

EzCatDB: the Enzyme Catalytic-mechanism Database

Nozomi Nagano*

Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST), 2-41-6 Aomi, Koto-ku, Tokyo 135-0064, Japan and PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho Kawaguchi-shi, Saitama 332-0012, Japan

Received August 15, 2004; Revised and Accepted October 8, 2004

ABSTRACT

The EzCatDB (Enzyme Catalytic-mechanism Database) specifically includes catalytic mechanisms of enzymes in terms of sequences and tertiary structures of enzymes, and proposed catalytic mechanisms, along with ligand structures. The EzCatDB groups enzyme data in the Protein Data Bank (PDB) and the SWISS-PROT database with identical domain compositions, Enzyme Commission (EC) numbers and catalytic mechanisms. The EzCatDB can be queried by the type of catalytic residue, name and type of ligand molecule that interacts with an enzyme as a cofactor, substrate or product. It can provide literature information, other database codes and EC numbers. The EzCatDB provides ligand annotation for enzymes in the PDB as well as literature information on structure and catalytic mechanisms. Furthermore, the EzCatDB also provides a hierarchic classification of catalytic mechanisms. This classification incorporates catalytic mechanisms and active-site structures of enzymes as well as basic reactions and reactive parts of ligand molecules. The EzCatDB is available at <http://mbs.cbrc.jp/EzCatDB/>.

INTRODUCTION

The organization of enzyme data is chaotic at various research stages. Complete structures have been determined for some enzymes. Nevertheless, for many enzymes, structural information is either completely unavailable or available only for non-catalytic domains. Even for those tertiary structures of catalytic domains that are available, it is extremely difficult to annotate catalytic sites and to propose a catalytic mechanism without site-directed mutagenesis results or structures of liganded forms of enzymes. In particular, the latter would lend the user some insight into enzyme catalysis. In contrast, catalytic mechanisms could be proposed for those enzymes

whose catalytic sites are annotated, and whose complex structures with ligand molecules have been determined. For these reasons, enzyme structure data should be sorted in terms of the research stage in which they are.

Meanwhile, the classification of enzymes has been defined by the Enzyme Commission (EC) (1), based mainly on the whole chemical structures of the substrates and products, and on the cofactors involved. However, the EC classification neglects protein sequence and structure information, which are extremely important in catalytic mechanisms. Non-homologous enzymes can catalyze similar reactions, while homologous enzymes sometimes adopt different strategies in terms of catalytic mechanisms (2,3). The EC classification does not reflect such detailed mechanisms in terms of protein structures. Moreover, some enzymes catalyze complex reactions comprising several basic reactions, such as oxidation/reduction, hydrolysis, transfer/elimination/addition of some groups and isomerization, which are very difficult to express using only one EC number.

To date, numerous enzyme structures have been determined using X-ray crystallography and NMR. They have been deposited in the Protein Data Bank (PDB) (4). Furthermore, various enzyme databases have been developed based on EC numbers and enzyme sequence/structure information: BRENDA (5,6), ExPASy (7), KEGG (8) and the Catalytic Site Atlas (CSA) (9). Although the BRENDA, ExPASy and KEGG databases annotate various enzyme data, including ligand molecules (cofactors, substrates, products, inhibitors and activators), reaction formulae and metabolic pathways, their detailed catalytic mechanisms are neither annotated nor classified in terms of sequence or structure information. The CSA database focuses on the catalytic residues of enzymes that are involved in catalysis. It also provides related literature. Nevertheless, the catalytic mechanisms are whole systems of active site residues and ligand molecules, including cofactors, substrates and products, which act interactively on the enzyme molecules. In some enzymes, cofactors or some moieties of substrates are involved directly in catalytic groups, instead of protein residues.

