Medical Informatics Europe '99
P. Kokol et al. (Eds.)
IOS Press, 1999

Metabolic Drug Pointing and Information Processing

Ralf Hofestädt, Ulrike Mischke, Manuela Prüß, Uwe Scholz
Institute of Technical and Business Information Systems
Otto-von-Guericke-Universität Magdeburg
P.O. Box 4120, D-39016 Magdeburg
E-mail: {hofestaedt|mischke|pruess|uscholz}@iti.cs.uni-magdeburg.de
Tel.: ++49-391-67-18659 Fax: ++49-391-67-12020

Abstract. Computer supported drug design is based on the biochemical information for the prediction of alternative bio-chemical pathways. Molecular information on genes, proteins, biochemical reactions, mutations, inborn errors and metabolic diseases are available via internet. Based on this information we developed an information retrieval and processing concept combining knowledge about metabolic diseases and biochemical effects.

1. Introduction

Up to now about 450 genetic metabolic defects are known. The most frequent conditions are diseases in the degradation pathway of amino acids or the just recently described fatty acid oxidation disorders. It is estimated that every 4000th newborn is affected with such a disease. For most of these defects dietary and/or medical treatment is available.
Thus by applying modern methods, for many of these metabolic diseases biochemical diagnosis is available, so a rapidly increasing number of diagnosed children and infants is to expect. The most challenging problem is the fact that inborn errors of metabolism are very heterogeneous. Prediction of the phenotype by knowing the genotype has become more and more important since the decision for initiating a certain treatment may be difficult using only biochemical data. The possibility for a genotype-phenotype correlation in a certain disease is dependent from several parameters. However, the number of patients is relatively small and they are widespread, so for making any correlation, an international metabolic diseases database for rare metabolic disorders is very useful.
A lot of different electronical information systems are available which represent the medical level of knowledge. Moreover, different databases concerning data on diseases, genes, enzymes, biochemical reactions, and biochemical effects of drugs are available. The idea of our project was and is to integrate the microscopic and macroscopic knowledge of metabolic diseases pointing out the biochemical effects of drugs.

2. Metabolic Diseases, Diagnosis, and Therapy

Inherited metabolic diseases are a group of disorders characterised by pathological abnormalities of metabolic pathways. Most of them are caused by a block regarding the corresponding metabolic pathway, others by a defect in a transport protein or a defect in the storage of products of the metabolism. The underlying defect leading to this type of disease is mostly an inborn enzyme deficiency, caused by a single gene mutation.
The knowledge of the pathogenesis of these diseases is very important, because it can explain the manifestations of a specific disorder, is necessary in order to reach a decision about rational treatment, and may even suggest that treatment is not feasible. Today, information on metabolic, molecular and genetic details for many metabolic diseases is

available, which brings with it new perspectives concerning the complex biological processes and resulting methods of diagnosis and therapy. At the same time, efforts to understand the biochemical and especially the molecular basis underlying human genetic variation are more vigorous then ever [1].
One of the most common inborn metabolic diseases is the Ornithine Transcarbamylase Deficiency, or OTC Deficiency. The deficient enzyme, Ornithine Transcarbamylase, or OTC, is involved in the urea cycle, and there are more than 60 known mutations of the OTC gene [2]. The OTC gene of the DNA, part of the whole genotype of a person, encodes the enzyme OTC. This enzyme's special task lies in the metabolic pathway, the urea cycle, where it catalyses the reaction from Ornithine to Citrulline. In case of a gene defect, the enzyme will not be produced and the urea cycle is interrupted. As a result, typical pathological effects occur concerning the phenotype of the involved person.
The diagnosis of metabolic diseases generally is based on the detection of metabolites. The reason is that specific metabolites can not be produced, basing on a genetic defect. Beyond these biochemical reactions new methods of biotechnology allow the molecular detection of this metabolites [3]. Therefore, the diagnosis of metabolic processes can be realised on different metabolic levels.
In general, however, efforts are made to intervene therapeutically into the involved pathway, although substitution of the lacking gene product mostly is difficult. In most of the cases a special dietetic treatment is used to prevent the accumulation of undegradable metabolites. If the enzyme activity is reduced, it sometimes can be activated by high-dosed intake of essential cofactors, e.g. of biotin. Another method of degradation of metabolites is the activation of alternative metabolic pathways. Such an activation can occur pharmacologically by supplementation of substances which bind and help to excrete toxic metabolites, or possibly gene therapeutically by the introduction of a new gene respectively a new enzyme which "nudges" a new pathway, resulting in a normal phenotype. Research is undertaken intensively, especially, in the field of metabolic diseases [4].

