J Mol Evol (1990) 31:325–329

Journal of Molecular Evolution

Letter to the Editor

Prokaryotic and Eukaryotic Pyridoxal-Dependent Decarboxylases are Homologous

F. Rob Jackson

Neurobiology Division, The Worcester Foundation For Experimental Biology, 222 Maple Avenue, Shrewsbury, Massachusetts 01545, USA

Summary. A database search has revealed significant and extensive sequence similarities among prokaryotic and eukaryotic pyridoxal phosphate (PLP)-dependent decarboxylases, including *Drosophila* glutamic acid decarboxylase (GAD) and bacterial histidine decarboxylase (HDC). Based on these findings, the sequences of seven PLP-dependent decarboxylases from five different organisms have been aligned to derive a consensus sequence for this family of enzymes. In addition, quantitative methods have been employed to calculate the relative evolutionary distances between pairs of the decarboxylases comprising this family. The multiple sequence analysis together with the quantitative results strongly suggest an ancient and common origin for all PLP-dependent decarboxylases. This analysis also indicates that prokaryotic and eukaryotic HDC activities evolved independently. Finally, a sensitive search algorithm (PROFILE) was unable to detect additional members of this decarboxylase family in protein sequence databases.

Key words: PLP-dependent decarboxylase — Evolution — Profile analysis

Introduction

In vertebrates and invertebrates, several pyridoxal (PLP)-dependent decarboxylases, including dopa decarboxylase (DDC), glutamic acid decarboxylase (GAD), and histidine decarboxylase (HDC), catalyze the synthesis of neurotransmitters and neuromodulatory compounds (reviewed in Siegel et al. 1989). Although these and other PLP-dependent decarboxylases have quite distinct substrate-binding specificities, significant sequence similarities have been shown to exist among some of the enzymes (Kobayashi et al. 1987; De Luca et al. 1989; Jackson et al. 1990). These sequence similarities and the results of stereochemical studies (Dunathan and Voet 1974) have been interpreted as evidence for a common origin of PLP-dependent decarboxylases. Extensive sequence comparisons, however, have been reported only for eukaryotic decarboxylases (mammals, insects, and plants), although similarities have been documented for short peptides encompassing the PLP-binding domains of decarboxylases from *Escherichia coli* and pig kidney (Bossa et al. 1977). In addition, multiple sequence alignments, which might be useful for identifying functional domains, have not been performed for this family of enzymes. During the course of studying *Drosophila* glutamic acid decarboxylase (GAD), it came to my attention that bacterial histidine decarboxylase (HDC) has significant and extensive similarity to the other members of this decarboxylase family. This observation suggests that PLP-dependent decarboxylases have a very ancient evolutionary origin, a conclusion supported by the multiple sequence alignment and quantitative sequence comparisons reported in the present paper.

Offprint requests to: F.R. Jackson

Materials and Methods

All sequence comparisons and database searches including "Profile" analyses (Gribskov et al. 1987) were performed using the software of the Genetics Computer Group (Devereux et al. 1984) implemented on a VAX 11/750 running VMS 4.7. A derivative of the Dayhoff evolutionary distances matrix was used for the database searches of the profile analysis. The analyzed sequences were obtained from published reports or from release 22.0 of the NBRF:NEW protein sequence database (National Biomedical Research Foundation).

Results and Discussion

My laboratory recently deduced the amino acid sequence of *Drosophila* GAD (Jackson et al. 1990). A search of the NBRF:NEW protein database (release 22.0), using the FASTA program (Pearson and Lipman 1988), revealed that *Drosophila* GAD (fGAD) shares significant similarities with other eukaryotic PLP-dependent decarboxylases. Surprisingly, this database search also revealed an extensive similarity between fGAD and bacterial (*Morganella morganii*) histidine decarboxylase (bHDC; Vaaler et al. 1986). The two sequences are approximately 24% identical within a common 366-residue domain that includes the putative PLP-binding site of decarboxylases (Fig. 1). This is a striking degree of similarity, given the evolutionary distance between prokaryotes and eukaryotes. It supports the idea that prokaryotic and eukaryotic PLP-dependent decarboxylases originated from a common ancestral protein.

