

Available online at www.sciencedirect.com



BBRC

Biochemical and Biophysical Research Communications 364 (2007) 53-59

www.elsevier.com/locate/ybbrc

EzyPred: A top-down approach for predicting enzyme functional classes and subclasses

Hong-Bin Shen¹, Kuo-Chen Chou^{*}

Gordon Life Science Institute, San Diego, CA 92130, USA Institute of Image Processing & Pattern Recognition, Shanghai Jiaotong University, Shanghai 200030, China

> Received 15 September 2007 Available online 2 October 2007

Abstract

Given a protein sequence, how can we identify whether it is an enzyme or non-enzyme? If it is, which main functional class it belongs to? What about its sub-functional class? It is important to address these problems because they are closely correlated with the biological function of an uncharacterized protein and its acting object and process. Particularly, with the avalanche of protein sequences generated in the Post Genomic Age and relatively much slower progress in determining their functions by experiments, it is highly desired to develop an automated method by which one can get a fast and accurate answer to these questions. Here, a top–down predictor, called **EzyPred**, is developed by fusing the results derived from the functional domain and evolution information. **EzyPred** is a 3-layer predictor: the 1st layer prediction engine is for identifying a query protein as enzyme or non-enzyme; the 2nd layer for the main functional class; and the 3rd layer for the sub-functional class. The overall success rates for all the three layers are higher than 90% that were obtained through rigorous cross-validation tests on the very stringent benchmark datasets in which none of the proteins has $\geq 40\%$ sequence identity to any other in a same class or subclass. **EzyPred** is freely accessible at http://chou.med.harvard.edu/bioinf/EzyPred/, by which one can get the desired 3-level results for a query protein sequence within less than 90 s.

Keywords: Enzyme functional class; Evolution; Functional domain; Fusion approach; EzyPred web-server

For a newly-found protein sequence the most interesting thing people wish to know is about its biological function, and hence the following questions are often asked: Is the query protein an enzyme or non-enzyme? If it is, which main functional class does it belong to? Or going further deeper, what about its sub-functional class? Although the answers to these questions can be found by conducting various biochemical experiments, it is both time-consuming and costly to do so solely by experimental approaches. With the explosion of newly-found protein sequences entering into databanks in the Post Genomic Age, it has become a major challenge to bridge the gap between the number of newly generated sequence entries and the number of functionally characterized protein entries. Actually, some efforts were made in this regard [1,2]. However, the investigation in [1] was limited within the scope of oxidoreductases while that in [2] limited among the main enzyme classes only. Particularly, no web-server was provided in either [1] or [2]. The present study was initiated in an attempt to develop a topdown approach to solve all these problems and make it accessible to the vast majority of experimental scientists by providing a user-friendly web-server.

Materials and methods

Materials

The ENZYME database at http://www.expasy.org/enzyme/ (released on 01-May-2007) was used to construct the benchmark datasets for the enzyme main functional classes and their subclasses.

^{*} Corresponding author. Present address: Gordon Life Science Institute, San Diego, CA 92130, USA. Fax: +1 858 484 1018.

E-mail addresses: hbshen@crystal.harvard.edu (H.-B. Shen), kcchou@ gordonlifesciene.org (K.-C. Chou).

¹ Present address: BCMP, Harvard Medical School, Boston, MA 02115, USA.

⁰⁰⁰⁶⁻²⁹¹X/\$ - see front matter @ 2007 Elsevier Inc. All rights reserved. doi:10.1016/j.bbrc.2007.09.098

Main functional classes. According to their main functions or enzyme commission (EC) numbers [3], enzymes are classified into the following six main classes: (1) oxidoreductase (EC.1), (2) transferase (EC.2), (3) hydrolase (EC.3), (4) lyase (EC.4), (5) isomerase (EC.5), and (6) ligase (EC.6). To get the high quality benchmark dataset, the data were curated strictly according to the following procedures. Step 1: to avoid fragment data, those enzymes whose sequences were annotated with "fragment" or had less than 50 amino acids were excluded. Step 2: for the uniqueness, those enzymes that occur in two or more classes were excluded. Step 3: to reduce the homology bias, a redundancy cutoff was operated by an in-house program to winnow those sequences which have $\ge 40\%$ sequence identity to any other in a same functional class. Finally, 9832 enzyme sequences classified into six main functional classes were obtained. Meanwhile, to construct a non-enzyme benchmark dataset, 9850 nonenzyme protein sequences were randomly collected from Swiss-Prot at http://www.ebi.ac.uk/swissprot/ (version 52.0 released on 7-Mar-2007); these non-enzyme proteins were also subject to the same screening procedures to exclude the fragment and redundancy sequences. For the convenience of formulation, suppose the benchmark dataset thus obtained is denoted by S, which consists of the enzyme dataset \mathbb{S}^{ezy} and the non-enzyme dataset $\mathbb{S}^{non-ezy};$ i.e.,

$$\begin{cases} \mathbb{S} = \mathbb{S}^{ezy} \cup \mathbb{S}^{non-ezy} \\ \mathbb{S}^{ezy} = \mathbb{S}_1^{ezy} \cup \mathbb{S}_2^{ezy} \cup \mathbb{S}_3^{ezy} \cup \mathbb{S}_4^{ezy} \cup \mathbb{S}_5^{ezy} \cup \mathbb{S}_6^{ezy} \end{cases}$$
(1)

where \cup is the symbol for union in the set theory, \mathbb{S}_2^{ezy} is the EC.1 subset containing 1618 oxidoreductase sequences, \mathbb{S}_2^{ezy} the EC.2 subset of 3450 transferase sequences, \mathbb{S}_5^{ezy} the EC.3 subset of 2791 hydrolase sequences, \mathbb{S}_5^{ezy} the EC.4 subset of 679 lyase sequences, \mathbb{S}_5^{ezy} the EC.5 subset of 518 isomerase sequences, and \mathbb{S}_6^{ezy} the EC.6 subset of 776 ligase sequences. The 9832 enzyme sequences classified into six subsets as well as the 9850 non-enzyme sequences are provided in Online Supporting Information A.

Sub functional classes. To reflect the enzyme functions at a deeper level, each of the aforementioned six main enzyme classes has been further classified into many sub-classes [3]. For instance, the oxidoreductase family of EC.1 has 18 sub-classes denoted by EC.1.1, EC.1.2, EC.1.3,...,EC.1.18 and their functions are briefly described in a highly condensed Fig. 1. To construct the subclass benchmark dataset for each of



Fig. 1. A schematic drawing to use tree branches to classify enzyme and non-enzyme as well as the six main functional classes of enzymes and their subclasses.

the six main enzyme families, the same procedures in the "Main functional classes" section were used. However, if the number of enzyme sequences thus obtained for a subclass was less than 10, the subclass and the sequences therein were left out because of lacking statistical significance. Similar to Eq. (1), the benchmark datasets thus obtained can be formulated as

$$\begin{cases} S_{1}^{exy} = S_{1.1}^{exy} \cup S_{1.2}^{exy} \cup S_{1.4}^{exy} \cup S_{1.4}^{exy} \cdots \cup S_{1.18}^{exy} \\ S_{2}^{exy} = S_{2.1}^{exy} \cup S_{2.2}^{exy} \cup S_{2.3}^{exy} \cup S_{2.4}^{exy} \cdots \cup S_{2.8}^{exy} \\ S_{3}^{exy} = S_{3.1}^{exy} \cup S_{3.2}^{exy} \cup S_{3.4}^{exy} \cup S_{3.5}^{exy} \cup S_{3.6}^{exy} \\ S_{4}^{exy} = S_{4.1}^{exy} \cup S_{4.2}^{exy} \cup S_{4.3}^{exy} \cup S_{4.6}^{exy} \cup S_{4.99}^{exy} \\ S_{5}^{exy} = S_{5.1}^{exy} \cup S_{5.2}^{exy} \cup S_{5.4}^{exy} \cup S_{5.4}^{exy} \cup S_{5.99}^{exy} \\ S_{6}^{exy} = S_{6.1}^{exy} \cup S_{6.2}^{exy} \cup S_{6.4}^{exy} \cdots \cup S_{6.6}^{exy} \end{cases}$$
(2)

where $S_{1.1}^{ezy}$ represents the EC.1.1 subset with the function acting on the CH–OH group of donors (Fig. 1), and so forth. Note that in Eq. (2) some subsets such as $S_{3.3}^{ezy}$ and $S_{4.5}^{ezy}$ are missing because the numbers of their sequences obtained through the above procedures were less than 10. All the sequences for each of the subsets in Eq. (2) are provided in Online Supporting Information B.

Method

To develop a top-down predictor, a novel technique was introduced by fusing the FunD (Functional Domain) approach and the Pse-PSSM (Pseudo Position-Specific Scoring Matrix) approach.

