TABLE 1 - List of EC-classified, PLP-dependent enzyme activities

EC	*Enzyme name*	*Activity known to occur in:* Archaea	Bacteria	Fungi	Plants	Metazoa	**Known genes**
1.4.4.2	Glycine dehydrogenase	✓	✓	✓	✓	✓	✓
2.1.2.1	Glycine hydroxymethyltransferase	✓	✓	✓	✓	✓	✓
2.1.2.5	Glutamate formimidoyltransferase					✓	✓
2.1.2.7	D-alanine-2-hydroxymethyltransferase		✓				
2.3.1.29	Glycine C-acetyltransferase		✓			✓	✓
2.3.1.37	5-aminolevulinic acid synthase		✓	✓	✓	✓	✓
2.3.1.47	8-amino-7-oxononanoate synthase		✓	✓			✓
2.3.1.50	Serine C-palmitoyltransferase		✓	✓	✓	✓	✓
2.4.1.1	Phosphorylase		✓	✓	✓	✓	✓
2.5.1.47	Cysteine synthase	✓	✓	✓	✓		✓
2.5.1.48	Cystathionine gamma-synthase		✓	✓	✓		✓
2.5.1.49	O-acetylomhoserine (thiol)-lyase	✓	✓	✓			✓
2.5.1.51	Beta-pyrazolylalanine synthase				✓		
2.5.1.52	L-mimosine synthase				✓		
2.5.1.53	Uracilylalanine synthase				✓		
2.6.1.1	Aspartate aminotransferase	✓	✓	✓	✓	✓	✓
2.6.1.2	Alanine aminotransferase	✓		✓	✓	✓	✓
2.6.1.3	Cysteine aminotransferase					✓	
2.6.1.4	Glycine aminotransferase		✓		✓	✓	
2.6.1.5	Tyrosine aminotransferase		✓	✓	✓	✓	✓
2.6.1.6	Leucine aminotransferase					✓	
2.6.1.7	Kynurenine--oxoglutarate aminotransferase		✓	✓		✓	✓
2.6.1.8	2,5-diaminovalerate aminotransferase		✓	✓		✓	
2.6.1.9	Histidinol-phosphate aminotransferase	✓	✓	✓			✓
2.6.1.11	Acetylornithine aminotransferase	✓	✓	✓	✓		✓
2.6.1.12	Alanine--oxo-acid aminotransferase		✓				
2.6.1.13	Ornithine--oxo-acid aminotransferase		✓	✓	✓	✓	✓
2.6.1.14	Asparagine--oxo-acid aminotransferase					✓	
2.6.1.15	Glutamine--pyruvate aminotransferase					✓	
2.6.1.17	Succinyldiaminopimelate aminotransferase		✓				✓
2.6.1.18	Beta-alanine--pyruvate aminotransferase		✓		✓		✓
2.6.1.19	4-aminobutyrate aminotransferase		✓	✓	✓	✓	✓
2.6.1.21	D-alanine aminotransferase		✓		✓		✓
2.6.1.22	(S)-3-amino-2-methylpropanoate aminotransferase					✓	
2.6.1.23	4-hydroxyglutamate aminotransferase					✓	
2.6.1.24	Diiodotyrosine aminotransferase					✓	
2.6.1.26	Thyroid-hormone aminotransferase					✓	
2.6.1.27	Tryptophan aminotransferase		✓	✓	✓	✓	
2.6.1.28	Tryptophan--phenylpyruvate aminotransferase		✓				
2.6.1.29	Diamine aminotransferase		✓				

Table 1 (continued)