The EzCatDB is a novel enzyme catalytic-mechanism database. It includes a search system that can retrieve some

*Tel: +81 3 3599 8043; Fax: +81 3 3599 8081; Email: n.nagano@aist.go.jp

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use permissions, please contact journals.permissions@oupjournals.org.

enzyme groups specifically by their respective types of catalytic residues, names or ligand molecule types that interact with enzymes as cofactors, substrates or products. It also allows querying of literature information, other database entry codes and a specific research stage. Furthermore, the EzCatDB specifically addresses the catalytic mechanisms of enzymes. It is intended to classify them based on structural information of enzymes and ligand molecules, and proposed mechanisms. The EzCatDB is available at <http://mbs.cbrc.jp/EzCatDB/>.

MAIN FEATURES OF EzCatDB

The contents of respective entries in the EzCatDB are summarized in Table 1. Depending on the research stage and other situations, additional information can be included. Main features of the EzCatDB are described in subsequent sections.

Search page

A search for enzyme data can be specified in various ways. Each EzCatDB entry has been assigned EC numbers and CATH numbers that represent hierarchic classification of domain structures (10). Enzyme data can be retrieved using those numbers. Although each number comprises four levels, some levels can be abbreviated for the search. With a specific entry code for the PDB (4), SWISS-PROT/ExpASY (7,11) and KEGG database (8), the enzyme data related to the database entries can be retrieved. Moreover, types of active-site residues can be specified along with ligand names or types that are related to enzymes as cofactors, substrates or products. An author name and key words in related literature can also be specified in addition to the code for literature in the PubMed database (12). Moreover, enzyme data can be retrieved with specific stages of data: 'Catalytic domain' determined/annotated; 'Active-site residues' annotated; complex structures with

ligand molecules determined; literature reporting catalytic mechanisms; catalytic mechanism classified in the EzCatDB; and three-dimensional (3D) models of catalytic mechanism constructed. In addition, any of those search items can be combined to create a more specific search.

Table of annotated ligand for the PDB data

The ligand information will be useful when considering catalytic mechanisms of the enzymes. Therefore, annotation of ligand molecules bound to the enzyme structures was performed manually for each PDB entry, then tabulated for corresponding cofactors, substrates and products. The table includes intermediate and transition-state data. Analogous ligand molecules were analyzed in addition to 'native' ligand molecules, then annotated as 'analogues' of cofactors, substrates, products or intermediates. Thereby, those entries with ligand molecules can be selected easily among many PDB entries. An example of the table of annotated ligands is shown in Figure 1. Annotated ligand data are linked to the compound data in KEGG COMPOUND (8) and PDBsum Ligand data (13), whereas the PDB entries are linked to the PDBsum (13). Moreover, in the search page noted above, those enzyme data with any PDB data, whose ligand molecules could be annotated as cofactors, substrates, products, intermediates or analogues of native ligand molecules, can be retrieved. Meanwhile, entries without any PDB data, for which the complex structures with ligand molecules have been determined, can also be retrieved. This capability will be useful for structural biologists as they search for target enzymes.

Collection of literature information on catalytic mechanisms

Information with reference to previous studies, especially those studies related to protein structures and catalytic mechanisms, have been collected for each enzyme in this

Table 1. Contents for each entry in EzCatDB

Information item	Description ^a	Entries that contain information (number of entries ^b)
DB code	Entry ID in this EzCatDB database	All entries (302)
RLCP catalysis type	Classification of catalytic mechanism	Entries with catalytic mechanism classified (163)
CATH	Domain composition described with CATH number (10)	All entries
Catalytic domain	Catalytic domain annotated for CATH domains	Entries with catalytic domain assigned (283)
EC	Enzyme commission number (1) and links to KEGG enzyme data (8)	All entries
3D-view of catalysis	3D-model of catalytic mechanisms	Entries with constructed 3D-model of catalytic mechanism (7)
Related DB codes	Links to other entries in the same family	Entries with links to related data within EzCatDB (homologous data)
Enzyme name	Nomenclature in the SWISS-PROT (11) and KEGG (8)	All entries
KEGG pathways	Related metabolic pathways with links to KEGG data (8)	Entries with links to related data in other database (184)
SWISS-PROT information	Basic information on activity, cofactor and links to SWISS-PROT (11)	Entries with links to related data in other database
Table of ligand annotation	Annotation of ligand molecules in the PDB structures	All entries
Active-site residues	Catalytic residues, cofactor-binding residues and modified residues	Entries with active site residues annotated (276)
Catalytic mechanism	References on catalytic mechanism are annotated	Entries with literature on catalytic mechanisms (274)
References	List of related references with links to PubMed (12)	All entries
Comments	Additional information for catalytic mechanisms based on literature	Entries with additional information (216)

^aReference numbers are given in parentheses.