Functional domain (FunD) composition. Proteins often contain several modules or domains, each with a distinct evolutionary origin and function. Based on such a fact, several FunD databases were developed, such as SMART [4], COG [5], KOG [5], CDD [6]. Pfam database is a large collection of multiple sequence alignments and hidden Markov models currently covering 8958 common protein domains and families [7]. With each of the 8958 domain sequences as a vector-base, a given protein sample can be defined as an 8958-D (dimensional) vector according to the following procedures. *Step 1*: use RPS-BLAST (Reverse PSI-BLAST) program [8] to compare the protein sequence with each of the 8958 domain sequences in Pfam database. *Step 2*: if the significance threshold value (expect value) is ≤ 0.01 for the *i*th profile in Pfam meaning a "hit" is found, then the *i*th component of the protein in the 8958-D space is assigned 1; otherwise, 0. *Step 3*: the protein sample **P** in the FunD space can thus be formulated as

$$\mathbf{P}_{\mathrm{FunD}} = \begin{bmatrix} \mathbb{D}_1 & \mathbb{D}_2 & \cdots & \mathbb{D}_i & \cdots & \mathbb{D}_{8958} \end{bmatrix}^{\mathrm{T}}$$
(3)

where T is the transpose operator, and

$$\mathbb{D}_{i} = \begin{cases} 1, & \text{when a hit is found for } \mathbf{P} \text{ in the } i\text{th profile of Pfam} \\ 0, & \text{otherwise} \end{cases}$$
(4)

Pseudo position-specific scoring matrix (Pse-PSSM). To incorporate the evolution information of proteins, the PSSM (Position-Specific Scoring Matrix) [8] was used; i.e., according to the concept of PSSM, the sample of a protein \mathbf{P} can be represented by:

$$\mathbf{P}_{\text{PSSM}} = \begin{bmatrix} \mathbb{V}_{1 \to 1} & \mathbb{V}_{1 \to 2} & \cdots & \mathbb{V}_{1 \to 20} \\ \mathbb{V}_{2 \to 1} & \mathbb{V}_{2 \to 2} & \cdots & \mathbb{V}_{2 \to 20} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbb{V}_{i \to 1} & \mathbb{V}_{i \to 2} & \cdots & \mathbb{V}_{i \to 20} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbb{V}_{L \to 1} & \mathbb{V}_{L \to 2} & \cdots & \mathbb{V}_{L \to 20} \end{bmatrix}$$
(5)

where $\mathbb{V}_{i\to j}$ represents the score of the amino acid residue in the *i*th position of the protein sequence being changed to amino acid type *j* during the evolution process. Here, the numerical codes 1, 2,...,20 are used to denote the 20 native amino acid types according to the alphabetical order of their single character codes. The $L \times 20$ scores in Eq. (5) were generated by using PSI-BLAST [8] to search the Swiss-Prot database (version 52.0

released on 6-March-2007) through three iterations with 0.001 as the E-value cutoff for multiple sequence alignment against the sequence of the protein **P**, followed by a standardization procedure given below:

$$\mathbb{V}_{i \to j} = \frac{\mathbb{V}_{i \to j}^{0} - \left\langle \mathbb{V}_{i}^{0} \right\rangle}{\mathrm{SD}(\mathbb{V}_{i}^{0})} \quad (i = 1, 2, \dots, L; \ j = 1, 2, \dots, 20)$$
(6)

where $\mathbb{V}_{i\rightarrow j}^{0}$ represent the original scores directly created by PSI-BLAST [8] that are generally shown as positive or negative integers; $\langle \mathbb{V}_{i}^{0} \rangle$ the mean of $\mathbb{V}_{i\rightarrow j}^{0}$ over 20 native amino acids; $SD(\mathbb{V}_{i}^{0})$ the standard deviation of $\mathbb{V}_{i\rightarrow j}^{0}$. The standardized scores will have a zero mean value over the 20 amino acids and will remain unchanged if going through the same conversion procedure again. The positive score means that the corresponding mutation occurs more frequently in the alignment than expected by chance, while the negative one means just the opposite. However, according to the PSSM descriptor (Eq. (5)), proteins with different lengths will correspond to row-different matrices. To make the PSSM descriptor become a size-uniform matrix, one possible approach is to represent a protein sample P by

$$\bar{\mathbf{P}}_{\text{PSSM}} = \begin{bmatrix} \bar{\mathbb{V}}_1 & \bar{\mathbb{V}}_2 & \cdots & \bar{\mathbb{V}}_{20} \end{bmatrix}^{\text{T}}$$

$$\tag{7}$$

where

$$\bar{\mathbb{V}}_{j} = \frac{1}{L} \sum_{i=1}^{L} \mathbb{V}_{i \to j} \quad (j = 1, 2, \dots, 20)$$
(8)

where $\bar{\mathbb{V}}_j$ represents the average score of the amino acid residues in the protein **P** being changed to amino acid type *j* during the evolution process. However, if $\bar{\mathbf{P}}_{\text{PSSM}}$ of Eq. (7) was used to represent the protein **P**, all the sequence-order information during the evolution process would be lost. To avoid complete loss of the sequence-order information, the concept of the pseudo amino acid (PseAA) composition as originally proposed in [9,10] was adopted; i.e., instead of Eq. (7), let us use the pseudo position-specific scoring matrix (Pse-PSSM) as given by

$$\mathbf{P}_{\text{Pse-PSSM}}^{\xi} = \begin{bmatrix} \bar{\mathbb{V}}_1 & \bar{\mathbb{V}}_2 & \cdots & \bar{\mathbb{V}}_{20} & \boldsymbol{\Phi}_1^{\xi} & \boldsymbol{\Phi}_2^{\xi} & \cdots & \boldsymbol{\Phi}_{20}^{\xi} \end{bmatrix}^{\mathrm{T}}$$
(9)

to represent the protein P, where

$$\mathbf{\Phi}_{j}^{\xi} = \frac{1}{L - \xi} \sum_{i=1}^{L - \xi} \left[\mathbb{V}_{i \to j} - \mathbb{V}_{(i+\xi) \to j} \right]^{2} \quad (j = 1, 2, \dots, 20; \ \xi < L)$$
(10)

meaning that Φ_j^1 is the correlation factor by coupling the most contiguous PSSM scores along the protein chain for the amino acid type j; Φ_j^2 that by coupling the second-most contiguous PSSM scores; and so forth. Note that, as mentioned in the Material section, the length of the shortest protein sequence in the benchmark dataset is L = 50, and hence the value allowed for ξ in Eq. (10) must be smaller than 50. When $\xi = 0$, Φ_j^{ξ} becomes a naught element and Eq. (9) is degenerated to Eq. (7).

Optimized evidence-theoretic k nearest neighbor (OET-KNN) classifier. The OET-KNN classifier is a very powerful classification engine as demonstrated by its role in enhancing the success rates of predicting protein subcellular localization [11], where a detailed mathematical formulation for OET-KNN was also provided in the Appendix B. Here, we just give a brief description of how to use it to identify enzyme, its main-class and subclass. First of all, let us consider the top-level problem, i.e., to identify a protein as enzyme or non-enzyme with the benchmark dataset $S = S^{ezy} \cup S^{non-ezy}(Eq. (1))$. Suppose the process in identifying the query protein **P** among the two classes by OET-KNN is formulated as

$$\mathbf{OET}\text{-}\mathbf{KNN} \triangleright \mathbf{P} = \begin{cases} \mathbf{OET}\text{-}\mathbf{KNN} \triangleright \mathbf{P}_{\text{FunD}} = \Lambda_1(K, i), & \text{for FunD frame} \\ \mathbf{OET}\text{-}\mathbf{KNN} \triangleright \mathbf{P}_{\text{Pse-PSSM}}^{\xi} = \Lambda_2(K, \xi, i), & \text{for Pse-PSSM} \end{cases}$$
(11)

where \triangleright represents an action operator, $\Lambda_1(K, i)$ the creditability score for the query protein believed in the *i*th class when it is defined in the FunD frame (Eq. (3)), K is the parameter selected for the OET-KNN classifier [11], $\Lambda_2(K, \xi, i)$ the corresponding creditability score when the prediction is operated in the Pse-PSSM frame (Eq. (9)), and ξ the parameter selected for defining $\mathbf{P}_{\text{pse-PSSM}}^{\xi}$ (Eqs. (9) and (10)). Accordingly, using different descriptors to represent protein samples may lead to different results; even if the same descriptor is adopted, selecting different parameters may lead to different results as well. In order to get a unique result, the fusion approach is introduced as formulated below.

Fusion approach. The parameter K in Eq. (11) is the number of the nearest proteins counted against the query protein during the prediction process [12]. Generally speaking, for most training datasets, when K > 10 the success rate drops down remarkably and hence we can narrow the scope of K from 1 to 10. Also, the parameter ξ must be smaller than 50, the number of amino acids for the shortest protein sequence in the benchmark dataset. Therefore, the final predicted result should be determined by a fusion approach through the following voting mechanism. According to Eq. (11), the voting score for the query protein **P** belonging to the *i*th class is given by

$$\Pi_{i} = \sum_{K=1}^{10} w_{K}^{1} \Lambda_{1}(K, i) + \sum_{K=1}^{10} \sum_{\xi=0}^{49} w_{K,\xi}^{2} \Lambda_{2}(K, \xi, i), \quad (i = 1, 2)$$
(12)

where i = 1 is for enzyme and i = 2 for non-enzyme, w_K^1 and $w_{K,\xi}^2$ are the weight factors and were set at 1 for simplicity, thus the query protein **P** is predicted belonging to the class or subset for which the score of Eq. (12) is the highest; i.e.,

$$\mu = \arg \max \{\Pi_i\}, \quad (i = 1, 2)$$
(13)

where μ is the argument of *i* that maximize Π_i . If there is a tie, then the final predicted result will be randomly assigned to one of their corresponding subsets although this kind of tie case rarely happens and actually was not observed in the current study.

By changing (i = 1, 2) to (i = 1, 2, ..., 6) and working on the benchmark dataset \mathbb{S}^{ezy} (Eq. (1)), Eqs. (11)–(13) can be automatically used to solve the 2nd-level problem; by changing to (i = 1, 2, ..., 18) and working on \mathbb{S}_1^{ezy} (Eq. (1)), solve the 1st problem at the 3rd-level; and so forth. Such a procedure is the so-called top–down approach, and the entire predictor called **EzyPred**.

The above fusion approach not only can incorporate both the functional domain information and the protein evolution information but will also automatically solve the problem caused by the incompleteness of the FunD database. For example, if a query protein has no hit whatsoever when searching the Pfam database, it will correspond to a naught vector according to Eq. (3). The creditability score for a naught vector is zero by default (i.e., $\Lambda_1(K, i) = 0$) according to Eq. (11), and the creditability score will be solely determined by $\Lambda_2(K, \xi, i)$ derived from the Pse-PSSM frame.

To provide an intuitive picture, a flowchart to show how to fuse the FunD approach and Pse-PSSM approach is given in Fig. 2A, and that to show the top-down approach process of the 3-layer predictor is given in Fig. 2B.