3. Metabolic Drug Pointing

A drug can be defined as a chemical compound, natural or synthetic, that is used in the prevention, diagnosis, treatment, or cure of disease, for the relief of pain, or to control or improve any physiological or pathological disorder in humans or animals. The effect of a drug depends on its concentration, its reaction with specific receptors and its interaction with different substances. The computer supported drug design is becoming more and more relevant [5].
In case of metabolic diseases, information on drug effects and on specific drug destinations is rare. Here different levels can be important: on the one hand the biochemical level, at which the conversion of the substances takes place. On the other hand, also the DNA level is of great interest, because here methods of gene therapy could be used to replace a defect or deficient gene. Looking at a specific metabolic disorder, e.g. the OTC deficiency, for its understanding and especially for the understanding of its therapy, it is important to know where and how a given drug intervenes in the individual's organism. This is what we call "drug pointing". In the case of OTC deficiency, the urea cycle is blocked at the conversion of ornithine to citrulline and nitrogen can not be excreted as urea. Normally, glutamine is degraded via the urea cycle. An alternative degradation of glutamine is possible: when sodium phenylacetate is given as a drug, it converts glutamine to phenylacetate glutamate, which can be excreted.
It is the aim of our Metabolic Diseases Database (MDDB), to point out not only the regular pathways related to different metabolic diseases, but also alternative pathways and biochemical drug effects, as far as there is information available.

4. Information Systems and Metabolic Diseases

For the detection of inborn errors the database system METAGENE is available [6]. This database system provides medical knowledge of inborn errors (macroscopic knowledge). Our group has extended the METAGENE database system by collecting the molecular

knowledge of metabolic diseases (microscopic knowledge). This new database system is called Metabolic Diseases Database (MDDB). Moreover, access to the OMIM database system provides medical knowledge of metabolic diseases, and access to the molecular knowledge (genes, proteins, and pathways) is realised by using the powerful information system KEGG [7] which allows access to every known metabolic pathway including the related genes and proteins. Besides the development of the MDDB, the design and implementation of an integrative molecular knowledge server, called BIOBENCH, for the detection of inborn errors is one of our research goals. The architecture of our system allows automatic access to different database systems.

4.1 Related Works

Today, a new update of P.O.S.S.U.M (Pictures Of Standard Syndromes and diagnoses Malformations) is available. The biochemical genetics unit of the Montreal Children's Hospital is in the process of implementing a database system, that will provide knowledge of Phenylalanine-Hydroxylase (PKU). The University of Nürnberg-Erlangen has started a new project called Computer-Mediated-Communication, which will implement a professional electronic mailing list for the discussion of topics related to inborn errors of the metabolism. Moreover, specific database systems exist which provide medical knowledge, a lot of them are available via the Internet. At a microscopic level, methods of metabolic engineering will allow the analysis and synthesis of metabolic pathways. Different database systems for genes, proteins, and pathways are already available.
Nowadays, specific integrative information systems are implemented which allow specific access to molecular data. For the detection of metabolic diseases typical gene sequences, proteins, and metabolic pathways must be identified and analysed. An electronical representation of the complete metabolic pathway data set does not exist. Boehringer collects all biochemical reactions and presents this data graphically on the so called Boehringer Biochemical Pathway Chart [8]. The electronical representation of this data is the main goal of the KEGG project.

4.2 MDDB - A Molecular Information System for the Prediction of Alternative Pathways

We developed and implemented a database system called Metabolic Diseases Database (MDDB) which provides microscopic and macroscopic knowledge for the detection of a specific group of inborn errors, the so called hyperammonemias. We want to present not only the medical data, like general information on the disease, symptoms, laboratory findings and therapy, but also the molecular data. This includes information on genes, gene variants and their description, and gene regulation elements, as far as there are data available. The data on enzymes include also general information like EC number, synonyms and structure of the enzyme and information on the catalysed biochemical reaction, with structural formulas of the substances. Additionally, we collect data about drugs which are of therapeutic use in different cases, and we try to show their effects on the metabolic pathways, as described in section 3. The related pathways are shown, and possible alternative pathways will be predicted.
Thus using different information systems and literature, we are extending medical knowledge (macroscopic knowledge) as presented in the METAGENE database systematically by adding information about molecular processes. The MDDB has been developed to simplify data collection and to support the persistent storage of information about metabolic diseases. Moreover the user of the MDDB can obtain access to the various data in a uniform and integrated form.