Other members of the PLP-dependent decarboxylase family, including dopa decarboxylase (DDC), show a comparable similarity to bHDC. This is worth mentioning, because it has recently been reported that rat DDC (rDDC) is not similar to bHDC (Tanaka et al. 1989). Figure 2 shows a multiple sequence alignment for seven different PLP-dependent decarboxylases, including bHDC, rat HDC (rHDC; Joseph et al. 1990), and rDDC, for which extensive primary sequence has been reported. This alignment includes plant tryptophan decarboxylase (pTDC; De Luca et al. 1989) as well as the *Drosophila* alpha-methyldopa hypersensitive protein (fAMD), which is evolutionarily related to DDCs (Eveleth and Marsh 1986; Marsh et al. 1986), although its precise function is unknown. The multiple sequence alignment was constructed after pairwise comparisons of the different sequences. It was then used to derive a consensus sequence for pyridoxal-dependent decarboxylases (Fig. 2). This consensus shows the extensive similarities that exist among all of these decarboxylases, in particular in a 90-residue segment (arrows, Fig. 2) encompassing the PLP-binding domain (rectangle, Fig. 2). Indeed, the putative PLP-binding residue (K) and the adjacent histidine (H)

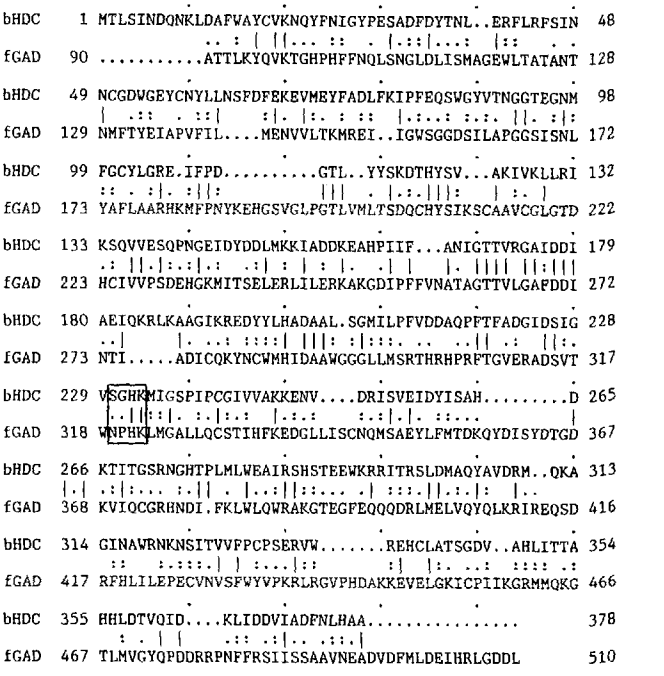

Fig. 1. Sequence alignment of *Drosophila* glutamic acid decarboxylase (fGAD) and bacterial histidine decarboxylase (bHDC). The numbering is, respectively, according to Jackson et al. (1990) and Vaaler et al. (1986). Vertical lines represent identities, whereas dots indicate similarities between residues. The box encompasses the putative pyridoxal phosphate-binding domain of fGAD.

are conserved in every one of these decarboxylases. Also of interest are the conserved Cys (C) residues at positions 385, 461, and 578, in light of evidence for disulfide linkage of decarboxylase subunits (Legay et al. 1987). Additional noteworthy identities are indicated by stars. The obvious similarities among these sequences strongly suggest that PLP-dependent decarboxylases have a common evolutionary origin. In addition to evolutionary considerations, these sequence alignments will be useful for identifying the regions that determine the diverse substrate specificities of PLP-dependent decarboxylases.

The similarity among rHDC and the other decarboxylases is also noteworthy. The rHDC enzyme is more closely related to members of the DDC/TDC subfamily of PLP-dependent decarboxylases, than to bHDC. The rHDC sequence, for example, is approximately 45% identical to either rDDC or fDDC (Joseph et al. 1990), but is only 15% identical to bHDC, the bacterial enzyme having the same substrate specificity. Thus, it appears that the prokaryotic and eukaryotic HDC activities evolved independently.