Table 1

Success rates by the jackknife test in identifying the enzyme proteins and non-enzyme proteins

Protein type	Number of proteins	Number of correct predictions	Success rate (%)
Enzyme	9832	9089	92.4
Non-enzyme	9850	8875	90.1
Overall	19,682	17,964	91.3

Table 2

Success rates by the jackknife test in identifying enzyme main functional classes

Enzyme main functional class	Number of proteins	Number of correct predictions	Success rate (%)
EC.1: Oxidoreductase	1618	1478	91.4
EC.2: Transferase	3450	3260	94.5
EC.3: Hydrolase	2791	2711	97.1
EC.4: Lyase	679	578	85.1
EC.5: Isomerase	518	433	83.6
EC.6: Ligase	776	749	96.5
Overall	9832	9209	93.7



Fig. 2. A flowchart to show (A) how to fuse the FunD approach and Pse-PSSM approach into a prediction engine, and (B) how the top-down approach of the 3-layer predictor works.

Table 3

Success rates by the jackknife test in identifying sub-classes of the six main functional classes

Sakkar of orderdename IEC11		Number of proteins	Number of correct predictions	Success rate (%)
EC.11. Avaing on the CH-OH group of doors 149 440 98.0 EC.12. Avaing on the CH-OH group of doors 158 138 87.3 EC.13. Avaing on the CH-OH group of doors 177 256 77.8 EC.15. Avaing on the CH-OH group of doors 177 92 78.6 EC.15. Avaing on abel and EC-H-NH group of doors 177 92 78.6 EC.15. Avaing on abeliar proof of doors 76 66 92.7 EC.15. Avaing on abeliar proof doors 76 66 64 92.7 EC.15. Avaing on abeliar proof doors 71 64 958 95.7 95.8 EC.112. Avaing on hydrogen as elonor 71 14 82.4 95.8 95.7 95.8	Subclass of oxidoreductases (EC 1)	*	Å	
EC.12. Acting on the cli-Cli group of donors 158 1.38 87.3 EC.13. Acting on the Cli-NH, group of donors 72 56 77.3 EC.14. Acting on the Cli-NH, group of donors 72 76 67.3 EC.15. Acting on The Cli-NH, group of donors 70 186 89.9 EC.16. Acting on The Cli-NH group of donors 76 65 85.5 EC.16. Acting on the ritrogroup of donors 76 65 85.5 EC.17. Acting on a sulfur group of donors 76 66 95.7 EC.11.0. Acting on a dube group of donors 71 46 81.4 EC.11.2. Acting on a dube group of donors 71 46 82.4 EC.11.2. Acting on patroid donors, with incorporation or roduction of molecular oxygen 71 14 82.4 EC.11.3. Acting on patroid donors, with incorporation or roduction of molecular oxygen 73 86.6 86.6 EC.11.4. Cution on patroid donors, with incorporation or roduction of molecular oxygen 73 86.6 EC.11.4. Cution on intro- suffix proteins as donors 26 16 61.5 Oreanll 1820 157 86.7 EC.2.1.16. Cutify more suffix sets (EC.2)	EC.1.1: Acting on the CH–OH group of donors	449	440	98.0
FC1.3. Acting on the CH-CH group of donors. 149 101 67.3 EC1.4. Acting on the CH-NH group of donors 117 92 78.6 EC1.6. Acting on the CH-NH group of donors 35 17 48.6 FC1.8. Acting on other attrogenous compounds as donors 35 17 48.6 FC1.8. Acting on other attrogenous compounds as donors 90 66 95.7 FC1.18. Acting on a heme group of donors 90 66 95.7 FC1.18. Acting on a heme group of donors 92 64 93.3 EC1.14. Acting on a heme group of donors 92 64 93.3 EC1.14. Acting on a heme group of donors 101 18 93.1 EC1.14. Acting on a peroxide as acceptor 71 68 93.3 EC1.14. Acting on phydropen as door acyperation or molecular oxygen 173 18.6 18.6 EC1.15. Acting on superoxide as acceptor 20 14 70.0 86 EC1.17. Acting on superoxide as acceptor 20 14 70.0 86 EC1.17. Acting on superoxide as acceptor 20 14 90.0 15.2 90.5 15.2 90.5 15.2	EC.1.2: Acting on the aldehyde or oxo group of donors	158	138	87.3
EC.14. Acting on the C1+NIL group of doors 72 56 77.8 EC.15. Acting on NADH or NADPH 207 186 89.9 EC.16. Acting on OADH or NADPH 207 186 89.9 EC.17. Acting on OADH or NADPH 207 186 89.9 EC.18. Acting on a sulfar group of doors 76 65 85.5 EC.10. Acting on other nitrogenous compounds as doors 42 44 81.0 EC.11. Acting on a perioxide as acceptor 71 168 95.8 EC.11.2. Acting on hydrogen as donor 71 14 82.4 EC.11.3. Acting on single doors with incorporation of molecular oxygen 17 13 70.8 EC.11.4. Acting on single doors with incorporation or reduction of molecular oxygen 73 93.8 82.6 EC.11.3. Acting on single doors with incorporation or reduction of molecular oxygen 17 14 82.4 EC.11.3. Acting on single doors with incorporation or reduction of molecular oxygen 17 85.7 90.8 EC.11.3. Acting on single doors with incorporation or reduction of molecular oxygen 17 86.3 86.6 EC.11.3. Acting on superoxide as acceptor 128 121 90.7 8	EC.1.3: Acting on the CH-CH group of donors.	149	101	67.8
EC.15. Acting on the CH-NH group of doors 117 92 78.6 EC.16. Acting on other mitogenous compounds as donors 35 17 48.6 FC.18. Acting on other mitogenous compounds as donors 60 66 98.7 EC.17. Acting on a near group of donors 60 66 98.7 EC.11. Acting on a period as acceptor 71 68 98.3 EC.11. Acting on a period as acceptor 71 68 98.3 EC.11. Acting on a period donors, with incorporation or molecular oxygen 17 18 88.6 EC.11. Acting on superiod donors, with incorporation or molecular oxygen 17 18 88.6 EC.11. S. Acting on role of donors, with incorporation or molecular oxygen 17 18 88.6 EC.11. S. Acting on iron-suffur proteins us donors 26 16 66.1.5 Overall 180 00.6 67.7 88 86.6 EC.21. Transferring alchydro kroter exidues 34 34 100 67.7 86.7 Stocksas of maniferases 239 512 94.8 62.2 94.8 62.2 94.8 62.2 94.8 94.7 94.8 94.7<	EC.1.4: Acting on the CH-NH ₂ group of donors	72	56	77.8
EC.16. Acting on NADH or NADPH 207 186 89.9 EC.17. Acting on Other microgenous compounds as donors 35 17 48.6 EC.18. Acting on a suftur group of donors 76 6.5 85.5 EC.10. Acting on a suftur group of donors 42 34 81.0 EC.11.10. Acting on a diplemob and relatel subtances as donors 42 34 81.0 EC.11.2. Acting on hydrogen as donor 17 14 82.4 EC.11.2. Acting on hydrogen as donor 17 14 82.4 EC.11.2. Acting on paired donors, with incorporation or reduction of molecular oxygen 17 18.0 90.0 EC.11.7. Acting on thor -suftur proteins as donors 26 16 61.5 0 EC.11.7. Acting on tron-suftur proteins as donors 29 512 96.8 62.2 Subclass of transforms quoteque as acceptor 17 84.3 100 100 157 86.7 EC.2.1.3. Acting on thor of thy groups 129 96.8 157 86.7 EC.2.2.3. Acyltransforases 34 34 100 162.1 161.5 0 EC.2.3. Acyltransforases 129	EC.1.5: Acting on the CH–NH group of donors	117	92	78.6
EC.12: Acting on other introgenous compounds as donors 35 17 48.6 EC.13: Acting on a shull group of donors 69 66 95.7 EC.10: Acting on a shull group of donors 69 66 95.7 EC.11: Acting on a proxide as acceptor 71 68 98.8 EC.11: Acting on inple donors with incorporation of molecular oxygen 77 14 82.4 EC.11: Acting on inple donors with incorporation or reduction of molecular oxygen 73 157 90.8 EC.11: Acting on inple donors with incorporation or reduction of molecular oxygen 73 157 90.8 EC.11: Acting on inde-adular proteins as doors 20 14 70.0 EC.11: Acting on inde-adular proteins as doors 26 16 61.5 Solchass of transfermages (EC.2) EC.2: Transfermage advector groups 29 512 96.8 EC.2: Arguing inco-adular proteins as doors 31 34 100 100 92.2 94.8 94.9 94.7 94.9 94.7 94.9 94.7 94.7 94.6 94.7 94.6 94.7 94.6 94.7 94.6 94.7 94.6 94.6 94.6	EC.1.6: Acting on NADH or NADPH	207	186	89.9
Ref. 18: Acting on a sulting group of donors 60 63 85.7 RC1.10: Acting on a home group of donors 64 85.7 RC1.11: Acting on a persoid as acceptor 71 168 95.8 RC1.12: Acting on inyclogen as donor 77 14 82.4 RC1.13: Acting on paired donors, with incorporation of molecular oxygen 17 13 157 90.8 RC1.14: Acting on naiged donors with incorporation or reduction of molecular oxygen 173 157 90.8 RC1.14: Acting on naiged donors, with incorporation or meduction of molecular oxygen 173 157 90.8 RC1.17: Acting on superxide as acceptor 20 14 70.0 RC1.17: Acting on icon-sultur proteins as donors 26 16 6.5 Subclass of transferaces (EC.2) EC.21: Transferring addekyde or kolone residues 34 400 RC2.2: Transferring addekyde or kolone residues 34 400 90.7 RC2.2: Transferring addekyde or kolone residues 34 94.9 90.7 RC2.2: Transferring addekyde or kolone residues 34 100 90.5 RC2.2: Transferring addekyde or kolone residues 36 37 90.5	EC.1.7: Acting on other nitrogenous compounds as donors	35	17	48.6
