

## WEB SITE REVIEW

# Review of the Enzymes and Metabolic Pathways (EMP) Database

Anthony P. Burgard and Costas D. Maranas<sup>1</sup>

*Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802*

Received April 19, 2001; accepted April 19, 2001

### INTRODUCTION

The Enzymes and Metabolic Pathways (EMP) database is an electronic compilation of enzymatic and metabolic data from over 15,000 literature sources covering over 1800 different organisms (Selkov *et al.*, 1996). The Metabolic Pathways Database (MPW) is a derivative of the larger EMP database containing over 3000 pathway diagrams of primary and secondary metabolism, membrane transport, signal transduction, intracellular traffic, translation, and transcription (Selkov *et al.*, 1998). The assembly of the encoded data and of the pictorial pathway diagrams is an ongoing project centered at the Laboratory of Mathematical Simulation of Multienzyme Systems at the Institute of Theoretical and Experimental Biophysics of the Russian Academy of Sciences, in Pushchino, Russia. These resources can be accessed free of charge by academic and nonprofit users at <http://www.empproject.com>.

### DESCRIPTION OF EMP AND MPW

Each record in the EMP database translates the entire factual content of a journal publication into a structured, indexed, and searchable form. Altogether over 300 subject fields are employed to describe the various categories of biological importance shown in Table 1. A comprehensive description of the subject fields is available at the EMP Web site (2001). The database can be queried by means of a whole text search, taxonomy search, by accession number (i.e., a unique key that identifies a record), or by enzyme commission (EC) number. In addition to the type of information shown in Table 1, each entry provides search links to the MEDLINE (Wheeler *et al.*, 2001), ENZYME (Bairoch, 2000), BRENDA (2001), and LIGAND (Goto *et al.*, 2000) databases for complementary information as well as the metabolic pathway name for easy retrieval from MPW.

<sup>1</sup> To whom correspondence and reprint requests should be addressed. Fax: (814)-865-7846. E-mail: [costas@psu.edu](mailto:costas@psu.edu).

The collection of pathways found in the MPW database originated as the “working notes” of Evgeni Selkov (Selkov *et al.*, 1997) and is currently updated with approximately 1000 pathways a year by the Pushchino team (EMP Web site, 2001). Pathway maps include not only the stoichiometry of the metabolic pathways, but also the substrate and coenzyme specificity of their enzymes, their subcellular locations, required prosthetic groups and cofactors, and the taxonomic occurrence of the pathways. All pathway names are generated in the format substrates-products-function-(coenzymes)-(locations)-[comment]. Figure 1 shows the MPW-generated pathway diagram for glutathione biosynthesis named L-glutamate, L-cysteine–glutathione anabolism (ATP) (cytosol). This pathway was also used as an example in the last issue’s Web site review (Maranas and Burgard, 2001) of the EcoCyc and MetaCyc databases (Karp *et al.*, 2000). As is the case with the EcoCyc and MetaCyc pathways, each compound or reaction number in the MPW pathways is linked to additional information. On the other hand, unlike the EcoCyc pathways which focus solely on the metabolism of *Escherichia coli*, species-specific regulatory structure information is not available through MPW.

TABLE 1

The Number of Subject Fields in each EMP Informational Category

356 Total Subject Fields			
Bibliography	16	Immunochemistry	3
Biochemical Genetics	18	Metabolism	6
Biological Source	19	Miscellaneous	31
Cell Cultivation	19	Modification	12
Enzyme Kinetics and Reaction Mechanisms	21	Non-catalytic Proteins	3
Enzyme Regulation	15	Phenotype	10
Enzyme Structure	30	Physical Chemistry and Spectral Properties	23
Enzyme and Reaction	28	Purification and Assay	21
Equilibrium and Thermodynamics	14	Thermodynamics of the Activated State	4
Host	19	Transporters	44

Cytosol

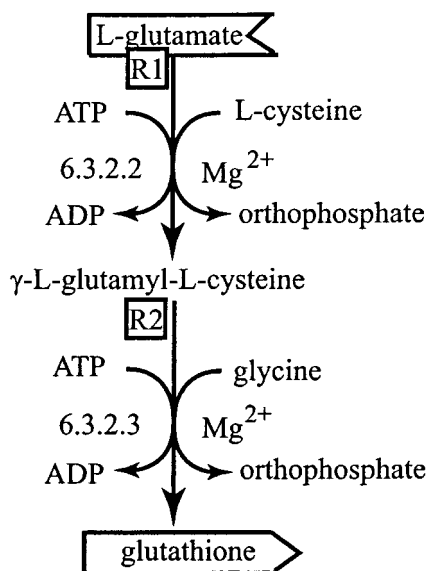


FIG. 1. Glutathione biosynthesis pathway obtained from the EMP/MPW Web site.