^bNumber of entries may change through updates.

	Cofactors	Substrates			Products			intermediates
KEGG-id	C00305	C00002	C00055	C00239	C00008	C00112	C00705	
Compound	Magnesium	ATP	CMP	dCMP	ADP	CDP	dCDP	
Type	divalent metal (Ca ²⁺ , Mg ²⁺)	amine group, nucleotide	amine group, nucleotide	amine group, nucleotide	amine group, nucleotide	amine group, nucleotide	amine group, nucleotide	
1ckeA	Unbound	Unbound	Unbound	Unbound	Unbound	Unbound	Unbound	Unbound
1gf9A	Bound: MG	Unbound	Bound: C5P	Unbound	Bound: ADP	Unbound	Unbound	Transition-state-analogue: ALF
1ukd	Unbound	Analogue: UP5(ATP)	Analogue: UP5(CMP)	Unbound	Unbound	Unbound	Unbound	Unbound
1uke	Bound: MG	Analogue: UP5(ATP)	Analogue: UP5(CMP)	Unbound	Unbound	Unbound	Unbound	Unbound
2cmkA	Unbound	Unbound	Unbound	Unbound	Unbound	Bound: CDP	Unbound	Unbound
2ukd	Bound: MG	Unbound	Bound: C5P	Unbound	Bound: ADP	Unbound	Unbound	Unbound
3ukd	Bound: MG	Unbound	Bound: C5P	Unbound	Bound: ADP	Unbound	Unbound	Transition-state-analogue: AF3
4ukd	Bound: MG	Unbound	Unbound	Unbound	Bound: ADP	Analogue: UDP	Unbound	Transition-state-analogue: BF2
5ukdA	Bound: MG	Unbound	Bound: C5P	Unbound	Bound: ADP	Unbound	Unbound	Transition-state-analogue: AF3

Figure 1. Table of annotated ligand for the PDB entries. This table is for cytidylate kinases (EC 2.7.4.14). The PDB entries are listed at left, whereas the ligand names are listed across the top row. Types of ligands are also classified. Whereas the PDB entries are linked to the PDBsum (13), the annotated ligand data are linked to KEGG COMPOUND data (8) and PDBsum ligand data (13).

database. A link to the abstract page allows users to access that literature, which is maintained in the PubMed database (12). The literature on catalytic mechanisms was annotated manually.

Hierarchic classification of catalytic mechanisms, RLCP

A novel classification of enzyme catalytic mechanisms, which clusters catalytic mechanisms at four levels, has also been developed:

- (i) Basic Reaction (R),
- (ii) Ligand group involved in catalysis (L),
- (iii) Type of Catalytic mechanism (C),
- (iv) Residues/cofactors located on Proteins (P).

'Basic Reaction' represents reaction types such as hydrolysis, phosphorolysis and transfer, which are mostly related to the primary number of each EC number. Reactive groups of substrates are classified at the second level (L). Whereas EC numbers have been based on whole chemical formulae of substrates and products, only the reactive parts of ligand molecules are considered at the 'L' level. At the third level (C), the catalytic mechanisms are classified systematically based on the types of catalysts, such as nucleophile, acid, base, stabilizer and modulator; existence of cofactors; S_N2/S_N1 (or associative/dissociative) reactions; and also on the way in which these catalytic groups function interactively. Table 2 summarizes these determinants of catalytic-mechanism classification. The types of catalysts are classified based on the reported definition by Bartlett *et al.* (14), which have been modified slightly.