	Enzyme name	***Activity known to occur in:***					**Known genes**
		Archaea	Bacteria	Fungi	Plants	Metazoa	
2.6.1.33	dTDP-4-amino-4,6-dideoxy-D-glucose aminotransferase		✓				
2.6.1.34	UDP-4-amino-2-acetamido-2,4,6-trideoxyglucose aminotransferase		✓				
2.6.1.35	Glycine--oxaloacetate aminotransferase		✓				
2.6.1.36	L-lysine aminotransferase		✓	✓			✓
2.6.1.37	2-aminoethylphosphonate--pyruvate transaminase		✓				✓
2.6.1.38	Histidine aminotransferase		✓				
2.6.1.39	2-aminoadipate aminotransferase			✓		✓	✓
2.6.1.40	(R)-3-amino-2-methylpropionate--pyruvate aminotransferase					✓	
2.6.1.41	D-methionine--pyruvate aminotransferase				✓		
2.6.1.42	Branched-chain amino acid aminotransferase	✓	✓	✓	✓	✓	✓
2.6.1.43	Aminolevulinate aminotransferase		✓	✓	✓	✓	
2.6.1.44	Alanine--glyoxylate aminotransferase		✓	✓	✓	✓	✓
2.6.1.45	Serine--glyoxylate aminotransferase		✓		✓		✓
2.6.1.46	Diaminobutyrate--pyruvate aminotransferase		✓				✓
2.6.1.47	Alanine--oxomalonate aminotransferase					✓	
2.6.1.48	5-aminovalerate aminotransferase		✓	✓			
2.6.1.49	Dihydroxyphenylalanine aminotransferase		✓			✓	
2.6.1.50	Glutamine--scyllo-inosose aminotransferase		✓				✓
2.6.1.51	Serine--pyruvate aminotransferase				✓	✓	✓
2.6.1.52	Phosphoserine aminotransferase	✓	✓	✓	✓	✓	✓
2.6.1.55	Taurine aminotransferase		✓				
2.6.1.56	1D-1-guanidino-3-amino-1,3-dideoxy-scyllo-inositol aminotransferase		✓				
2.6.1.57	Aromatic amino acid transaminase	✓	✓	✓		✓	✓
2.6.1.58	Phenylalanine(histidine) aminotransferase					✓	
2.6.1.59	dTDP-4-amino-4,6-dideoxygalactose aminotransferase		✓				
2.6.1.60	Aromatic-amino-acid--glyoxylate aminotransferase					✓	
2.6.1.61	(R)-3-amino-2-methylpropanoate aminotransferase					✓	
2.6.1.62	Adenosylmethionine--8-amino-7-oxononanoate aminotransferase	✓	✓	✓			✓
2.6.1.63	Kynurenine--glyoxylate aminotransferase					✓	
2.6.1.64	Glutamine--phenylpyruvate aminotransferase		✓			✓	
2.6.1.65	N6-acetyl-beta-lysine aminotransferase		✓				
2.6.1.66	Valine--pyruvate aminotransferase		✓				✓
2.6.1.67	2-aminohexanoate aminotransferase			✓			
2.6.1.70	Aspartate--phenylpyruvate aminotransferase		✓				
2.6.1.71	Lysine--pyruvate 6-aminotransferase			✓			
2.6.1.72	D-4-hydroxyphenylglycine aminotransferase		✓				
2.6.1.73	Methionine--glyoxylate transaminase		✓		✓		
2.6.1.74	Cephalosporin-C transaminase		✓				
2.6.1.75	Cysteine-conjugate transaminase					✓	

Table 1 (continued)