E.1.F. Acting on a figure 3 and related substances as donors 42 34 81.0 EC.111 Acting on a provide as acceptor 71 68 92.5 EC.112 Acting on inplot donard solutions stances as donors 42 34 81.0 EC.113 Acting on inplot donard solution of molecular oxygen 17 14 82.4 EC.113 Acting on inplot donard solution of molecular oxygen 17 14 82.4 EC.115 Acting on inplot donard solution of molecular oxygen 17 14 82.4 EC.115 Acting on input donard solution of molecular oxygen 17 14 82.4 EC.115 Acting on input donard solution of molecular oxygen 17 14 82.6 EC.115 Acting on incon-ular provide as acceptor 26 16 66.15 Overall 18.20 157 86.7 Soldats of transferring aldehyde or lectone residues 24 44 100 EC.21 Transferring aldehyde or lectone residues 24 244 90.7 EC.22 Transferring aldehyde or lectone residues 467 433 94.0 EC.24 Circondyntronformers 267 264 264 96.8 EC.24 Transfering aldehyde or lectone	EC.1.8: Acting on a sulfur group of donors	/6	65	85.5
L2.117. Acting on approximate as acceptor 71 64 92.8 EC.112. Acting on particities as acceptor 71 64 92.8 EC.112. Acting on sigle doors with incorporation of molecular oxygen 17 13 65.1 EC.112. Acting on sigle doors with incorporation or reduction of molecular oxygen 173 157 92.6 EC.114. Acting on sigle doors with incorporation or reduction of molecular oxygen 173 157 92.6 EC.116. Acting on sigle doors with incorporation or reduction of molecular oxygen 173 157 92.6 EC.116. Acting on supervised as acceptor 26 16 6.5.7 Subclass of transferrang one carbon groups 26 15 65.7 Subclass of transferrang acting to restore residues 34 44 100 EC.21: Transferring aldelyde or ketone residues 34 34 100 EC.22: Transferring aldelyde or ketone residues 34 34 100 EC.23: Transferring aldelyde or ketone residues 34 34 100 EC.24: Clycosylaras 467 433 94.9 EC.25: Transferring aldelyde or ketone residues 36 37 90.3 EC	EC.1.9: Acting on a neme group of donors	42	00 24	95.7
11 01 22.5 12 Acting on spirage and solver of the solution of molecular oxygen 17 14 82.4 12 Acting on single donors with incorporation or nonecular oxygen 17 15 05 12 Acting on single donors, with incorporation or nonecular oxygen 17 13 157 08.8 12 Acting on a uper oxide as acceptor 23 23 92.0 12 Acting on diard on ron-sulfur proteins as donors 26 16 66 61.5 12 Transferrating one carbon groups 259 512 98.8 86.6 12 Transferrating one-carbon groups 259 512 98.8 86.6 12 Transferrating one-carbon groups 32 29.4 90.7 86.7 12 Transferring advice or ketone residues 34 34 100 76.2 76.7 96.8 12 Transferring advice or ketone residues 34 24 94.9 90.7 26.2 77.7 267 96.4 122.5 Transferring advice or ketone residues 36.3 57 90.5 100.7 96.9	EC.1.10. Acting on a perovide as acceptor	42	54 68	81.0 95.8
C113: Acting on single donors with incorporation of molecular oxygen 47 32 68.1 C1.14: Acting on single donors with incorporation or reduction of molecular oxygen 173 157 90.8 EC.14: Acting on supersoid as acceptor 25 23 92.0 EC.16: Oxidizing metal ions 20 14 70.0 EC.17: Acting on ion-suffur proteins as donors 26 16 61.5 Overall 182.0 1577 86.6 EC.21: Transferring one-carbon groups 29 512 96.8 EC.22: Transferring aldehyde or ketone residues 34 34 100 EC.23: Transferring aldehyde or ketone residues 324 294 90.7 EC.24: Carbonsferring aldehyde or groups 277 267 96.4 EC.25: Transferring aldehyde or groups or ports 139 1007 96.9 EC.24: Carbonsferring aldehyde or groups or ports 128 124 98.9 EC.25: Transferring aldehyde or groups or ports 128 124 98.9 EC.24: Carbonsferring aldehyde or groups 1039 1007 96.9 Sociation or action-introg nulfur-containing groups 164 446	EC.1.11. Acting on hydrogen as donor	17	14	95.8 82.4
EC.114. Acting on paired donors, with incorporation or reduction of molecular oxygen 173 157 90.8 EC.115. Acting on supervoide as acceptor 20 14 70.0 EC.116. Oxidizing metal incorporation or reduction of molecular oxygen 16 86.6 EC.116. Oxidizing metal incorporation as donors 20 14 70.0 EC.117. Acting on the northur proteins as donors 20 16 66.1.5 Subclass of transferring on carbon groups 529 512 96.8 EC.2.1: Transferring onc-carbon groups 529 512 96.8 EC.2.2: Transferring alkylor oxistone residues 34 34 1000 EC.2.2: Transferring alkylor and groups (other than methyl groups) 277 267 96.4 EC.2.3: Criticarisering introgenous groups 114 112 98.3 EC.2.3: Transferring phosphorous-containing groups 103 1007 96.9 EC.2.3: Transferring phosphorous-containing groups 128 124 98.9 EC.2.3: ransferring boshphorous-containing groups 134 128 25.5 C.2.1: Transferring phosphorous-containing inous	EC 1.13: Acting on single donors with incorporation of molecular oxygen	47	32	68.1
EC.115. Acting on superviside as acceptor 1 25 23 92.0 EC.116. Oxiding metal ions 20 14 70.0 EC.117. Acting on CH or CH ₂ groups 67 58 86.6 EC.118. Nating on iron-sulfur proteins as donors 26 16 61.5 Overall 1820 1577 86.7 Subclass of transferring alchyde or ketone residues 34 34 100 EC.21. Transferring alchyde or ketone residues 324 294 90.7 EC.23. Transferring alchyde or ketone residues 324 294 90.7 EC.24. Grucsoftransferrases 467 443 94.9 EC.25. Transferring hiotopsonous groups 114 112 96.3 EC.26. Transferring hiotopsonous groups 1039 1007 96.9 EC.27. Transferring hiotopsonous groups 114 112 98.3 EC.27. Transferring hiotopsonous groups 1039 1007 96.9 EC.28. Transferring hiotopsonous containing groups 163 57 90.5 Overall 2847 2726 95.8 Subclass of plaster bonds 122 95.8	EC.1.14: Acting on paired donors, with incorporation or reduction of molecular oxygen	173	157	90.8
EC.116: Oxidizing metal ions 20 14 70.0 EC.117: Acting on CH or CHg poups 26 16 61.5 Overall 1820 1577 86.7 Subclass of transferring one-carbon groups 529 512 96.8 EC.21: Transferring one-carbon groups 34 34 100 EC.23: Transferring allelyde or ketone residues 34 34 94.9 EC.24: Transferring allelyde or ketone residues 34 404 94.9 EC.23: Transferring allelyde or ketone residues 467 443 94.9 EC.24: Transferring allelyde or ketone residues 467 443 94.9 EC.25: Transferring allelyde or araly groups (other than methyl groups) 114 112 98.3 EC.27: Transferring balfur-containing groups 1039 1007 96.6 EC.38: Acting on seter bonds 128 1214 98.9 EC.34: Acting on seter bonds 128 1214 98.9 EC.35: Carbon-nitrogen bonds other than peptide bonds 436 408 93.6 EC.35: Carbon-nitrogen bonds other than peptide bonds 365 35.9 95.9 Subclass of	EC.1.15: Acting on superoxide as acceptor	25	23	92.0
EC.1.7: Acting on CH groups 67 58 86.6 EC.1.8: Acting on iron-sulfur proteins as donors 26 16 61.5 Overall 1820 1577 86.7 Subclass of transferases (EC.2) 512 96.8 EC.2.1: Transferring addryde recentives 34 34 100 EC.2.3: Transferring addryde recentives 324 294 90.7 EC.2.4: Carbon Scipturnasferases 324 294 90.7 EC.2.5: Transferring aldryde recentives 34 34 94.0 EC.2.5: Transferring aldryde recentives 367 96.4 EC.2.6: Transferring phosphorous-containing groups 1039 10007 96.9 EC.2.8: Transferring aldryde recentaining groups 63 57 90.5 Overall 2847 2726 95.8 Subclass of hydrolaxes (EC.3) EC.3: Carbon carbon EC.3: Carbon carbon and phydrolaxes (EC.3) EC.3: Carbon carbon and phydrolaxes (EC.3) EC.3: Carbon carbon soft pases 665 632 95.0 Subclass of bases (EC.4) EC.4: Carbon carbon soft pases 36 36 95.9 Subclass of bases (EC.4) EC.4: Carbon-ca	EC.1.16: Oxidizing metal ions	20	14	70.0