Types of catalytic residues and cofactors with the ligating residues are classified at the fourth level (P). Even though enzymes have got the same reaction mechanisms, they can have different catalytic residues. Information on catalytic

Table 2. Summary of major determinants of catalytic-mechanism classification for hydrolysis reactions

Determinants	Possible types for determinants
1st Nucleophile	Water ^a /substrate ^b /catalytic residue
Catalytic groups	Only cofactor/enzyme residues and substrates/both cofactor and enzyme residues and substrates
S_N2/S_N1 reactions	S_N2 (associative) reaction/ S_N1 (dissociative) reaction/unclear
Base	Catalytic residue/substrate-assisted/non-existent
Acid	Catalytic residue/substrate-assisted/non-existent
Target of 1st base	Water ^c /nucleophile, other than water ^d /other site/non-existent
Stabilizer	Catalytic residue/substrate-assisted/non-existent
Modulator	Catalytic residue/substrate-assisted/non-existent

^aFor transfer reactions, this type will be replaced by 'acceptor group of substrate'. For phosphorolysis reaction, this will be 'phosphate'.

^bFor transfer reactions, this type will be replaced by 'group of substrate, other than acceptor group'.

^cFor transfer reactions, this type will be replaced by 'transferred group of substrate'. For phosphorolysis reaction, this will be 'phosphate'.

^dFor transfer reactions, this type will be replaced by 'nucleophile (catalytic residue or acceptor group of substrate)'.

mechanism and active sites has been manually collected from the related literature, the PDB and the SWISS-PROT/ExpASY data (1,7,11). Whereas the conventional EC numbers are a kind of nomenclature of enzymes, the RLCP classifies the catalytic reactions. Therefore, some enzymes could have several RLCP classes, rather than only one class, if they catalyze more than one reaction. Furthermore, whereas whole chemical structures of substrates and products were examined for EC classification, the RLCP specifically addresses the reactive part of the ligand molecules. For that reason, the first two levels of the RLCP classification, the 'R' and 'L' levels, can be correlated with the EC numbers.

Hydrolysis and transfer reactions have been the primary focus of the RLCP classification effort, as hydrolases and transferases present the majority of enzymes. Moreover, elucidation of basic reactions engenders better understanding of complicated reactions that comprise several reactions. For example, many ligases catalyze two successive transfer reactions: phosphoryl transfer and acyl group transfer. Indeed, active sites and catalytic mechanisms of ligases are quite similar to those of kinases (phosphoryl transferases). Other reaction mechanisms such as elimination (EC numbers; 4.-.-) and isomerization (EC numbers; 5.-.-) will be available soon.

An example of the RLCP view is shown in Figure 2. This RLCP site is linked to the EzCatDB entries, catalytic mechanisms of which have been classified, but this page can also be viewed independently from the EzCatDB, at <http://mbs.cbrc.jp/EzCatDB/RLCP/>.

Three-dimensional view of catalytic mechanisms

Proposed catalytic mechanisms of the enzymes tend to be indicated in schematic diagrams in papers, rather than in 3D graphical diagrams. To elucidate detailed functions of enzymes, 3D models of the catalytic mechanisms should be constructed and presented. For some enzyme data, their catalytic mechanisms were modeled from the PDB structure data, based on the

proposed mechanism reported in the literature. Those model structures were constructed using the Insight II/Discover application (Accelrys Inc.). Such mechanism models can be viewed with a 3D graphics application, 3D-EzCat, which was developed for this database, on a Java applet platform with Java3D. The application allows users to load and view a catalytic mechanism as a series of snapshots, with 3D arrows indicated between catalytic atoms, and with half-bonds and one-and-a-half-bonds as well as double-bonds between the reactive atoms to show the processes of transition-states or intermediates. Figure 3 shows an example of the 3D-view of catalytic mechanism. Such 3D-catalytic-mechanisms will be helpful to elucidate catalysis and to design novel drugs that can inhibit or activate the enzymes.