EC	Enzyme name	Activity known to occur in: Archaea	Bacteria	Fungi	Plants	Metazoa	Known genes
2.6.1.76	Diaminobutyrate-2-oxoglutarate transaminase		✓				✓
2.9.1.1	L-seryl-tRNA(Sec) selenium transferase		✓				✓
3.5.99.7	1-aminocyclopropane-1-carboxylate deaminase		✓	✓			✓
3.7.1.3	Kynureninase		✓	✓		✓	✓
4.1.1.12	Aspartate 4-decarboxylase		✓	✓		✓	✓
4.1.1.14	Valine decarboxylase		✓		✓		
4.1.1.15	Glutamate decarboxylase		✓	✓	✓	✓	✓
4.1.1.16	Hydroxyglutamate decarboxylase		✓				
4.1.1.17	Ornithine decarboxylase		✓	✓	✓	✓	✓
4.1.1.18	Lysine decarboxylase		✓		✓		✓
4.1.1.19	Arginine decarboxylase		✓		✓	✓	✓
4.1.1.20	Diaminopimelate decarboxylase	✓	✓		✓		✓
4.1.1.22	Histidine decarboxylase		✓		✓	✓	✓
4.1.1.24	Aminobenzoate decarboxylase		✓				
4.1.1.25	Tyrosine decarboxylase		✓		✓		✓
4.1.1.28	Aromatic-L-amino-acid decarboxylase				✓	✓	✓
4.1.1.29	Sulfinoalanine decarboxylase					✓	✓
4.1.1.53	Phenylalanine decarboxylase		✓		✓	✓	
4.1.1.57	Methionine decarboxilase		✓		✓		
4.1.1.64	2,2-dialkylglycine decarboxylase (pyruvate)		✓	✓			✓
4.1.2.5	Threonine aldolase		✓	✓		✓	✓
4.1.2.26	Phenylserine aldolase					✓	
4.1.2.27	Sphinganine-1-phosphate aldolase			✓		✓	✓
4.1.99.1	Tryptophanase		✓				✓
4.1.99.2	Tyrosine phenol-lyase		✓				✓
4.2.1.20	Tryptophan synthase	✓	✓	✓	✓		✓
4.2.1.22	Cystathionine beta-synthase		✓	✓		✓	✓
4.2.1.50	Pyrazolylalanine synthase				✓		
4.2.3.1	Threonine synthase		✓	✓	✓		✓
4.2.3.2	Ethanolamine-phosphate phospho-lyase		✓			✓	
4.3.1.9	Glucosaminate ammonia-lyase		✓				
4.3.1.13	Carbamoyl-serine ammonia-lyase					✓	
4.3.1.15	Diaminopropionate ammonia-lyase		✓				✓
4.3.1.16	Threo-3-hydroxyaspartate ammonia-lyase		✓				
4.3.1.17	L-serine ammonia-lyase		✓	✓		✓	✓
4.3.1.18	D-serine ammonia-lyase		✓			✓	✓
4.3.1.19	Threonine ammonia-lyase		✓	✓	✓	✓	✓
4.3.1.20	Erythro-3-hydroxyaspartate ammonia-lyase		✓				
4.3.1.21	Aminodeoxygluconate ammonia-lyase		✓				
4.4.1.1	Cystathionine gamma-lyase		✓	✓	✓	✓	✓
4.4.1.2	Homocysteine desulfhydrase		✓				
4.4.1.4	Alliin lyase				✓		✓
4.4.1.6	S-alkylcysteine lyase		✓		✓		
4.4.1.8	Cystathionine beta-lyase		✓	✓	✓		✓
4.4.1.4	Alliin lyase				✓		✓

Table 1 (continued)

EC	*Enzyme name*	*Activity known to occur in:* Archaea	Bacteria	Fungi	Plants	Metazoa	**Known genes**
4.4.1.6	S-alkylcysteine lyase		✓		✓		
4.4.1.8	Cystathionine beta-lyase		✓	✓	✓		✓
4.4.1.9	L-3-cyanoalanine synthase		✓		✓		✓
4.4.1.10	Cysteine lyase					✓	
4.4.1.11	Methionine gamma-lyase		✓				✓
4.4.1.13	Cysteine-S-conjugate beta-lyase		✓	✓	✓	✓	
4.4.1.14	1-aminocyclopropane-1-carboxylate synthase				✓		✓
4.4.1.15	D-cysteine desulfhydrase		✓		✓		✓
4.4.1.16	Selenocysteine lyase		✓			✓	✓
4.5.1.2	3-chloro-D-alanine dehydrochlorinase		✓				
4.5.1.5	S-carboxymethylcysteine synthase		✓				
5.1.1.1	Alanine racemase		✓	✓		✓	✓
5.1.1.2	Methionine racemase		✓				
5.1.1.6	Threonine racemase		✓				
5.1.1.9	Arginine racemase		✓				
5.1.1.10	Amino-acid racemase		✓				
5.1.1.13	Aspartate racemase	✓				✓	
5.1.1.15	2-aminohexano-6-lactam racemase		✓				
5.1.1.17	Isopenicillin N-epimerase		✓				✓
5.4.3.2	Lysine 2,3-aminomutase		✓				✓
5.4.3.3	Beta Lysine 5,6-aminomutase		✓				✓
5.4.3.4	D-lysine 5,6-aminomutase		✓				✓
5.4.3.5	D-ornithine 4,5-aminomutase		✓				✓
5.4.3.8	Glutamate-1-semialdehyde 2,1-aminomutase	✓	✓		✓		✓