STRENGTHS AND WEAKNESSES

A key strength of the EMP database is the depth of enzyme-specific biochemical data for an extensive list of organisms. This information finds applications ranging from the analysis and mathematical simulation of metabolic networks (Goryanin *et al.*, 1999) to metabolic modification (Klinke *et al.*, 2000) and drug development (Black and Hare, 2000; Selzer *et al.*, 2000). Although the high degree of detail for each reference summary may be overwhelming for novice users, a quite useful “help” program and tutorial are available at the EMP Web site. The major strength of the MPW part of the EMP database is its usefulness in metabolic reconstructions. Once the functional assignment of a sequenced genome is complete, genes can be attached to templates from the MPW database allowing the majority of an organism’s metabolic network to be reconstructed *in silico* (Covert *et al.*, 2001). The WIT2 Pro system (Overbeek *et al.*, 2000) of Integrated Genomics, Inc. (<http://igweb.integratedgenomics.com/WIT2/>) is an example of the successful application of MPW for the metabolic reconstruction and annotation of over 250 sequenced genomes (EMP Web site, 2001). However, because the MPW database is not tailored to a particular organism, such as EcoCyc, or its future counterpart, EcoReg (Arkin *et al.*, 2001), it cannot directly incorporate regulatory structure specifics onto the pathway diagrams or provide the genes responsible for the participating enzyme-catalyzed reactions.

Nevertheless, information along these lines can be obtained by querying the EMP database literature sources. In summary, the wealth of biochemical data accessible through the EMP/MPW database makes it a very good starting point for identifying detailed information on enzymes, compounds, or pathways.

REFERENCES

Arkin, A., Conway, T., Karp, P., Saier, M., and Valle, F. (2001). EcoReg: The *Escherichia coli* Regulation Consortium. <http://gobi.lbl.gov/~ecoreg/index.html>.

Bairoch, A. (2000). The ENZYME database in 2000. *Nucleic Acids Res.* **28**, 304–305.

Black, T., and Hare, R. (2000). Will genomics revolutionize antimicrobial drug discovery? *Curr. Opin. Microbiol.* **3**, 522–527.

BRENDA: The Comprehensive Enzyme Information System (2001). <http://www.brenda.uni-koeln.de>.

Covert, M. W., Schilling, C. H., Famili, I., Edwards, J. S., Goryanin, I. I., Selkov, E., and Palsson, B. O. (2001). Metabolic modeling of microbial strains *in silico*. *Trends Biochem. Sci.* **26**, 179–186.

The EMP Project, Inc. Web site (2001). <http://www.empproject.com>.

Goryanin, I., Hodgman, T. C., and Selkov, E. (1999). Mathematical simulation and analysis of cellular metabolism and regulation. *Bioinformatics* **9**, 749–758.

Goto, S., Nishioka, T., and Kanehisa, M. (2000). LIGAND: Chemical database of enzyme reactions. *Nucleic Acids Res.* **28**, 380–382.

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

Klinke, S., Dauner, M., Scott, G., Kessler, B., and Witholt, B. (2000). Inactivation of isocitrate lyase leads to increased production of medium-chain-length poly(3-hydroxyalkanoates) in *Pseudomonas putida*. *Appl. Environ. Microbiol.* **66**, 909–913.

Maranas, C. D., and Burgard, A. P. (2001). Review of EcoCyc and MetaCyc databases. *Metab. Eng.* **3**, 98–99.

Overbeek, R., Larsen, N., Pusch, G. D., D’Souza, M., Selkov, E., Jr., Kyrpides, N., Fonstein, M., Maltsev, N., and Selkov, E. (2000). WIT: Integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Res.* **28**, 123–125.

Selkov, E., Basmanova, S., Gaasterland, T., Goryanin, I., Gretchkin, Y., Maltsev, N., Nenashev, V., Overbeek, R., Panyushkina, E., Pronevitch, L., Selkov, E., Jr., and Yunus, I. (1996). The metabolic pathway collection from EMP: The enzymes and metabolic pathways database. *Nucleic Acids Res.* **24**, 26–28.

Selkov, E., Galimova, M., Goryanin, I., Gretchkin, Y., Ivanova, N., Komarov, Y., Maltsev, N., Mikhailova, N., Nenashev, V., Overbeek, R., Panyushkina, E., Pronevitch, L., and Selkov, E., Jr. (1997). The metabolic pathway collection: An update. *Nucleic Acids Res.* **25**, 37–38.

Selkov, E., Jr., Gretchkin, Y., Mikhailova, N., and Selkov, E. (1998). MPW: The metabolic pathways database. *Nucleic Acids Res.* **26**, 43–45.

Selzer, P. M., Brutsche, S., Wiesner, P., Schmid, P., and Mullner, H. (2000). Target-based drug discovery for the development of novel anti-infectives. *Int. J. Med. Microbiol.* **290**, 191–201.

Wheeler, D. L., Church, D. M., Lash, A. E., Leipe, D. D., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Tatusova, T. A., Wagner, L., and Rapp, B. A. (2001). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **29**, 11–16.