Although the Insight II/Discover application is usually used for homology modeling, it can also be used for 3D-catalytic-mechanism modeling. This application has several kinds of force-field, such as AMBER (15) and extensible systematic force field (ESFF) (16). The ESFF was used for the 3D-catalytic-mechanism modeling because it covers many kinds of atoms, including even magnesium ions and other metals, during optimization of structures. Some metals are essential for reactions such as phosphoryl transfer, so that the ESFF is the most useful force-field. The 3D-coordinates from the PDB structures with substrate molecules, or their analogous molecules, usually started as initial structures. Using the

RLCP
Hierarchic Classification of Catalytic Mechanisms

[TOP](#)
[SEARCH](#) [tutorial](#)
[PDB entry list](#)
[RLCP](#) [tutorial](#)
[about EzCatDB](#)

R 1.-.- : Hydrolysis (bond cleavage by water) (Reaction)

L 1.13.-.- : Amide bond (One of Carbonyl Ester bonds; including peptide bond) (Ligand group involved)

C 1.13.30000.- : Trypsin-like reaction (Nucleophile/Acid-Base/Modulator/Stabilizer); covalent bond between enzyme and substrate; followed by protonation of the leaving group; hydrolysis by activated water by base (Catalytic mechanism)

P 1.13.30000.10 : Ser/His/Asp + mainchain amide (Residues/cofactors in Protein)

1st Nucleophile : a catalytic residue
 Catalytic groups : groups in residue/substrate
 General Base : a catalytic residue
 General Acid : a catalytic residue

Related Enzymes

There are 9 entries in this class.

- [S00295](#)
- [S00353](#)
- [S00374](#)
- [D00189](#)
- [D00424](#)
- [D00430](#)
- [M00122](#)
- [M00133](#)
- [M00139](#)

Figure 2. RLCP page for serine hydrolases. Catalytic mechanisms are clustered at four levels in the RLCP classification. Related enzyme entries are listed at the bottom of this page.