The Enzyme Commission (EC) numbering was updated on February 28, 2003 (http://www.chem.qmul.ac.uk/iubmb/enzyme/). The list only includes activities that were experimentally shown (or strongly suggested) to depend on pyridoxal phosphate. Some activities that could be reasonably assumed to be PLP-dependent [for example, Valine--3-methyl-2-oxovalerate aminotransferase, EC 2.6.1.32] but for which no sufficient experimental data could be found in the literature, were not included. Although biotin synthase (EC 2.8.1.6) was recently claimed to be a PLP-dependent enzyme, the issue is still under debate (Ollagnier-de-Choudens *et al*., 2002) and this activity, too, has been omitted from the list.

Note that some of the activities that have been included can be carried out both by PLP-dependent and by PLP-independent enzymes: pyruvoyl-dependent (PLP-independent) histidine decarboxylases (EC 4.1.1.22) and arginine decarboxylases (EC 4.1.1.19) are found in some prokaryotes (Gallagher *et al*., 1993; Graham *et al*., 2002); PLP-independent L-serine ammonia-lyases (EC 4.3.1.17) are used by some bacteria (Grabowski *et al*., 1993); and, finally, bacterial aspartate racemases (EC 5.1.1.13) do not employ cofactors although PLP-dependent enzymes have been recently described in some Archaea and metazoa (Long *et al*., 2001; Shibata *et al*., 2003). Note also that several characterized PLP-dependent enzymes have not yet been assigned an EC number. These include, for example, serine racemase (Wolosker *et al*., 1999), 4-amino-4-deoxychorismate lyase (Nakai *et al*., 2000) and nicotianamine aminotransferase (Takahashi *et al*., 1999).

Information about the occurrence of specific activities in different organisms has been taken from the specialized database BRENDA (Schomburg *et al*., 2002a; Schomburg *et al*., 2002b), and integrated by a direct survey of the literature. The occurrence of known sequences referring to a given EC number has been verified by querying the protein database in GenBank.

METHODS

Definition of PLP-dependent families

Sequences of PLP-dependent enzymes whose activity had been experimentally determined were assigned to the corresponding EC numbers. This initial assignment was based on information retrieved from the BRENDA database (Schomburg *et al*., 2002a; Schomburg *et al*., 2002b) and from the relevant literature. Sequences were then grouped into families. Families were defined as monophyletic groups of sequences all possessing the same enzymatic activity, according to Mehta and Christen (2000). Each EC number was assigned to one or more families based on this criterion. Families for eleven characterized PLP-dependent enzymes that have not yet been assigned an EC number were also considered: O-succinylhomoserine sulfhydrylase (Tate *et al*., 1999), nicotianamine aminotransferase (Takahashi *et al*., 1999), 4-aminobutyrate-pyruvate aminotransferase (Van Cauwenberghe *et al*., 2002), cystine C-S lyase (C-DES) (Lang and Kessler, 1999), 4-amino-4-deoxychorismate lyase (Nakai *et al*., 2000), cysteine desulfhydrase (Chu *et al*., 1997), cysteine desulfurase (NIFs-like enzymes) (Zheng *et al*., 1993), D-threonine aldolase (Liu *et al*., 1998), L-2,4-diaminobutyrate decarboxylase (Ikai and Yamamoto, 1994), L-threonine-*O*-3-phosphate decarboxylase (Brushaber *et al*., 1998) and serine racemase (Wolosker *et al*., 1999).