EC.1.18: Acting on iron-sulfur proteins as donors 26 16 6.1.5 Overall 1820 1577 86.7 Subclass of transferring concearbon groups 529 512 96.8 EC.2.1: Transferring oncearbon groups 324 294 90.7 EC.2.3: Argitransferrases 324 294 90.7 EC.2.4: Construint concerning alkylor aryl groups (other than methyl groups) 277 267 96.4 EC.2.5: Transferring sulfur-containing groups 114 112 98.3 EC.2.6: Transferring sulfur-containing groups 63 57 90.5 C2.6: Transferring on perfore bonds 284 246 96.9 EC.3: Construing end phyloidases (EC.3) 1228 1214 98.9 EC.3: Construing end phyloidases (EC.3) 1228 1214 98.9 EC.3: Acting on ester bonds 1228 1214 98.9 EC.3: Acting on acid anhydrides 665 632 95.0 Overall 3279 3146 95.9 Subclass of hases (EC.4) 222 96.8 24.3 24.3 26.3 25.0 25.0 25.0 25.	EC.1.17: Acting on CH or CH ₂ groups	67	58	86.6
Overall 1820 1577 86.7 Subclass of transferring one-achong props 529 512 96.8 EC.2.1: Transferring aldehyde or ketone residues 34 34 100 EC.2.3: Acyttransferrases 324 294 90.7 EC.2.4: Grossyltransferrases 324 294 90.7 EC.2.5: Transferring aldehyde or ketone residues 324 294 90.7 EC.2.5: Transferring introgenous groups 114 112 98.3 EC.2.6: Transferring sulfur-containing groups 103 1007 96.9 Coverall 2847 2726 95.8 Subclass of hydrolaxes (EC.3) EC.3: Acting on ester bonds 128 1214 98.9 EC.3.4: Acting on acid anhydrides 265 652 95.0 0 95.0 Subclass of lyase (EC.4) EC.3: Acting on acid anhydrides 266 632 95.0 Coverall 229 3146 96.5 95.0 Subclass of lyase (EC.4) EC.4: Carbon-arbon lyases 36 30 95.0	EC.1.18: Acting on iron-sulfur proteins as donors	26	16	61.5
Subclass of transferraces (EC.2) 529 512 96.8 EC.2.1: Transferring aldehydo restore residues 34 34 100 EC.2.3: Transferring aldehydo restore residues 324 294 90.7 EC.2.4: Transferring aldhydo restore residues 324 294 90.7 EC.2.5: Transferring alkyl or aryl groups (other than methyl groups) 277 267 96.4 EC.2.5: Transferring ploxphorus-containing groups 114 112 98.3 EC.2.6: Transferring sulfur-containing groups 1039 1007 96.9 EC.2.7: Transferring sulfur-containing groups 63 57 90.5 Overall 284 214 98.9 EC.3.1: Acting on ester bonds 1228 1214 98.9 EC.3.2: Citycosylases 464 446 91.8 EC.3.4: Acting on actid anhydrides 665 632 95.0 Overall 3279 3146 95.9 Subclass of hyses (EC.4) EC.4: Carbon-nitrogen bonds other than peptide bonds 436 408 EC.3.4: Acting on actid anhydrides 365 350 95.9 Subclass of hyses (EC.4) <td< td=""><td>Overall</td><td>1820</td><td>1577</td><td>86.7</td></td<>	Overall	1820	1577	86.7
EC.2.1: Transferring one-carbon groups 529 512 96.8 EC.2.2: Transferring aldehyde or ketone residues 34 34 100 EC.2.3: Acyltransferring alkylor aryl groups (other than methyl groups) 277 267 96.4 EC.2.5: Transferring nitrogenous groups 114 112 98.3 EC.2.6: Transferring nitrogenous-containing groups 1039 1007 96.9 EC.2.8: Transferring nitrogenous-containing groups 63 57 90.5 Coverall 2847 2726 95.8 Subclass of hydrolaxes (EC.3) 2121 98.9 95.0 EC.3.1: Acting on ester bonds 1228 1214 98.9 EC.3.2: Glycosylases 464 446 96.1 EC.3.4: Acting on carbon-mitrogen bonds other than peptide bonds 436 408 93.6 EC.3.5: Acting on carbon-mitrogen bonds other than peptide bonds 365 350 95.9 Subclass of lyaces (EC.4) EC.4: Carbon-actonon lyases 365 350 95.9 Subclass of lyaces (EC.4) EC.4: Carbon-actonon lyases 365 350 95.9 Subclass of lyaces (EC.4) EC.4: Carbon-actonon ly	Subclass of transferases (EC.2)			
EC.2.2 transferring aldehyde or ketone residues 34 34 34 94 90.7 EC.2.3 regittmansferases 324 294 90.7 EC.2.4: Glycosyltransferases 467 443 94.9 EC.2.5: Transferring nitrogenous groups 114 112 98.3 EC.2.6: Transferring phosphorous-containing groups 103 1007 96.9 EC.2.7: Transferring phosphorous-containing groups 63 57 90.5 Overall 2847 2726 95.8 Subclass of hydrolases (EC.3) 2 2 124 98.9 EC.3.1: Acting on ester bonds 123 1214 98.9 93.6 EC.3.2: Glycosylases 464 446 91.8 93.6 EC.3.5: Acting on earbon-mitrogen bonds other than peptide bonds 436 408 93.6 EC.3.6: Acting on acid anhydrides 665 652 95.9 Subclass of Jyses (EC.4) 2 42 94.9 EC.4.1: Carbon-oxygen lyses 36 350 95.9 Subclass of Jyses (EC.4) 2 56 56 100 EC.4.2: Carbon-oxygen ly	EC.2.1: Transferring one-carbon groups	529	512	96.8
EC.2.3: Acyltransferases 324 294 90.7 EC.2.4: Glycosyltransferases 467 443 49.9 EC.2.5: Transferring nitrogenous groups 114 112 98.3 EC.2.6: Transferring notyperous-containing groups 1039 1007 96.9 EC.2.6: Transferring notyperous-containing groups 63 57 90.5 Overall 2847 2726 95.8 Subclass of hydrolases (EC.3) 5 90.5 90.5 EC.3.1: Acting on setre bonds 1228 1214 98.9 EC.3.2: Citing on carbon-nitrogen bonds other than peptide bonds 464 446 96.1 EC.3.5: Acting on acid anhydrides 665 632 95.0 Overall 329 3146 95.9 Subclass of hyses (EC.4) 50 632 95.0 Overall 329 96.8 96.8 95.9 Subclass of hyses (EC.4) 50 80.7 95.9 EC.4.1: Carboncarbon lyses 62 50 80.7 EC.4.2: Carbonarbon payses 62 50 80.7 EC.4.2: Carbonsurgen lyses	EC.2.2: Transferring aldehyde or ketone residues	34	34	100
EC.2.4: Glycosyltransferases 467 443 94.9 EC.2.5: Transferring alky or aryl groups (other than methyl groups) 277 267 96.4 EC.2.6: Transferring nitrogenous groups 114 112 98.3 EC.2.7: Transferring phosphorous-containing groups 63 57 90.5 Overall 2847 2726 95.8 Subclass of hydrolases (EC.3) 84.4 98.9 EC.3.1: Acting on estro bonds 1228 1214 98.9 EC.3.4: Acting on earbon-nitrogen bonds other than peptide bonds 436 446 91.8 EC.3.5: Acting on acid anhydrides 665 632 95.0 Overall 3279 3146 95.9 Subclass of lyases (EC.4) 22 96.8 EC.4.1: Carbon-actron lyases 365 350 95.9 Subclass of lyases (EC.4) 23 74.2 EC.4.2: Carbon-actron lyases 365 350 95.9 Subclass of lyases 31 23 74.2 EC.4.2: Carbon-actron lyases 36 36 30 95.9 <t< td=""><td>EC.2.3: Acyltransferases</td><td>324</td><td>294</td><td>90.7</td></t<>	EC.2.3: Acyltransferases	324	294	90.7
EC.2.5. Transferring altyd or aryl groups (other than methyl groups) 277 267 96.4 EC.2.6. Transferring nitogenous groups 114 112 98.3 EC.2.7: Transferring nitogenous groups 1039 1007 96.9 EC.2.8. Transferring nitogenous groups 63 57 90.5 Overall 2847 2726 95.8 Subclass of hydrolases (EC.3) EC.3.1: Acting on ester bonds 1228 1214 98.9 EC.3.2: Citycosylases 464 446 91.8 8 EC.3.4: Cating on exter bonds 436 446 91.8 EC.3.5: Acting on acid anhydrides 665 632 95.9 Overall 3279 3146 95.9 Subclass of lyases (EC.4) EC.4.1: Carbon-carbon lyases 62 50 80.7 EC.4.2: Carbon-surgen lyases 62 50 80.7 82.4 89.5 Overall 3279 314 92 94.4 89.5 Subclass of lyases (EC.4) 82 84 89.5 80.6 100 EC.4.2: Carbon-surgen lyases 56 56 100 <	EC.2.4: Glycosyltransferases	467	443	94.9
EC.2.6. Transferring nitrogenous groups 114 112 98.3 EC.2.7. Transferring nosphorous-containing groups 1039 1007 96.9 EC.2.8. Transferring nosphorous-containing groups 63 57 90.5 Overall 2847 2726 95.8 Subclass of hydrolases (EC.3) EC.3.1: Acting on selter bonds 1228 1214 98.9 EC.3.4. Acting on selter bonds 163 446 446 96.1 EC.3.5. Acting on action-nitrogen bonds other than peptide bonds 436 446 95.9 Subclass of lyases (EC.4) 2279 3146 95.9 Subclass of lyases (EC.4) EC.4.1: Carbon-atrop lyases 366 350 95.9 Subclass of lyases (EC.4) EC.4.2: Carbon-nitrogen lyases 366 50 95.9 Subclass of lyases (EC.4) EC.4.2: Carbon-nitrogen lyases 366 50 95.9 Subclass of isomerases 31 23 74.2 24.4 89.5 Overall 892 842 94.4 89.5 56 56 100 EC.4.2: Carbon-sult lyases 56 56 56 <	EC.2.5: Transferring alkyl or aryl groups (other than methyl groups)	277	267	96.4