5. BIOBENCH - a Knowledge Server for the Detection of Inborn Errors

The MDDB is integrated into the BIOBENCH system; a prototype of it is available via Internet. Its key idea is to develop and implement an information tool for the detection of inborn errors which is based on new technology, using methods of *information fusion* [9]. Therefore, we will combine the basic knowledge of medicine and molecular biology using a multi-database system. The basic module of our molecular knowledge server is the MDDB

database system. Moreover, automatic access to the KEGG information system and the TRANSFAC database system is implemented. KEGG represents the molecular knowledge of metabolic pathways and TRANSFAC the molecular knowledge of gene regulation processes. Using our knowledge server, the information stored in these systems can easily be used. Communication between the user and our knowledge based server is founded on an uniform user interface. Regarding the current version, the user can ask for diseases, enzymes, and pathways. Asked for example for the OTC deficiency, the system will automatically create agents which will contact all the linked databases and information systems. The created agents will automatically identify every relevant entry, which will be sent to the server of our system, followed by the server's realisation of the information fusion between the data collected by the agents. At the end of that process an information window, which represents the selected information, will be presented to the user.

6. Discussion

Today, methods of biotechnology enable the systematic investigation of metabolic diseases, and gene therapy is becoming popular. This does not only influence the molecular diagnostics in a positive way, but also the therapy of this diseases. As a result of this research a huge amount of data arose and is still growing. There are different information systems available, which manage and offer this data via Internet.
The idea of our work is to expand the macroscopic knowledge about metabolic diseases by collecting molecular information, too. This process is what we call information fusion. Therefore, we have developed and implemented the Metabolic Disease Database (MDDB). Regarding a metabolic disease, our information system presents information about the corresponding metabolic pathways, genes, enzymes, and medicine (drugs). Moreover, relevant references and therapy plans are offered. The special feature of the MDDB we are working at will be the realisation of the so called drug pointing. This data will be of great relevance for the computation of alternative pathways and for the tuning of gene therapy.
We embedded our MDDB system in an integrative molecular information system, the BIOBENCH, which provides a connection to the KEGG information system and the TRRD database system. It enables an easy access to relevant biochemical and genetic data concerning metabolic diseases.

Acknowledgement

This work is supported by the Kurt-Eberhard-Bode-Stiftung in the Stifterverband für die Deutsche Wissenschaft.

References

[1] A. L. Beaudet *et al.*, Genetics, biochemistry, and Molecular Basis of Variant Human Phenotypes. In: C. R. Scriver *et al.* (eds.), The Metabolic and Molecular Bases of Inherited Disease (7. ed.). McGraw-Hill, New York, 1995, pp. 53-118.
[2] M. Tuchmann and R. J. Plante, Mutations and Polymorphisms in the Human Ornithine Transcarbamylase Gene: Mutation Update Addendum. *Hum. Mut.* **5** (1995) 293-295.
[3] C. Perez-Carda *et al.*, An Unusual Late-Onset Case of Propionic Acidaemia: Biochemical Investigations, Neuroradiological Findings and Mutation Analysis. *Eur. J. Pediatr.* **157** (1998) 50-52.
[4] K. Overturf *et al.*, Ex Vivo Hepatic Gene Therapy of a Mouse Model of Hereditary Tyrosinemia Type I. *Hum. Gene Ther.* **9** (1998) 295-304.
[5] H.-J. Böhm *et al.*, Wirkstoffdesign. Spektrum Akademischer Verlag, Heidelberg, Berlin, Oxford, 1996.
[6] U. Mischke *et al.*, KBS-DIAMET: Database and Expert System for Diagnosis and Treatment of Patients with Inborn Errors of Metabolism. *J. Inherit. Metab. Dis.* **18** (1995) 224-226.
[7] S. Goto *et al.*, Organizing and Computing Metabolic Pathway Data. In: R. Altman et al. (eds.), Proceedings of the Pacific Symposium on Biocomputing. (World Scientific), 1997, pp. 175-186.
[8] G. Michal (ed.), Biochemical Pathways. Böhringer Mannheim, Penzberg, 1993.
[9] R. Hofestädt and U. Scholz, Information Processing for the Analysis of Metabolic Pathways and Inborn Errors. *Biosystems* **47** (1998) 91-102.