The number of sequences in individual families was then increased by homology searches. Uncharacterized sequences were added if they were found to be members of the group after phylogenetic analysis and visual inspection of the multiple alignment. Multiple alignments were constructed with ClustalW (Thompson *et al*., 1994). The ProDom program (Corpet *et al*., 2000) was used for alignment inspection and phylogenetic analysis. Family alignments were used to build Hidden Markov Models (HMM) with programs of the HMMER suite (Eddy, 1998). The score of each sequence was calculated with respect to a HMM including all family members except the sequence under examination. On the basis of this procedure, distributions of HMM scores for all families were obtained and then used for sequence classification.

Search for PLP-dependent enzymes in genomic sets of predicted proteins

Complete sets of protein sequences deduced from genomic data were generally obtained from the NCBI ftp repository (ftp://ftp.ncbi.nih.gov/genomes). The *N. crassa* set was obtained from the Center for Genomic Research at the Whitehead Institute/MIT (http://www-genome.wi.mit.edu). The

human set, updated on March 27, 2003, was obtained from the Ensembl ftp repository (ftp://ftp.ensembl.org/pub/current_human).

The classification of hypothetical proteins was achieved using a two-step procedure. First, each sequence was compared with our database of PLP-dependent enzymes (1,255 sequences) using BLAST ($E \leq 10^{-3}$). This step served as a quick filter to identify candidate genes coding for PLP-dependent enzymes. Candidates were subsequently compared with the library of family HMMs. This step was more time-consuming and served for the elimination of false positive hits and sequence classification.

The score of sequences significantly similar to a given family model ($E \leq 10^{-3}$) was compared with the previously calculated score distribution in order to determine the distance from the model in terms of standard deviations (s.d.) from the mean. Sequences with a score within 3 s.d. were assigned to the corresponding enzymatic activity. Sequences with a score below this threshold were marked as 'low-score' to indicate their low similarity to the family model. These sequences were not considered to possess the enzymatic function of the family, but were assumed to have an uncharacterized, possibly related, activity.

According to this analysis, few sequences exhibited highly significant similarity in pairwise comparison ($E \leq 10^{-12}$) without having significant similarity to any HMM. In such cases, sequences were considered to be potential PLP-dependent enzymes with an uncharacterized catalytic activity.

References

Brushaber, K.R., O'Toole, G.A. and Escalante-Semerena, J.C. (1998) CobD, a novel enzyme with L-threonine-O-3-phosphate decarboxylase activity, is responsible for the synthesis of (R)-1-amino-2-propanol-O-2-phosphate, a proposed new intermediate in cobalamin biosynthesis in *Salmonella typhimurium* LT2. *J. Biol. Chem.*, 273: 2684-2691.

Chu, L., Ebersole, J.L., Kurzban, G.P. and Holt, S.C. (1997) Cystalysin, a 46-kDa cysteine desulfhydrase from *Treponema denticola*, with hemolytic and hemoxidative activities. *Infect. Immun.*, 65: 3231-3238.

Corpet, F., Servant, F., Gouzy, J. and Kahn, D. (2000) ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Res.*, 28: 267-269.

Eddy, S.R. (1998) Profile hidden Markov models. *Bioinformatics*, 14: 755-63.

Gallagher, T., Rozwarski, D.A., Ernst, S.R. and Hackert, M.L. (1993) Refined structure of the pyruvoyl-dependent histidine decarboxylase from *Lactobacillus 30a*. *J. Mol. Biol.*, 230: 516-528.

Grabowski, R., Hofmeister, A.E. and Buckel, W. (1993) Bacterial L-serine dehydratases: a new family of enzymes containing iron-sulfur clusters. *Trends Biochem. Sci.*, 18: 297-300.

Graham, D.E., Xu, H. and White, R.H. (2002) *Methanococcus jannaschii* uses a pyruvoyl-dependent arginine decarboxylase in polyamine biosynthesis. *J. Biol. Chem.*, 277: 23500-23507.

Ikai, H. and Yamamoto, S. (1994) Cloning and expression in *Escherichia coli* of the gene encoding a novel L-2,4-diaminobutyrate decarboxylase of *Acinetobacter baumannii*. *FEMS Microbiol. Lett.*, 124: 225-228.