EC.2.7: Transferring phosphorous-containing groups 1039 1007 96.9 EC.2.8: Transferring sulfur-containing groups 63 57 90.5 Overall 2847 2726 95.8 Subclass of hydrolases (EC.3) EC.3.1: Acting on seter bonds 1228 1214 98.9 EC.3.2: Glycosylases 464 446 96.1 8 EC.3.4: Acting on peptide bonds (peptide hydrolases) 486 446 91.8 EC.3.5: Acting on actoon-nitrogen bonds other than peptide bonds 665 632 95.0 Overall 3279 3146 95.9 Subclass of lyases (EC.4) EC.4.1: Carbon-carbon hyases 62 50 80.7 EC.4.2: Carbon-nitrogen lyases 62 50 80.7 EC.4.4: Carbon-nitrogen lyases 56 56 100 EC.4.4: Carbon-sorgen lyases 56 56 100 EC.4.4: Carbon-sorgen lyases 56 56 56 100 EC.4.6: Phosphorus-oxygen lyases 111 102 91.9 EC.4.6: Phosphorus-oxygen lyases 110 109 99.1 EC.5.1: Racenses and ep	EC.2.6: Transferring nitrogenous groups	114	112	98.3
EC.2.8: Transferring sulfur-containing groups 63 57 90.5 Overall 2847 2726 95.8 Subclass of hydrolases (EC.3) EC.31: Acting on ester bonds 1228 1214 98.9 EC.3.1: Acting on ester bonds 1228 1214 98.9 EC.3.2: Glycosylases 464 446 96.1 EC.3.2: Glycosylases 464 446 91.8 EC.3.5: Acting on acid anhydrides 665 632 95.0 Overall 3279 3146 95.9 Subclass of lyases (EC.4) EC.41: Carbon-carbon lyases 340 329 96.8 EC.4.2: Carbon-actor lyases 365 350 95.9 EC.4.3: Carbon-actor lyases 31 23 74.2 EC.4.6: Phosphorus-oxygen lyases 56 56 100 EC.4.1 82 842 94.4 Subclass of isomerases (EC.5) EC.5: I: Kacemases and epimerases 111 102 91.9 EC.5.2: cis-trams-Isomerases 110 109 99.1 125.7 133 95.7 EC.5.1: Racemases and epimerases 111	EC.2.7: Transferring phosphorous-containing groups	1039	1007	96.9
Overall 2847 2726 95.8 Subclass of hydrolases (EC.3) 5 5 EC.3.1: Acting on ester bonds 1228 1214 98.9 EC.3.2: Glycosylases 464 446 91.1 EC.3.4: Acting on peptide bonds (peptide hydrolases) 486 446 91.8 EC.3.5: Acting on acbon-nitrogen bonds other than peptide bonds 436 408 93.6 EC.3.6: Acting on acbon-nitrogen bonds other than peptide bonds 665 632 95.0 Overall 3279 3146 95.9 Subclass of lyases (EC.4) EC.4: Carbon-actron lyases 340 32.9 96.8 EC.4: Carbon-nitrogen lyases 62 50 80.7 EC.4: Carbon-nitrogen lyases 31 23 74.2 EC.4: Carbon-surgen lyases 56 56 100 EC.4: Carbon-surgen lyases 38 34 89.5 Overall 892 842 94.4 Subclass of isomerases (EC.5) EC.5.1: Racemases and epimerases 111 100 99.1 EC.	EC.2.8: Transferring sulfur-containing groups	63	57	90.5
Subclass of hydrolases (EC.3) 1228 1214 98.9 EC.3.1: Acting on seter bonds 1264 446 96.1 EC.3.2: Glycosylases 486 446 91.8 EC.3.4: Acting on peptide bonds (peptide hydrolases) 486 446 91.8 EC.3.5: Acting on acrbon-nitrogen bonds other than peptide bonds 436 408 93.6 C.3.5: Acting on acrbon-nitrogen bonds other than peptide bonds 665 652 95.0 Overall 3279 3146 95.9 Subclass of lyases (EC.4) EC.4: Carbon-actron lyases 340 329 96.8 EC.4: Carbonnitrogen lyases 62 50 80.7 EC.4: Carbonnitrogen lyases 62 50 80.7 EC.4: Carbonnitrogen lyases 31 23 74.2 EC.4: Carbonnitrogen lyases 38 34 89.5 Overall 892 842 94.4 Subclass of isomerases (EC.5) EC.5: Racemases and epimerases 111 100 99.9 EC.3: Itramolecular oxidoreductases 207 186 89.9 25.7 EC.3: Itramolecular oxidoreductases <td>Overall</td> <td>2847</td> <td>2726</td> <td>95.8</td>	Overall	2847	2726	95.8
EC.3.1: Acting on exter bonds 1228 1214 98.9 EC.3.2: Glycosylases 464 446 96.1 EC.3.4: Acting on peptide bonds (peptide hydrolases) 486 446 91.8 EC.3.5: Acting on carbon-nitrogen bonds other than peptide bonds 436 408 93.6 EC.3.5: Acting on acid anhydrides 665 632 95.0 Overall 3279 3146 95.9 Subclass of lyases (EC.4) EC.41: Carbon-carbon lyases 340 329 96.8 EC.4.2: Carbon-autorn lyases 365 350 95.9 EC.4.3: Carbon-nitrogen lyases 62 50 80.7 EC.4.4: Carbon-sulfur lyases 56 56 100 EC.4.5: Acting on sufficiency lyases 38 34 89.5 Overall 892 842 94.4 Subclass of isomerases (EC.5) EC.5.1: Racemases and epimerases 111 102 91.9 EC.5.2: eix-trans-lsomerases 110 109 99.1 133 95.7 EC.5.4: Intramolecular transferases (mutases) 139 133 95.7 45.5 EC.5.3: Int	Subclass of hydrolases (EC.3)			
EC.3.2: Glycosylases 464 446 96.1 EC.3.4: Acting on peptide bonds (peptide hydrolases) 486 446 91.8 EC.3.5: Acting on acid anhydrides 665 632 95.0 Overall 3279 3146 95.9 Subclass of lyases (EC.4) EC.4.1: Carbon-axpen lyases 365 350 95.9 Subclass of lyases (EC.4) EC.4.2: Carbon-axpen lyases 365 350 95.9 EC.4.3: Carbon-sulfur lyases 62 50 80.7 EC.4.4: Carbon-sulfur lyases 31 23 74.2 EC.4.6: Phosphorus-oxygen lyases 36 56 100 EC.4.9: Other lyases 38 34 89.5 Overall 892 842 94.4 Subclass of isomerases (EC.5) 110 109 99.1 EC.5.2: circ-trans-Isomerases 110 109 99.1 133 95.7 45.5 56.5 56.5 56.5 56.5 56.5 56.5 56.5 56.5 56.5	EC.3.1: Acting on ester bonds	1228	1214	98.9
EC.3.4: Acting on peptide bonds (peptide hydrolases) 486 446 91.8 EC.3.5: Acting on acidon-nitrogen bonds other than peptide bonds 665 632 95.0 EC.3.6: Acting on acid anhydrides 665 632 95.0 Overall 3279 3146 95.9 Subclass of lyases (EC.4) EC.4.1: Carbon-carbon lyases 340 32.9 96.8 EC.4.2: Carbon-arbon-nitrogen lyases 62 50 80.7 EC.4.3: Carbon-arbon-nitrogen lyases 62 50 80.7 EC.4.4: Carbon-auffur lyases 31 23 74.2 EC.4.6: Phosphorus-oxygen lyases 56 56 100 EC.4.4: Carbon-sulfur lyases 38 34 89.5 Overall 892 82 94.4 Subclass of isomerases (EC.5) EC EC 56 56 EC.5.1: Racemases and epimerases 111 102 91.9 EC.5.2: cis-trans-Isomerases 110 109 99.1 EC.5.2: harmolecular transferases (mutases) 139 133 95.7 EC.5.1: Bracemases and epimerases 11 5 45.5 </td <td>EC.3.2: Glycosylases</td> <td>464</td> <td>446</td> <td>96.1</td>	EC.3.2: Glycosylases	464	446	96.1
EC.3.5: Acting on carbon-nitrogen bonds other than peptide bonds 436 408 93.6 EC.3.6: Acting on acid anhydrides 665 632 95.0 Overall 3279 3146 95.9 Subclass of lyases (EC.4) EC.4.1: Carbon-carbon lyases 340 329 96.8 EC.4.2: Carbon-oxygen lyases 365 350 95.9 EC.4.4: Carbon-nitrogen lyases 62 50 80.7 EC.4.5: Actinon-sulfur lyases 56 56 100 EC.4.99: Other lyases 38 34 89.5 Overall 892 842 94.4 Subclass of isomerases (EC.5) EC.5.1: Racemases and epimerases 111 102 91.9 EC.5.2: cis-trans-Isomerases 110 109 99.1 EC.5.3: Intramolecular transferases (mutases) 139 133 95.7 EC.5.5: Intramolecular transferases (mutases) 11 5 45.5 EC.5.9: Other isomerases 61 61 100 Overall 639 596 93.3 Subclass of ligases (EC.6) EC.5.2: Forming carbon-oxygen bonds 496 493<	EC.3.4: Acting on peptide bonds (peptide hydrolases)	486	446	91.8
EC.3.6: Acting on acid annyarides 665 652 95.0 Overall 3279 3146 95.9 Subclass of lyases (EC.4) $=$ $=$ EC.4.1: Carbon-carbon lyases 340 329 96.8 EC.4.2: Carbon-oxygen lyases 365 350 95.9 EC.4.3: Carbon-nitrogen lyases 62 50 80.7 EC.4.4: Carbon-sulfur lyases 62 50 80.7 EC.4.4: Carbon-sygen lyases 62 56 100 EC.4.9: Other lyases 56 56 100 EC.4.9: Other lyases 38 34 89.5 Overall 892 842 94.4 Subclass of isomerases (EC.5) 111 102 91.9 EC.5.2: cis-trans-loomerases 110 109 99.1 EC.5.2: cis-trans-loomerases 110 109 99.1 EC.5.3: Intramolecular transferases (mutases) 139 133 95.7 EC.5.5: Intramolecular lyases 11 5 45.5 EC.5.99: Other isomerases 61 61 100 Overall 639 596 93.3 Subclass of ligases (EC.6) $=$ $=$ EC.6.1: Forming carbon-sulfur bonds 496 493 99.4 EC.6.2: Forming carbon-sulfur bonds 36 34 94.4 (continued on next page) 36 34 94.4	EC.3.5: Acting on carbon–nitrogen bonds other than peptide bonds $EC.2.6$ A (i.e., a circle of the left)	436	408	93.6
Overall 3279 3146 95.9 Subclass of lyases (EC.4) EC.4.1: Carbon-carbon lyases 340 329 96.8 EC.4.2: Carbon-arbon lyases 365 350 95.9 EC.4.3: Carbon-nitrgen lyases 62 50 80.7 EC.4.4: Carbon-sulfur lyases 62 50 80.7 EC.4.4: Carbon-sygen lyases 56 56 100 EC.4.9: Other lyases 38 34 89.5 Overall 892 842 94.4 Subclass of isomerases (EC.5) EC.5.1: Racemases and epimerases 111 102 91.9 EC.5.1: Racemases and epimerases 110 109 99.1 EC.5.3: Intramolecular oxidoreductases 207 186 89.9 EC.5.4: Intramolecular transferases (mutases) 139 133 95.7 EC.5.5: Intramolecular types 61 61 100 Overall 639 596 93.3 Subclass of ligases (EC.6) EC.6.1: Forming carbon-oxygen bonds 496 493 99.4	EC.3.6: Acting on acid anhydrides	665	632	95.0
