Metabolic Engineering 3, 98–99 (2001)
doi:10.1006/mben.2001.0180, available online at http://www.idealibrary.com on IDEAL®

WEB SITE REVIEW

Review of EcoCyc and MetaCyc Databases

Costas D. Maranas[1] and Anthony P. Burgard

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

Received January 17, 2001; accepted February 1, 2001

INTRODUCTION

The EcoCyc and MetaCyc databases are compilations of online reference sources for metabolic information (Karp *et al*., 2000). EcoCyc is an organism-specific database that provides detailed information about the metabolic pathways, signal-transduction pathways, enzymes, and transport proteins of *Escherichia coli K-12*. MetaCyc is a database that describes the metabolic pathways and enzymes of multiple organisms with an emphasis on microorganisms, but includes some human and other mammalian pathways as well. Both databases are accessed from a WWW server at http://ecocyc.DoubleTwist.com/ecocyc/. All users must apply for an account to access EcoCyc and MetaCyc, and access to the WWW server is free for all individuals at academic or government institutions.

EcoCyc DESCRIPTION

The EcoCyc database describes the biochemical machinery of *E. coli* from the level of individual genes up to the complex metabolic pathways governing cellular metabolism (Karp *et al*., 1998). EcoCyc contains the most up to date annotations of the *E. coli* genome based on new functional characterizations found in the literature and based on an ongoing sequence analysis of the *E. coli* genome by the EcoCyc project. The location of each gene on the *E. coli* chromosome (Blattner *et al*., 1997) and a gene-reaction schematic depicting the relationships between a gene, its protein products, and the reactions catalyzed by those proteins are provided. A detailed description of how to retrieve and interpret the metabolic data contained in EcoCyc is found in Karp (1999). A typical EcoCyc pathway diagram is illustrated in Fig. 1. Each diagram contains links to the compounds, reactions, enzymes, and genes involved in the pathway. Regulated reactions are denoted by a "+" and/or "−" symbol for activation and/or inhibition, and dragging the cursor over either symbol shows which metabolites are involved in the regulation.

MetaCyc DESCRIPTION

The MetaCyc database contains all enzymes, reactions, and metabolic pathways of *E. coli* plus a conglomerate of other organisms (Karp *et al*., 2000). As with EcoCyc, each pathway contains a literature citation to the source where it was reported along with the organism in which it is known to occur based on wet-lab experiments. MetaCyc includes every reaction from the enzyme classification system of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) as well as information regarding the 1900+ metabolites participating in these reactions. A summary of the information contained in both databases is provided in Table 1.

STRENGTHS AND WEAKNESSES

The EcoCyc and MetaCyc databases are accessed using the same software environment, called Pathway Tools,

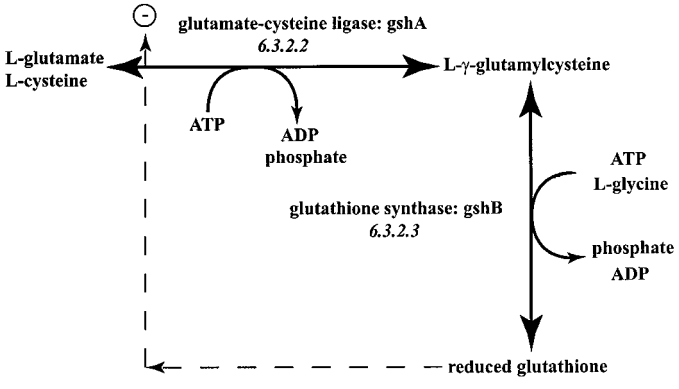

FIG. 1. Glutathione biosynthesis pathway obtained from the EcoCyc Web site.

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

1096-7176/01 $35.00

Metabolic Engineering **3**, 98–99 (2001)
doi:10.1006/mben.2001.0180

TABLE 1

The Number of Objects in the Latest 5.0 Version of the EcoCyc and MetaCyc Databases

	EcoCyc	MetaCyc
Citations	1944	315
Compounds	1868	1963
Enzymatic reactions	946	3786
Enzymes	692	82
Genes	4390	0
Pathways	159	305
Polypeptides	974	194
Protein complexes	452	104
Signaling pathways	20	0
Transport reactions	16	2
Transporters	13	0
tRNAs	79	0

which provides query, editing, and visualization capabilities. The information can be browsed by classification hierarchy or queried by protein, pathway, reaction, compound, or gene. The major strength of the EcoCyc and MetaCyc databases is that their pathways are not merely collections of related reaction steps from a variety of organisms, but are complete pathways elucidated in specific organisms. Thus, these pathways are typically of smaller size than those of other metabolic pathway tools while being much more clear and informative. Another key strength is the wealth of literature citations and the depth of the commentary on each enzyme and pathway. Weaknesses stem from the fact that one database cannot possibly encompass the complete metabolic picture of every sequenced organism. However, links to the BRENDA (BRENDA, 2000) database can be used for more detailed enzyme information, and links to SWISS-PROT (Bairoch *et al*., 2000), GenBank (Benson *et al*., 2000), and KEGG (Kanehisa *et al*., 2000) can be used for function-based retrieval of nucleotide or protein sequences. Summarizing, the EcoCyc and MetaCyc databases are excellent online reference sources on metabolic pathways providing a valuable interface between multiple databases for obtaining any information eluding their scope.

REFERENCES

Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A., and Wheeler, D. L. (2000). GenBank. *Nucleic Acids Res.* **28**, 15–18.

Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997). The complete genome sequence of *Escherichia coli K-12*. *Science* **277**, 1453–1474.

BRENDA: The Comprehensive Enzyme Information System. http://www.brenda.uni-koeln.de/.

Kanehisa, M., and Goto, S. (2000). KEGG: The Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28**, 29–34.

Karp, P. D., Riley, M., Paley, S. M., Pellegrini-Toole, A., and Krummenacker, M. (1998). EcoCyc: Encyclopedia of *Escherichia coli* genes and metabolism. *Nucleic Acids Res.* **26**, 50–53.

Karp, P. D. (1999). Using the EcoCyc database, *in* "Genetics Databases" (M. J. Bishop, Ed.), pp. 269–280, Academic Press, San Diego.

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.