The multiple sequence alignment of Fig. 2 has been used with the Genetics Computer Group (GCG) "Distances" program (Devereux et al. 1984) to derive quantitative estimates of sequence relatedness for the PLP-dependent decarboxylase family. Table 1 shows a matrix displaying the pairwise genetic

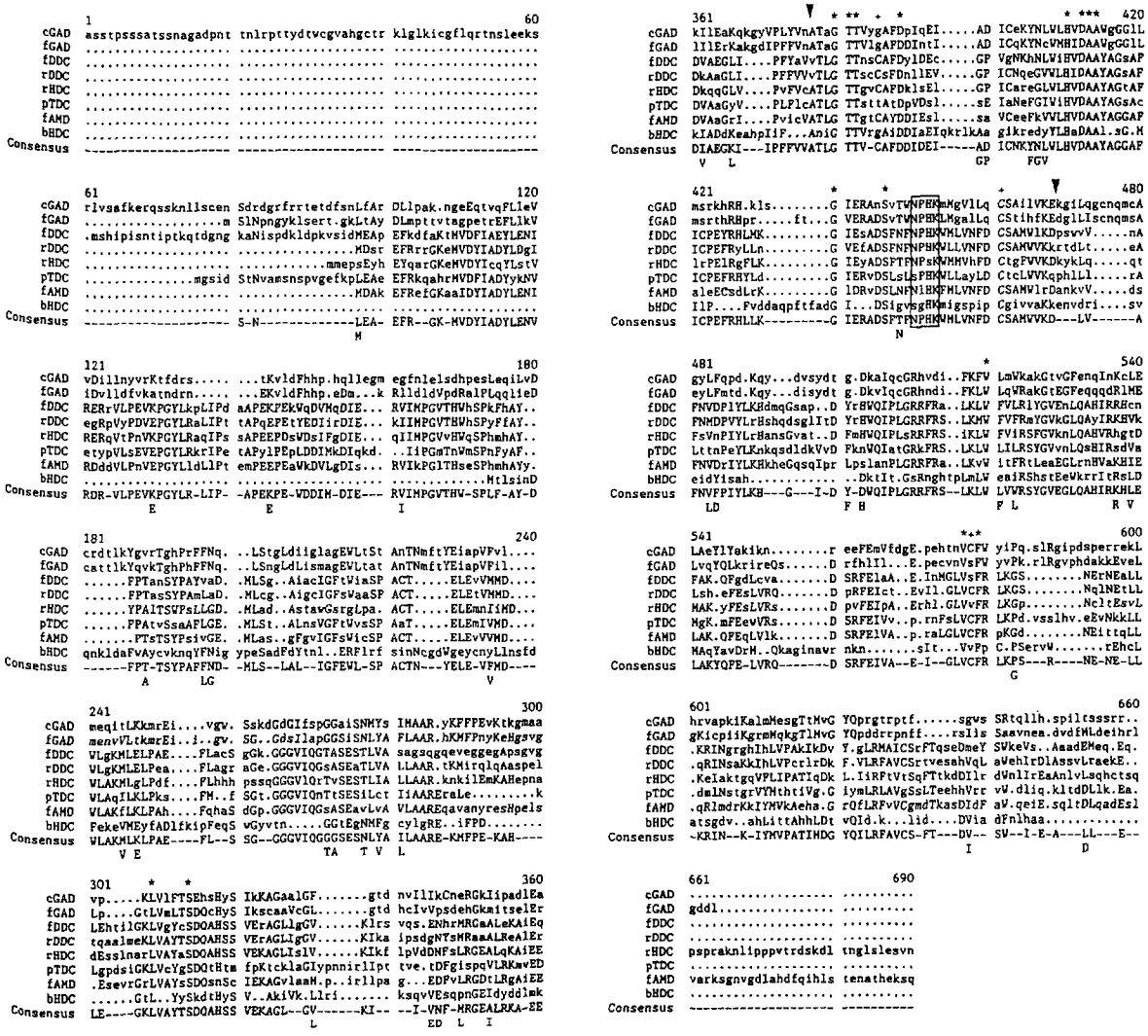

Fig. 2. Multiple sequence alignment of the pyridoxal phosphate (PLP)-dependent decarboxylase family. The indicated alignments are based upon pairwise comparisons of the PLP-dependent decarboxylases. Dots indicate pads that were inserted to optimize alignments. Uppercase letters indicate functional similarities or identities. Consensus residues represent at least three (out of eight) identities or three similarities (with at least two being identical). The rectangle encompasses the putative PLP-binding domain of these decarboxylases (the K residue of DDC is known to bind PLP; Bossa et al. 1977). Residues that are identical in all eight proteins are indicated with stars. Three conserved Cys (C) residues at positions 385, 461, and 578 are identified by a plus (+). The lengths of the various proteins, excluding pads, are given in Table 1. Note that the final 141 residues of rHDC are not included in this figure, as they cannot be aligned with any other decarboxylase. The group alignment is numbered beginning with residue 1 of cGAD. cGAD = cat glutamic acid decarboxylase (GAD) (Kobayashi et al. 1987); fGAD = fly (*Drosophila*) GAD (Jackson et al. 1990); fDDC = *Drosophila* dopa decarboxylase (Eveleth et al. 1986); rDDC = rat DDC (Tanaka et al. 1989); pTDC = *Catharanthus roseus* (periwinkle) tryptophan decarboxylase (De Luca et al. 1989); fAMD = *Drosophila* alpha-methyldopa hypersensitive protein (Marsh et al. 1986; Marsh, personal communication); bHDC = bacterial (*Morganella morganii*) histidine decarboxylase (Vaaler et al. 1986); rHDC = rat HDC (Joseph et al. 1990).