Subclass of lyases (EC.4) 340 329 96.8 EC.4.1: Carbon-acrbon lyases 365 350 95.9 EC.4.2: Carbon-nitrogen lyases 62 50 80.7 EC.4.4: Carbon-sulfur lyases 31 23 74.2 EC.4.9: Carbon-sulfur lyases 56 56 100 EC.4.6: Phosphorus-oxygen lyases 56 56 100 EC.4.9: Other lyases 38 34 89.5 Overall 892 842 94.4 Subclass of isomerases (EC.5) 207 186 89.9 EC.5.1: Racemases and epimerases 111 102 91.9 EC.5.2: cis-trans-Isomerases 207 186 89.9 EC.5.3: Intramolecular transferases (mutases) 139 133 95.7 EC.5.4: Intramolecular tyases 11 5 45.5 EC.5.9: Other isomerases 61 61 100 Overall 639 596 93.3 Subclass of ligases (EC.6) 205 94.4 55 EC.5.2: Forming carbon-oxygen bonds 496 493 99.4 <td< td=""><td>Overall</td><td>3279</td><td>3146</td><td>95.9</td></td<>	Overall	3279	3146	95.9
LC.4.1: Carbon-oxygen lyases34032990.8EC.4.2: Carbon-oxygen lyases36535095.9EC.4.3: Carbon-sulfur lyases625080.7EC.4.4: Carbon-oxygen lyases5656100EC.4.9: Other lyases5656100EC.4.9: Other lyases383489.5Overall89284294.4Subclass of isomerases (EC.5)11110291.9EC.5.1: Racemases and epimerases11010999.1EC.5.2: cis-trans-Isomerases20718689.9EC.5.3: Intramolecular oxidoreductases20718689.9EC.5.5: Intramolecular lyases11545.5EC.5.9: Other isomerases11545.5EC.5.9: Other isomerases6161100Overall63959693.3Subclass of ligases (EC.6)2699.4EC.6.1: Forming carbon-oxygen bonds3634EC.6.2: Forming carbon-oxygen bonds	Subclass of lyases (EC.4)	240	220	06.9
Inc. 4.2100100100EC.4.3: Carbon-nitrogen lyases625080.7EC.4.4: Carbon-sulfur lyases312374.2EC.4.4: Carbon-suygen lyases5656100EC.4.9: Other lyases5656100EC.4.9: Other lyases383489.5Overall89284294.4Subclass of isomerases (EC.5)EC.5.1: Racemases and epimerases11110291.9EC.5.2: cis-trans-Isomerases11010999.113395.7EC.5.3: Intramolecular transferases (mutases)13913395.795.7EC.5.5: Intramolecular lyases11545.545.5EC.5.9: Other isomerases6161100Overall63959693.333Subclass of ligases (EC.6)563499.4EC.6.1: Forming carbon-oxygen bonds363494.4(continued on next page)363494.4	EC.4.1. Carbon oxygen lyases	340	329	90.8
C.4.c Carbon-sulfur lyases 31 23 74.2 EC.4.t Carbon-sulfur lyases 36 56 100 EC.4.9: Other lyases 38 34 89.5 Overall 892 842 94.4 Subclass of isomerases (EC.5) EC.5.1: Racemases and epimerases 111 102 91.9 EC.5.2: cis-trans-Isomerases 110 109 99.1 EC.5.3: Intramolecular oxidoreductases 207 186 89.9 EC.5.1: Intramolecular transferases (mutases) 111 5 45.5 45.5 EC.5.2: Intramolecular transferases (mutases) 111 5 45.5 EC.5.99: Other isomerases 61 61 100 Overall 639 596 93.3 Subclass of ligases (EC.6) EC.6.1: Forming carbon-oxygen bonds 496 493 99.4 EC.6.1: Forming carbon-sulfur bonds 36 34 94.4	EC.4.2. Carbon-nitrogen lyases	62	50	95.9 80.7
EC.4.6: Phosphorus-oxygen lyases 56 56 100 EC.4.9: Other lyases 38 34 89.5 Overall 892 842 94.4 Subclass of isomerases (EC.5) EC.5.1: Racemases and epimerases 111 102 91.9 EC.5.2: cis-trans-Isomerases 110 109 99.1 EC.5.3: Intramolecular oxidoreductases 207 186 89.9 EC.5.4: Intramolecular oxidoreductases 139 133 95.7 EC.5.5: Intramolecular transferases (mutases) 11 5 45.5 EC.5.99: Other isomerases 61 61 100 Overall 639 596 93.3 Subclass of ligases (EC.6) EC.6.1: Forming carbon-oxygen bonds 496 493 99.4 EC.6.2: Forming carbon-sulfur bonds 36 34 94.4	EC 4 4: Carbon–sulfur lyases	31	23	74.2
EC.4.99: Other lyases383489.5Overall 892 842 94.4 Subclass of isomerases (EC.5)EC.5.1: Racemases and epimerases 111 102 91.9 EC.5.2: cis-trans-Isomerases 110 109 99.1 EC.5.3: Intramolecular oxidoreductases 207 186 89.9 EC.5.4: Intramolecular transferases (mutases) 139 133 95.7 EC.5.5: Intramolecular lyases 11 5 45.5 EC.5.99: Other isomerases 61 61 100 Overall 639 596 93.3 Subclass of ligases (EC.6) $EC.6.1$: Forming carbon-oxygen bonds 496 493 99.4 EC.6.2: Forming carbon-sulfur bonds 36 34 94.4	EC.4.6: Phosphorus-oxygen lyases	56	56	100
Overall89284294.4Subclass of isomerases (EC.5)EC.5.1: Racemases and epimerases11110291.9EC.5.2: cis-trans-Isomerases11010999.1EC.5.3: Intramolecular oxidoreductases20718689.9EC.5.4: Intramolecular transferases (mutases)13913395.7EC.5.5: Intramolecular lyases11545.5EC.5.99: Other isomerases6161100Overall63959693.3Subclass of ligases (EC.6)49649399.4EC.6.1: Forming carbon-oxygen bonds49649399.4EC.6.2: Forming carbon-sulfur bonds363494.4363494.4	EC.4.99: Other lyases	38	34	89.5
Subclass of isomerases (EC.5) EC.5.1: Racemases and epimerases 111 102 91.9 EC.5.2: cis-trans-Isomerases 110 109 99.1 EC.5.3: Intramolecular oxidoreductases 207 186 89.9 EC.5.4: Intramolecular transferases (mutases) 139 133 95.7 EC.5.5: Intramolecular lyases 11 5 45.5 EC.5.99: Other isomerases 61 61 100 Overall 639 596 93.3 Subclass of ligases (EC.6) EC.6.1: Forming carbon–oxygen bonds 496 493 99.4 EC.6.2: Forming carbon–sulfur bonds 36 34 94.4 (continued on next page) 36 34 94.4	Overall	892	842	94.4
EC.5.1: Racemases and epimerases11110291.9EC.5.2: cis-trans-Isomerases11010999.1EC.5.3: Intramolecular oxidoreductases20718689.9EC.5.4: Intramolecular transferases (mutases)13913395.7EC.5.5: Intramolecular lyases11545.5EC.5.99: Other isomerases6161100Overall63959693.3Subclass of ligases (EC.6)206496493EC.6.1: Forming carbon-oxygen bonds49649399.4EC.6.2: Forming carbon-sulfur bonds363494.4(continued on next page)	Subclass of isomerases (EC.5)			
EC.5.2: cis-trans-Isomerases11010999.1EC.5.3: Intramolecular oxidoreductases20718689.9EC.5.4: Intramolecular transferases (mutases)13913395.7EC.5.5: Intramolecular lyases11545.5EC.5.99: Other isomerases6161100Overall63959693.3Subclass of ligases (EC.6)206496493EC.6.1: Forming carbon-oxygen bonds49649399.4EC.6.2: Forming carbon-sulfur bonds363494.4(continued on next page)	EC.5.1: Racemases and epimerases	111	102	91.9
EC.5.3: Intramolecular oxidoreductases20718689.9EC.5.4: Intramolecular transferases (mutases)13913395.7EC.5.5: Intramolecular lyases11545.5EC.5.99: Other isomerases6161100Overall63959693.3Subclass of ligases (EC.6)206496493EC.6.1: Forming carbon–oxygen bonds49649399.4EC.6.2: Forming carbon–sulfur bonds363494.4(continued on next page)	EC.5.2: cis-trans-Isomerases	110	109	99.1
EC.5.4: Intramolecular transferases (mutases)13913395.7EC.5.5: Intramolecular lyases11545.5EC.5.99: Other isomerases6161100Overall63959693.3Subclass of ligases (EC.6)22EC.6.1: Forming carbon–oxygen bonds496493EC.6.2: Forming carbon–sulfur bonds363494.4(continued on next page)	EC.5.3: Intramolecular oxidoreductases	207	186	89.9
EC.5.5: Intramolecular lyases11545.5EC.5.99: Other isomerases6161100Overall63959693.3Subclass of ligases (EC.6)49649399.4EC.6.2: Forming carbon–sulfur bonds363494.4(continued on next page)	EC.5.4: Intramolecular transferases (mutases)	139	133	95.7
EC.5.99: Other isomerases6161100Overall63959693.3Subclass of ligases (EC.6)49649399.4EC.6.1: Forming carbon–oxygen bonds49649399.4EC.6.2: Forming carbon–sulfur bonds363494.4(continued on next page)	EC.5.5: Intramolecular lyases	11	5	45.5
Overall63959693.3Subclass of ligases (EC.6)49649399.4EC.6.1: Forming carbon-oxygen bonds49649399.4EC.6.2: Forming carbon-sulfur bonds363494.4(continued on next page)	EC.5.99: Other isomerases	61	61	100
Subclass of ligases (EC.6)49649399.4EC.6.1: Forming carbon-oxygen bonds363494.4EC.6.2: Forming carbon-sulfur bonds363494.4(continued on next page)	Overall	639	596	93.3
EC.6.1: Forming carbon-oxygen bonds49649399.4EC.6.2: Forming carbon-sulfur bonds363494.4(continued on next page)	Subclass of ligases (EC.6)	10.0	402	60 A
(continued on next page)	EC.0.1: Forming carbon-oxygen bonds EC.6.2: Forming carbon-sulfur bonds	496 36	493 34	99.4 04 4
	20.0.2. I oming our our our outful outful	50	(continued)	on next page)

