The ENZYME data bank

Amos Bairoch

Department of Medical Biochemistry, University of Geneva, 1 rue Michel Servet, 1211 Geneva 4, Switzerland

ABSTRACT

The ENZYME data bank is a repository of information relative to the nomenclature of enzymes. It is primarily based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) [1] and it contains the following data for each type of characterized enzyme for which an EC (Enzyme Commission) number has been provided:

- **EC number**
- **Recommended name**
- **Alternative names (if any)**
- **Catalytic activity**
- **Cofactors (if any)**
- **Pointers to the SWISS-PROT [2] protein sequence entrie(s) that correspond to the enzyme (if any)**
- **Pointers to human disease(s) associated with a deficiency of the enzyme (if any)**

INTRODUCTION

We believe that the ENZYME data bank can be useful to anybody working with enzymes and that it can be of help in the development of computer programs involved with the manipulation of metabolic pathways.

As stated above the main source for the data in the ENZYME data bank comes from the recommendations of the IUBMB, but additional information has been extracted from the literature. The information concerning human diseases originates from the MIM data base of Victor McKusick [3].

We do not assign EC numbers for newly characterized enzymes, this is the responsibility of the Nomenclature Committee of IUBMB (NC-IUBMB). To contact the committee one should write to:

Prof. K. Tipton
Department of Biochemistry
Trinity College
Dublin 2
Republic of Ireland

He can also be contacted by electronic mail at the following address:

ktipton@vax1.tcd.ie

By phone at the number:
+353−1−702 1608

And by fax at the number:
+353−1−677 2400

The ENZYME data bank is distributed with a form that can be used to fill in the information necessary for the NC-IUBMB to assign an EC number. The commission regularly sends us updates and additions to the nomenclature so that they can be integrated into the data bank in a timely manner.

FORMAT

The entries in the database are structured so as to be usable by human readers as well as by computer programs. An entry in the database is composed of defined line types, each with its own format; they are used to record the various types of data which make up the entry. For standardization purposes the format of ENZYME follows as closely as possible that of the SWISS-PROT [2] protein and EMBL [4] nucleotide sequence databases. Two sample ENZYME entries are shown below:

```
ID    1.14.17.3
DE    PEPTIDYLGLYCINE MONOOXYGENASE.
AN    PEPTIDYL ALPHA-AMIDATING ENZYME.
AN    PEPTIDYLGLYCINE 2-HYDROXYLASE.
CA    PEPTIDYLGLYCINE + ASCORBATE + O(2) = PEPTIDYL(2-
CA    HYDROXYGLYCINE) + DEHYDROASCORBATE + H(2)O.
CF    COPPER.
CC    -!- PEPTIDYLGLYCINES WITH A NEUTRAL AMINO ACID
CC        RESIDUE IN THE PENULTIMATE POSITION ARE THE
CC        BEST SUBSTRATES FOR THE ENZYME.
CC    -!- THE ENZYME ALSO CATALYZES THE DISMUTATION OF
CC        THE PRODUCT TO GLYOXYLATE AND THE
CC        CORRESPONDING DESGLYCINE PEPTIDE AMIDE.
DR    P10731, AMD__BOVIN ; P19021, AMD__HUMAN ; P14925,
      AMD__RAT;
DR    P08478, AMD1__XENLA; P12890, AMD2__XENLA;
//
ID    2.3.1.43
DE    PHOSPHATIDYLCHOLINE--STEROL ACYLTRANSFERASE.
AN    LECITHIN--CHOLESTEROL ACYLTRANSFERASE.
AN    LCAT
AN    PHOSPHOLIPID--CHOLESTEROL ACYLTRANSFERASE.
CA    PHOSPHATIDYLCHOLINE + STEROL = STEROL ESTER +
CA    1-ACYLGLYCEROPHOSPHOCHOLINE.
CC    -!- PALMITOYL, OLEOYL, AND LINOLEOYL CAN BE
CC        TRANSFERRED; A NUMBER OF STEROLS, INCLUDING
CC        CHOLESTEROL, CAN ACT AS ACCEPTOR.
DI    NORUM DISEASE; MIM:245900.
DI    FISH-EYE DISEASE; MIM:136120.
DR    P10480, GCAT__AERHY; P04180, LCAT__HUMAN; P16301,
      LCAT__MOUSE;
DR    P18424, LCAT__RAT ;
//
```