distances among these enzymes (a distance value represents the number of identical matches between a pair, excluding gaps, divided by the shorter sequence length; therefore, a value of 1 indicates a perfect match). This quantitative assessment confirms that cGAD (from cat) and fGAD are more closely related to one another than to members of the DDC/TDC subfamily. Likewise, pTDC (from periwinkle), fAMD, and the DDCs are highly related to one another (also see De Luca et al. 1989). Importantly, this analysis also underscores the relatedness of bacterial HDC to eukaryotic decarboxylases.

All of the enzymes of Fig. 2, including bHDC, have small segments in common with mouse ornithine decarboxylase (mODC; Kahana and Nathans 1985), another PLP-dependent decarboxylase (data not shown and De Luca et al. 1989). At best, however, mODC has a 10–12% overall identity with any of the other decarboxylases. Although mouse ODC and human ODC sequences are present in the NBRF database, they were not detected by FASTA

Table 1. Distance matrix

	1	2	3	4	5	6	7	8
1	1.00	0.53	0.17	0.17	0.17	0.17	0.17	0.19
2		1.00	0.19	0.18	0.17	0.17	0.17	0.24
3			1.00	0.58	0.45	0.40	0.45	0.15
4				1.00	0.51	0.41	0.45	0.13
5					1.00	0.36	0.39	0.15
6						1.00	0.37	0.13
7							1.00	0.13
8								1.00

Key for column and row indices: 1, cGAD, length without gaps: 585; 2, fGAD, length without gaps: 510; 3, fDDC1, length without gaps: 511; 4, rDDC, length without gaps: 480; 5, rHDC, length without gaps: 514; 6, pTDC, length without gaps: 500; 7, fAMD, length without gaps: 510; 8, bHDC, length without gaps: 378. See Fig. 2 for abbreviations

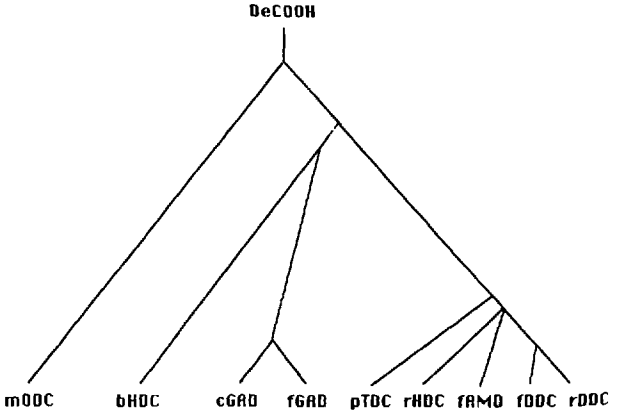

Fig. 3. Schematic representation of decarboxylase evolution. Abbreviations are the same as in Fig. 2, excluding mODC which represents mouse ornithine decarboxylase (Kahana and Nathans 1985).

searches with multiple individual decarboxylase sequences. Therefore, among the known PLP-dependent decarboxylases, the ODCs appear to be the most divergent. Figure 3 shows a possible phylogeny of the characterized PLP-dependent decarboxylases, including mODC. This pictorial representation summarizes the relationships among these decarboxylases, but cannot be thought of as quantitative because of likely differences in the rates of evolution among the different enzymes and species.

Profile analysis is possibly the most sensitive method for detecting distantly related members of a family of proteins; it utilizes dynamic programming algorithms to search databases for proteins with similarities to an aligned group of sequences (Gribskov et al. 1987). To perform such searches, the information in the multiple sequence alignment is represented quantitatively in a table of position-specific symbol comparison values (a profile), which can then be compared to existing protein sequence databases.