Table 3 (continued)

	Number of proteins	Number of correct predictions	Success rate (%)
EC.6.3: Forming carbon-nitrogen bonds	364	358	98.4
EC.6.4: Forming carbon-carbon bonds	13	11	84.6
EC.6.5: Forming phosphoric ester bonds	46	44	95.7
EC.6.6: Forming nitrogen-metal bonds	10	9	90.0
Overall	965	949	98.3

Results and discussion

In statistical prediction the independent dataset test, sub-sampling test, and jackknife test are often used in literatures for examining the accuracy of a predictor. Among these three, the jackknife test is deemed the most rigorous and objective [13], and hence has been increasingly adopted by investigators in examining the quality of various prediction methods (see, e.g., [14–32] as well as a recent review [33] in this regard).

The jackknife cross-validation results by **EzyPred** on the datasets S and S^{ezy} (cf. Eq. (1) and Online Supporting Information A) are given in Tables 1 and 2, respectively, from which we can see that the overall success rate in identifying the proteins as enzymes or non-enzymes is 91.3%, and that the overall success rate in identifying the enzymes among their six main functional classes is 93.7%. The corresponding results by **EzyPred** on the datasets S_1^{ezy} , S_2^{ezy} , S_3^{ezy} , S_4^{ezy} , S_5^{ezy} , and S_6^{ezy} (cf. Eq. (2) and Online Supporting Information B) are given in Table 3, from which we can see that the overall success rates in identifying the subfamily classes of oxidoreductase, transferases, hydrolases, lyases, isomerases, and ligases are 86.7%, 95.8%, 95.9%, 94.4%, 93.3%, and 98.3%, respectively.

It was reported [34] that even for the pair fragments with >50% sequence identity the probability of having a same EC number (enzymatic function) is <30%, meaning that enzyme function is much less conserved than anticipated. However, for the current datasets in which none of enzymes has $\ge 40\%$ sequence identity to any others in a same subset, the overall success rates by the **EzyPred** in identifying the main functional classes of enzymes and their subclasses are very high. As is well known, the more the number of classes to be identified, the less the success rate will be. However, even for the oxidoreductase dataset \mathbb{S}_1^{exy} consisting of 18 subfamily classes, the overall success rate obtained by the **EzyPred** is above 86%, indicating that **Ezy-Pred** is a very powerful predictor in identifying enzymes, their main classes, and their subclasses.

Conclusion

The reason why **EzyPred** predictor can yield so high success rates is because it operates by fusing the FunD approach and Pse-PSSM approach. The former is closely related to the functions of proteins, while the latter can incorporate their evolution information. It is anticipated

that with more data available in the ENZYME database, the current top-down **EzyPred** predictor can be extended to cover sub-subclass and sub-sub-subclass of enzymes as well. **EzyPred** is available to the public at the site http:// chou.med.harvard.edu/bioinf/EzyPred/.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc. 2007.09.098.

References

- [1] K.C. Chou, D.W. Elrod, Prediction of enzyme family classes, Journal of Proteome Research 2 (2003) 183–190.
- [2] K.C. Chou, Y.D. Cai, Predicting enzyme family class in a hybridization space, Protein Science 13 (2004) 2857–2863.
- [3] A. Bairoch, The ENZYME Database in 2000, Nucleic Acids Research 28 (2000) 304–305.
- [4] I. Letunic, R.R. Copley, B. Pils, S. Pinkert, J. Schultz, P. Bork, SMART 5: domains in the context of genomes and networks, Nucleic Acids Research 34 (2006) D257–D260.
- [5] R.L. Tatusov, N.D. Fedorova, J.D. Jackson, A.R. Jacobs, B. Kiryutin, E.V. Koonin, D.M. Krylov, R. Mazumder, S.L. Mekhedov, A.N. Nikolskaya, B.S. Rao, S. Smirnov, A.V. Sverdlov, S. Vasudevan, Y.I. Wolf, J.J. Yin, D.A. Natale, The COG database: an updated version includes eukaryotes, BMC Bioinformatics 4 (2003) 41.
- [6] A. Marchler-Bauer, J.B. Anderson, M.K. Derbyshire, C. DeWeese-Scott, N.R. Gonzales, M. Gwadz, L. Hao, S. He, D.I. Hurwitz, J.D. Jackson, Z. Ke, D. Krylov, C.J. Lanczycki, C.A. Liebert, C. Liu, F. Lu, S. Lu, G.H. Marchler, M. Mullokandov, J.S. Song, N. Thanki, R.A. Yamashita, J.J. Yin, D. Zhang, S.H. Bryant, CDD: a conserved domain database for interactive domain family analysis, Nucleic Acids Research 35 (2007) D237–D240.
- [7] R.D. Finn, J. Mistry, B. Schuster-Bockler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S.R. Eddy, E.L. Sonnhammer, A. Bateman, Pfam: clans, web tools and services, Nucleic Acids Research 34 (2006) D247– D251.
- [8] A.A. Schaffer, L. Aravind, T.L. Madden, S. Shavirin, J.L. Spouge, Y.I. Wolf, E.V. Koonin, S.F. Altschul, Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements, Nucleic Acids Research 29 (2001) 2994–3005.
- K.C. Chou, Prediction of protein cellular attributes using pseudo amino acid composition, Proteins: Structure, Function, and Genetics 43 (2001) 246–255 (Erratum: Proteins: Structure, Function, and Genetics 44 (2001) 60).
- [10] K.C. Chou, Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes, Bioinformatics 21 (2005) 10–19.
- [11] K.C. Chou, H.B. Shen, Predicting eukaryotic protein subcellular location by fusing optimized evidence-theoretic K-nearest neigh-

bor classifiers, Journal of Proteome Research 5 (2006) 1888-1897.

- [12] K.C. Chou, H.B. Shen, Hum-PLoc: a novel ensemble classifier for predicting human protein subcellular localization, Biochemical and Biophysical Research Communication 347 (2006) 150–157.
- [13] K.C. Chou, C.T. Zhang, Review: prediction of protein structural classes, Critical Reviews in Biochemistry and Molecular Biology 30 (1995) 275–349.
- [14] G.P. Zhou, An intriguing controversy over protein structural class prediction, Journal of Protein Chemistry 17 (1998) 729–738.
- [15] K.C. Chou, Y.D. Cai, Using functional domain composition and support vector machines for prediction of protein subcellular location, Journal of Biological Chemistry 277 (2002) 45765–45769.
- [16] Y. Huang, Y. Li, Prediction of protein subcellular locations using fuzzy k-NN method, Bioinformatics 20 (2004) 21–28.
- [17] C. Chen, X. Zhou, Y. Tian, X. Zou, P. Cai, Predicting protein structural class with pseudo-amino acid composition and support vector machine fusion network, Analytical Biochemistry 357 (2006) 116–121.
- [18] C. Chen, Y.X. Tian, X.Y. Zou, P.X. Cai, J.Y. Mo, Using pseudoamino acid composition and support vector machine to predict protein structural class, Journal of Theoretical Biology 243 (2006) 444–448.
- [19] P. Du, Y. Li, Prediction of protein submitochondria locations by hybridizing pseudo-amino acid composition with various physicochemical features of segmented sequence, BMC Bioinformatics 7 (2006) 518.
- [20] K.C. Chou, H.B. Shen, MemType-2L: a Web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM, Biochemical and Biophysical Research Communication 360 (2007) 339–345.
- [21] S. Mondal, R. Bhavna, R. Mohan Babu, S. Ramakumar, Pseudo amino acid composition and multi-class support vector machines approach for conotoxin superfamily classification, Journal of Theoretical Biology 243 (2006) 252–260.
- [22] H.B. Shen, K.C. Chou, Virus-PLoc: a fusion classifier for predicting the subcellular localization of viral proteins within host and virusinfected cells, Biopolymers 85 (2007) 233–240.
- [23] H. Lin, Q.Z. Li, Predicting conotoxin superfamily and family by using pseudo amino acid composition and modified Mahalanobis discrim-

inant, Biochemical and Biophysical Research Communication 354 (2007) 548-551.

- [24] H. Lin, Q.Z. Li, Using pseudo amino acid composition to predict protein structural class: approached by incorporating 400 dipeptide components, Journal of Computational Chemistry 28 (2007) 1463–1466.
- [25] H.B. Shen, K.C. Chou, Hum-mPLoc: an ensemble classifier for largescale human protein subcellular location prediction by incorporating samples with multiple sites, Biochemical and Biophysical Research Communication 355 (2007) 1006–1011.
- [26] Y. Cao, S. Liu, L. Zhang, J. Qin, J. Wang, K. Tang, Prediction of protein structural class with Rough Sets, BMC Bioinformatics 7 (20) (2006).
- [27] Q.B. Gao, Z.Z. Wang, Classification of G-protein coupled receptors at four levels, Protein Engineering, Design and Selection 19 (2006) 511–516.
- [28] K.D. Kedarisetti, L.A. Kurgan, S. Dick, Classifier ensembles for protein structural class prediction with varying homology, Biochemical and Biophysical Research Communication 348 (2006) 981–988.
- [29] K.C. Chou, H.B. Shen, Euk-mPLoc: a fusion classifier for large-scale eukaryotic protein subcellular location prediction by incorporating multiple sites, Journal of Proteome Research 6 (2007) 1728–1734.
- [30] S. Jahandideh, P. Abdolmaleki, M. Jahandideh, E.B. Asadabadi, Novel two-stage hybrid neural discriminant model for predicting proteins structural classes, Biophysical Chemistry 128 (2007) 87– 93.
- [31] K.C. Chou, H.B. Shen, Signal-CF: a subsite-coupled and windowfusing approach for predicting signal peptides, Biochemical and Biophysical Research Communication 357 (2007) 633–640.
- [32] X.B. Zhou, C. Chen, Z.C. Li, X.Y. Zou, Using Chou's amphiphilic pseudo-amino acid composition and support vector machine for prediction of enzyme subfamily classes, Journal of Theoretical Biology 248 (2007) 546–551.
- [33] K.C. Chou, H.B. Shen, Review: recent progresses in protein subcellular location prediction, Analytical Biochemistry 370 (2007) 1–16.
- [34] B. Rost, Enzyme function less conserved than anticipated, Journal of Molecular Biology 318 (2002) 595–608.