PRACTICAL INFORMATION

a) Content of the current release

Release 16.0 of ENZYME (June 1994) contains information relevant to 3546 enzymes. The data file (ENZYME.DAT)

requires about 1.1 Mb of disk storage space. The database is distributed with a user's manual (ENZUSER.TXT); a file describing the various classes, subclasses and sub-subclasses of enzymes (ENZCLASS.TXT); and a file that describes how the database can be obtained (ENZYME.GET).

b) How to obtain ENZYME

ENZYME is distributed on magnetic tape and on CD-ROM by the EMBL Data Library. For all enquiries regarding the subscription and distribution of ENZYME one should contact:

EMBL Data Library
European Molecular Biology Laboratory
Postfach 10.22.09, Meyerhofstrasse 1
D-69012 Heidelberg, Germany
Telephone: (+49 6221) 387 258
Telefax: (+49 6221) 387 519 or 387 306
Electronic network address: datalib@EMBL-heidelberg.de

ENZYME can be obtained from the EMBL File Server [5]. Detailed instructions on how to make the best use of this service, and in particular on how to obtain ENZYME, can be obtained by sending to the network address netserv@EMBL-heidelberg.de the following message:

HELP
HELP ENZYME

If you have access to a computer system linked to the Internet you can obtain ENZYME using FTP (File Transfer Protocol), from the following file servers:
EMBL anonymous FTP server
Internet address: ftp.EMBL-heidelberg.de (or 192.54.41.33)
NCBI Repository (National Library of Medicine, NIH, Washington D.C., U.S.A.)
Internet address: ncbi.nlm.nih.gov (130.14.20.1)
ExPASy (Expert Protein Analysis System) server, University of Geneva, Switzerland
Internet address: expasy.hcuge.ch (129.195.254.61)
National Institute of Genetics (Japan) FTP server
Internet address: ftp.nig.ac.jp (133.39.16.66)
A version of the database in the ASN.1 data exchange format compatible with the databases and software developed by the National Center for Biotechnology Information (NCBI) [6] is also available on some of the above servers.

You can also browse through ENZYME using various Internet Gopher servers that specialize in biosciences (biogophers) [7]. Gopher is a distributed document delivery service that allows a neophyte user to access various types of data residing on multiple hosts in a seamless fashion.

We plan, in the very near future, to make ENZYME available on the ExPASy World-Wide Web (WWW) molecular biology server [8].

The present distribution frequency is four releases per year. No restrictions are placed on use or redistribution of the data.

REFERENCES

1. Enzyme Nomenclature, Recommendations of the Nomenclature Committe of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes, NC-IUBMB, Academic Press, New-York, (1992).
2. Bairoch A., Boeckmann B. Nucleic Acids Res. 21:3093−3096(1993).
3. McKusick V.A. Mendelian Inheritance in Man. Catalogs of autosomal dominant, autosomal recessive, and X-linked phenotypes; 11th edition; Johns Hopkins University Press, Baltimore, (1994).
4. Rice C.M., Fuchs R., Higgins D.G., Stoehr P.J., Cameron G.N. Nucleic Acids Res. 21:2967−2971(1993).
5. Stoehr P.J., Omond R.A. Nucleic Acids Res. 17:6763−6764(1989).
6. Benson D.A., Boguski M.S., Lipman D.J., Ostell J. Genomics 6:389−391(1990).
7. Gilbert D. Trends Biochem. Sci. 18:107−108(1993).
8. Appel R.D., Bairoch A., Hochstrasser D.F. Trends Biochem. Sci. 19:258−260(1994).