et al. Association between the expression of inducible nitric oxide synthase by chondrocytes and its nitric oxide-generating activity in adjuvant arthritis in rats. Nitric Oxide 2003;8(3):164–9.
[42] Iwashige K, Kouda K, Kouda M, et al. Calorie restricted diet and urinary pentosidine in patients with rheumatoid arthritis. J Physiol Anthropol Appl Human Sci 2004;23(1): 19–24.
[43] Batshaw ML, Brusilow S, Waber L, et al. Treatment of inborn errors of urea synthesis: activation of alternative pathways of waste nitrogen synthesis and excretion. N Engl J Med 1982;306:1387–92.
[44] Batshaw ML, MacArthur RB, Tuchman M. Alternative pathway therapy for urea cycle disorders: twenty years later. J. Pediatr. 2001;138(1 Suppl.):S46–55.