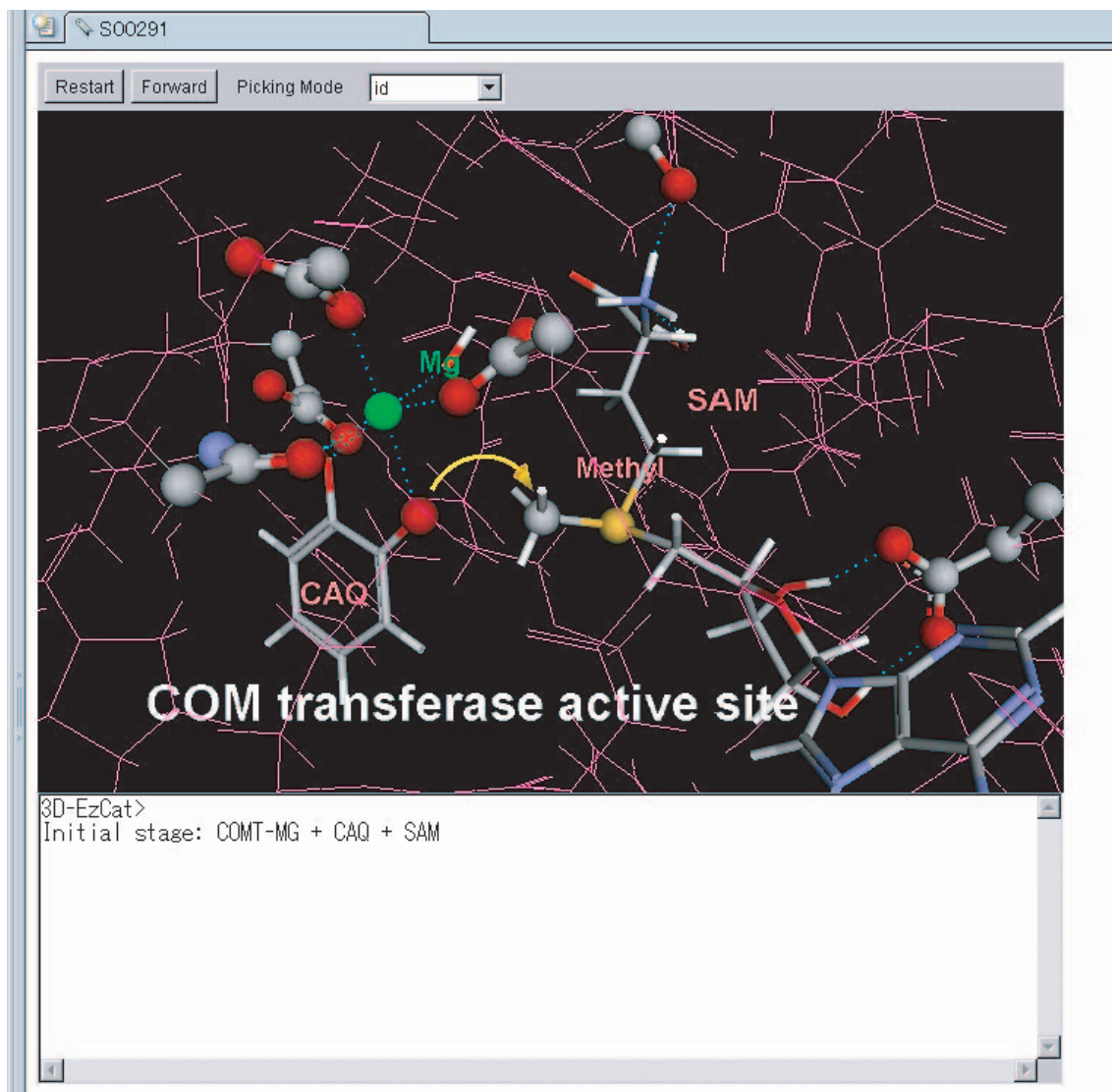


Figure 3. A snapshot of 3D-view of catalytic mechanism (3D-EzCat applet): in this snapshot, the yellow arc arrow indicates that the oxygen atom of 'CAQ' molecule makes a nucleophilic attack on the methyl group of 'SAM' molecule.

'Biopolymer' module of InsightII, some groups and atoms can be modified. Hydrogen atoms can be added to the PDB structures easily. Considering the charge and protonation states of active-site residues (based on literature information), the residue types (protonated or non-protonated residue) can also be replaced. During optimization, the distances, angles and dihedral angles between reacting atoms and groups can be specified to make them closer to each other, using the 'Restraint' option of the software. Use of the 'Restraint' option is especially important to model transition-state structures, which have energetically unlikely conformations.

DATA IN EzCatDB

Enzyme data with tertiary structures deposited in the PDB, to which EC numbers and CATH domain classification (version 2.4) had been assigned, have been analyzed and annotated in

the EzCatDB (4,10). The EzCatDB groups enzyme data in the PDB and the SWISS-PROT database with identical domain compositions, EC numbers and catalytic mechanisms (4,7,11). When the EC numbers in the PDB entries are inconsistent with those in SWISS-PROT data, those numbers annotated in SWISS-PROT/ExPASy data are assigned to the EzCatDB entries (4,7,11). Some enzyme sequences from different organisms can be homologous. Those PDB data and their corresponding SWISS-PROT codes from different organisms can be included in the same EzCatDB entries if they have the same EC numbers, catalytic mechanisms and domain compositions, and structures in terms of the CATH classification (1,4,7,10,11). In the case of a multienzyme having more than one chain, non-homologous SWISS-PROT data can be included in the same EzCatDB entries if they are found to function together as enzyme complexes (7,11). Moreover, in the case of some enzymes that have more than one EC number, the corresponding EzCatDB entries could have more than one

EC number (1). In contrast, some enzymes catalyze complicated reactions that are composed of several 'Basic Reactions'. For example, many ligases (EC 6.-.-.-) catalyze two successive transfer reactions. In such cases, EzCatDB entries could have more than one RLCP class. In this manner, clustering allows some EzCatDB entries to contain more than one SWISS-PROT code, EC number or RLCP class.