Lang, T. and Kessler, D. (1999) Evidence for cysteine persulfide as reaction product of L-Cyst(e)ine C-S-lyase (C-DES) from *Synechocystis*. Analyses using cystine analogues and recombinant C-DES. *J. Biol. Chem.*, 274: 189-195.

Liu, J.Q., Dairi, T., Itoh, N., Kataoka, M., Shimizu, S. and Yamada, H. (1998) A novel metal-activated pyridoxal enzyme with a unique primary structure, low specificity D-threonine aldolase from *Arthrobacter sp.* Strain DK-38. Molecular cloning and cofactor characterization. *J. Biol. Chem.*, 273: 16678-85.

Long, Z., Lee, J.A., Okamoto, T., Sekine, M., Nimura, N., Imai, K., Yohda, M., Maruyama, T., Sumi, M., Kamo, N., Yamagishi, A., Oshima, T. and Homma, H. (2001) Occurrence of D-Amino Acids and a pyridoxal 5'-phosphate-dependent aspartate racemase in the acidothermophilic archaeon, *Thermoplasma acidophilum*. *Biochem. Biophys. Res. Commun.*, 281: 317-321.

Mehta, P.K. and Christen, P. (2000) The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. *Adv. Enzymol.*, 74: 129-184.

Nakai, T., Mizutani, H., Miyahara, I., Hirotsu, K., Takeda, S., Jhee, K.H., Yoshimura, T. and Esaki, N. (2000) Three-dimensional structure of 4-amino-4-deoxychorismate lyase from *Escherichia coli*. *J. Biochem.*, 128: 29-38.

Ollagnier-de-Choudens, S., Mulliez, E. and Fontecave, M. (2002) The PLP-dependent biotin synthase from *Escherichia coli*: mechanistic studies. *FEBS Lett.*, 532: 465-8.

Schomburg, I., Chang, A., Hofmann, O., Ebeling, C., Ehrentreich, F. and Schomburg, D. (2002a) BRENDA: a resource for enzyme data and metabolic information. *Trends Biochem. Sci.*, 27: 54-56.

Schomburg, I., Chang, A. and Schomburg, D. (2002b) BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.*, 30: 47-49.

Shibata, K., Watanabe, T., Yoshikawa, H., Abe, K., Takahashi, S., Kera, Y. and Yamada, R.H. (2003) Purification and characterization of aspartate racemase from the bivalve mollusk *Scapharca broughtonii*. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.*, 134: 307-314.

Takahashi, M., Yamaguchi, H., Nakanishi, H., Shioiri, T., Nishizawa, N.K. and Mori, S. (1999) Cloning two genes for nicotianamine aminotransferase, a critical enzyme in iron acquisition (Strategy II) in graminaceous plants. *Plant Physiol.*, 121: 947-956.

Tate, R., Riccio, A., Caputo, E., Iaccarino, M. and Patriarca, E.J. (1999) The *Rhizobium etli* metZ gene is essential for methionine biosynthesis and nodulation of *Phaseolus vulgaris*. *Mol. Plant Microbe Interact.*, 12: 24-34.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22: 4673-4680.

Van Cauwenberghe, O.R., Makhmoudova, A., McLean, M.D., Clark, S.M. and Shelp, B.J. (2002) Plant pyruvate-dependent gamma-aminobutyrate transaminase:Identification of an *Arabidopsis* cDNA and its expression in *Escherichia coli*. *Can. J. Bot.*, 80: 933-941.

Wolosker, H., Blackshaw, S. and Snyder, S.H. (1999) Serine racemase: a glial enzyme synthesizing D-serine to regulate glutamate-N-methyl-D-aspartate neurotransmission. *Proc. Natl. Acad. Sci. U.S.A.*, 96: 13409-13414.

Zheng, L., White, R.H., Cash, V.L., Jack, R.F. and Dean, D.R. (1993) Cysteine desulfurase activity indicates a role for NIFS in metallocluster biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.*, 90: 2754-2758.