BiotechPro: A DATABASE FOR MICROBIOLOGICALLY SYNTHESIZED PRODUCTS OF BIOTECHNOLOGICAL VALUE

Ibragimova S.S.*, Smirnova O.G., Grigorovich D.A., Khlebodarova T.M.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia
* Corresponding author: isola@bionet.nsc.ru

Key words: database, microbial organism, industrially important products, microbial biotechnology, strain improvement

SUMMARY

Motivation: Development of bacterial strains producing biotechnologically important substances is a complex and many-staged process. For a better progress in the area, a knowledge base of biotechnologically valuable molecular-genetic bacterial systems and processes is required. This knowledge base should contain information on the organisms, which are the sources of a particular target product; on its biosynthesis pathways; on the genes whose products are involved in the metabolic pathways; and on the constants and dynamic variables of a particular process to enable a computer simulation of the process under various environmental conditions, given a desired yield of the target product. Eventually, on the basis of these data, the knowledge base should help scientists develop a gene engineering strategy for modification of a particular strain.

Results: We have developed a database of industrially important products (BiotechPro) produced by microbes. The database includes information about species producing a particular product, the metabolic pathway or functional components of the operon regulating this pathway, industrial strains, and associated substrates for each strain. BiotechPro is aimed to provide information for applied biotechnology and strain optimization projects.

Availability: the database is available at http://www.mgs.bionet.nsc.ru/mgs/dbases/biotech/

INTRODUCTION

Many products of bacterial cellular metabolism possess physiological activity and are of practical value for various industries. Advancements in molecular biology and recombinant DNA technology allow the number of genes for a particular bacterial product and the level of its expression in that particular strain to be enhanced as desired (Stephanopoulos, 2002). To handle such tasks, informational support is required in the form of computer analysis of experimental data and computer simulation of the processes involved. Information on the metabolites synthesized by various bacterial species, on the strains in which a particular species can appear and on the molecular-genetic systems controlling its biosynthesis is scattered across multiple publications, each addressing but few aspects of the synthesis of the target product.

A specialized database compiling the information about biotechnologically important products, synthesized by microorganisms, their biosynthesis pathways, and source strains used in biotechnologies may be used for solving this problem.
The primary navigator in the field is “The Prokaryotes”, an electronic resource (http://141.150.157.117:8080/prokPUB/index.htm), which can be used as an external source of data on the most biotechnologically important products produced by microorganisms. The database contains information largely on industrial source strains that have extensively been used in biotechnology. BSD databases: the Biodegradative Strain Database and the Biocatalysis/Biodegradation Database are narrowly focused databases containing information on strains producing enzymes that utilize hazardous chemicals (Urbance et al., 2003, Ellis et al., 2006).

We have developed a database of industrially important products (BiotechPro) produced by microbes; we have since been collecting information about the species that are sources of a particular product, the metabolic pathway or functional component of the operon that regulates the corresponding pathway, industrial strains and associated substrates for each strain. Information of industrial products, strains and substrates is collected from scientific literature.

BiotechPro consists of two databases BiotechProduct and BiotechStrain cross-linked with one another. This provides a possibility to select a potential bacterial producer of a biotechnologically important product and information on how much of the product the strain can synthesize and on the associated substrate for that strain, with regard to data on its biosynthetic pathway.

BiotechPro will contain consolidated data on microorganisms whose products are or could be used in different biotechnological processes.

IMPLEMENTATION AND RESULTS

BiotechPro database is implemented on the SRS platform and consists of 2 subdatabases: database on product and database on strain. Both databases have cross-reference fields.

**Structure and description of product database.** This database compiles information of industrially important products produced by different microbial species, which are or could be used in biotechnology (Fig. 1a).

The PRODUCT field contains the full and short names of the product. The SPECIES field contains the Latin names of the organisms that synthesize this product. The PATHWAY field contains information on the metabolic pathway for the target product, with the different references to the KEGG, MetaCyc databases (if the pathway is known) or the functional components involved in the synthesis of the target product (Caspi et al., 2006; Kanchev et al., 2006). The field COMMENT details the biosynthetic pathways of the target product. The APPLICATION field contains information about biotechnological processes in which the product could be used. BiotechProduct subdatabases contain two cross-reference fields – PRODUCT_ID and STRAIN_ID to subdatabases BiotechStrain.

BiotechProduct database has links to databases containing additional information on product:
- GeneNet, a database that contains descriptions of the local gene networks and biosynthetic signal transduction pathways generating the target product (Ananko et al., 2005);
- ProkaTEX, a database that is developed on the background of EcoTRRD, contains structural descriptions of the promoters of prokaryotic genes (including E. coli) whose products are involved in biotechnological processes (Khlebodarova et al., this issue);
- KiNet, a database that contains the descriptions of the constants and dynamic variables required for modeling the process as a whole and in part (Khlebodarova et al., this issue);
- ModelER, a database that contains computerized models for the functioning of the molecular-genetic and metabolic processes being investigated in different environments for different source organisms (unpublished data).
Modelling of molecular genetic systems in bacterial cell

Figure 1. Example of entries of BiotechPro database: (a) – product description, (b) – strain description.

Structure and description of strain database. This database accumulates information about strains that produce industrially important products. The strain description contains 3 fields. Each strain has an individual strain data page (Fig. 1b).

The STRAIN field contains the Latin name of the strain, the strain designation and a brief description of the origin of the strain. The PRODUCTIVITY field contains information on how much of the product the strain can synthesize and on the associated substrate for that strain. The reference stating that the strain produced this product, or degrades or transformed this substrate is included in the REFERENCE field. References are also provided as links to PubMed abstracts when available. BiotechStrain subdatabases contain cross-reference field to subdatabases BiotechProduct – PRODUCT_ID.

The BiotechPro database is to be populated with data on the strains that synthesize products, which are or could be used in biotechnology. With BiotechPro, the biotechnologists can pick out source strains either to address particular biotechnological tasks or to modify the strain as deemed desirable.

SRS tools allow database fields to be indexed and searched using a query system.

The main field for the search of the subdatabase BiotechProduct is the PRODUCT field. The search by the field SPECIES provides information on the organisms that synthesize the product of your interest. The search by the field APPLICATION allows you to find the strains for particular tasks.
The fields STRAIN and PRODUCTIVITY are the main fields for search in BiotechStrain subdatabases. The search by the field PRODUCTIVITY allows one to find the references to the microorganisms that utilize a particular substrate.

At present, BiotechPro contains information on 50 industrially important products produced by microbes. It also contains information on more than 60 associated strains that are sources of amino acids, vitamins, food additives, industrially important thickeners and enzymes.

We plan to expand the content of BiotechPro in future.

ACKNOWLEDGEMENTS

The work was supported by the Siberian Branch of the Russian Academy of Sciences (Integration Project on Fundamental Research of SB RAS No. 24 “Role of microorganisms in live system functioning: fundamental aspects and bioengineering applications”) and by Innovation project of Federal Agency of Science and Innovation IT-CP.5/001 “Development of software for computer modeling and design in postgenomic system biology (system biology in silico)”. The authors are grateful to I.V. Lokhova for bibliographical support.

REFERENCES

Khlebodarova T.M. et al. (2006) In silico cell II. Information sources for modeling Escherichia coli metabolic pathways. This issue.