Currently, ~300 enzyme entries, mainly for hydrolases (EC numbers; 3.-.-.-) and transferases (EC numbers; 2.-.-.-) related to ~1500 PDB data, are deposited in the EzCatDB. Regarding 3D-views of catalytic mechanisms, such models have been prepared for seven entries. The EzCatDB is updated on a weekly basis at the rate of roughly 10 entries a week.

FUTURE PERSPECTIVES

Only hydrolysis and transfer reactions have been classified in the RLCP. However, the reactions that are catalyzed by ligases (EC numbers; 6.-.-.-) and some isomerases (EC numbers; 5.-.-.-) that catalyze intramolecular transfer reactions can be included in the category. Other reactions that are catalyzed by oxidoreductases, lyases and isomerases must be classified into several types of 'Basic Reaction' in the RLCP. Through the classification of catalytic mechanisms and ligand annotation, the relationships between the active-site structures and functions of enzymes can be elucidated in detail.

ACKNOWLEDGEMENTS

I would like to thank Yuko Hasegawa, Keitarou Nonaka, Kenji Morita, Munehiro Sugiyama and Junko Some, who assisted in the annotation of enzyme data and collection of the literature. I also would like to thank Dr Yutaka Akiyama and Dr Tamotsu Noguchi for supporting this project, and Dr Paul Horton for correcting the manuscript. This project was supported by a grant from PRESTO, organized by the Japan Science and Technology Corporation (JST).

REFERENCES

1. Webb,E.C. (1992) *Enzyme Nomenclature 1992. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology*. Academic Press, New York, NY.
2. Todd,A.E., Orengo,C.A. and Thornton,J.M. (2001) Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.*, **307**, 1113–1143.
3. Todd,A.E., Orengo,C.A. and Thornton,J.M. (2002) Plasticity of enzyme active sites. *Trends Biochem. Sci.*, **27**, 419–426.
4. Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
5. Schomburg,I., Chang,A. and Schomburg,D. (2002) BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.*, **30**, 47–49.
6. Schomburg,I., Chang,A., Hofmann,O., Ebeling,C., Ehrentreich,F. and Schomburg,D. (2002) BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem. Sci.*, **27**, 54–56.
7. Gasteiger,E., Gattiker,A., Hoogland,C., Ivanyi,I., Appel,R.D. and Bairoch,A. (2003) ExpASY: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.*, **31**, 3784–3788.
8. Kanehisa,M., Goto,S., Kawashima,S., Okuno,Y. and Hattori,M. (2004) The KEGG resources for deciphering the genome. *Nucleic Acids Res.*, **32**, D277–D280.
9. Porter,C.T., Bartlett,G.J. and Thornton,J.M. (2004) The Catalytic Site Atlas: a resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res.*, **32**, D129–D133.
10. Orengo,C.A., Michie,A.D., Jones,S., Jones,D.T., Swindells,M.B. and Thornton,J.M. (1997) CATH—a hierarchic classification of protein domain structures. *Structure*, **5**, 1093–1108.
11. Bairoch,A., Boeckmann,B., Ferro,S. and Gasteiger,E. (2004) SWISS-PROT: juggling between evolution and stability. *Brief Bioinformatics*, **5**, 39–55.
12. McEntyre,J. and Lipman,D. (2001) PubMed: bridging the information gap. *CMAJ*, **164**, 1317–1319.
13. Laskowski,R.A. (2001) PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res.*, **29**, 221–222.
14. Bartlett,G.J., Porter,C.T., Borkakoti,N. and Thornton,J.M. (2002) Analysis of catalytic residues in enzyme active sites. *J. Mol. Biol.*, **324**, 105–121.
15. Weiner,S.J., Kollman,P.A., Case,D.A., Singh,U.C., Ghio,C., Alagona,G., Profeta,S., Jr and Weiner,P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, **106**, 765–784.
16. Shi,S., Yan,L., Yang,Y., Fisher-Shaulsky,J. and Thacher,T. (2003) An extensible and systematic force field, ESFF, for molecular modeling of organic, inorganic, and organometallic systems. *J. Comput. Chem.*, **24**, 1059–1076.