This method was used to derive a profile for the decarboxylase alignment shown in Fig. 2. This "decarboxylase profile" was then employed to search the PIR:NEW database for additional sequences with similarity to the group. This search found fDDC, fAMD, and bHDC (already known to be in the database), but did not detect ODCs or any other proteins with significant similarity to the PLP-dependent decarboxylase group (note that the GADs and pTDC are not present in this database). Importantly, the search did not detect numerous non-PLP-dependent decarboxylases present in the database. Thus, it appears that the family of PLP-dependent decarboxylases is unrelated to other characterized proteins.

In conclusion, the results of these analyses suggest an ancient evolutionary origin for the gene duplications that gave rise to the PLP-dependent decarboxylase family. Interestingly, the data also imply that histidine decarboxylase activity evolved independently within eukaryotic and prokaryotic species.

Acknowledgments. I thank L. Newby for help with some of the sequence analyses; Larry Marsh for information regarding the fAMD sequence; L. Newby, R. O'Connell, and S. Kulkarni for reading drafts of the manuscript; H. Bayley for a consult on sequence evolution; and D. Yang for assistance with the VAX/VMS system.

References

Bossa F, Martini F, Barra D, Voltattorni B, Minelli A, Turano C (1977) The chymotryptic phosphopyridoxyl peptide of dopa decarboxylase from pig kidney. Biochem Biophys Res Comm 78:177–184

De Luca V, Marineau C, Brisson N (1989) Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopa decarboxylases. Proc Natl Acad Sci USA 86:2582–2586

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Dunathan HC, Voet JG (1974) Stereochemical evidence for the evolution of pyridoxal-phosphate enzymes of various function from a common ancestor. Proc Natl Acad Sci USA 71: 3888–3891

Eveleth DD, Marsh JL (1986) Evidence for evolutionary duplication of genes in the dopa decarboxylase region of *Drosophila*. Genetics 114:469–483

Eveleth DD, Gietz RD, Spencer CA, Nargang FE, Hodgetts RB, Marsh JL (1986) Sequence and structure of the dopa decarboxylase gene of *Drosophila*: evidence for novel RNA splicing variants. EMBO J 5:2663–2672

Gribskov M, McLachlan AD, Eisenberg D (1987) Profile analysis: detection of distantly related proteins. Proc Natl Acad Sci USA 84:4355–4358

Jackson FR, Newby LM, Kulkarni SJ (1990) *Drosophila* GABAergic systems: sequence and expression of glutamic acid decarboxylase. J. Neurochem 54:1068–1078

Joseph DR, Sullivan PM, Wang Y-M, Kozak C, Fenstermacher DA, Behrendsen ME, Zahnow CA (1990) Characterization and expression of the complementary DNA encoding rat histidine decarboxylase. Proc Natl Acad Sci USA 87:733–737

Kahana C, Nathans D (1985) Nucleotide sequence of murine ornithine decarboxylase mRNA. Proc Natl Acad Sci USA 82: 1673–1677

Kobayashi Y, Kaufman DL, Tobin A (1987) Glutamic acid decarboxylase cDNA: nucleotide sequence encoding an enzymatically active fusion protein. J Neurosci 7:2768–2772

Legay F, Henry S, Tappaz M (1987) Evidence for two distinct forms of native glutamic acid decarboxylase in rat brain soluble extract: an immunoblotting study. J Neurochem 48:1022–1026

Marsh JL, Erfle M, Leeds C (1986) Molecular localization, developmental expression and nucleotide sequence of the *alpha-methyldopa hypersensitive* gene of *Drosophila*. Genetics 114:453–467

Pearson WR, Lipman DJ (1988) Improved tools for biological sequence analysis. Proc Natl Acad Sci USA 85:2444–2448

Siegel G, Agranoff B, Albers RW, Molinoff P (1989) Basic neurochemistry, ed 4. Raven Press, New York

Tanaka T, Horio Y, Taketoshi M, Imamura I, Ando-Yamamoto M, Kangawa K, Matsuo H, Kuroda M, Wada H (1989) Molecular cloning and sequencing of a cDNA of rat dopa decarboxylase: partial amino acid homologies with other enzymes synthesizing catecholamines. Proc Natl Acad Sci USA 86: 8142–8146

Vaaler GL, Brasch MA, Snell EE (1986) Pyridoxal 5'-phosphate-dependent histidine decarboxylase: nucleotide sequence of the *hdc* gene and the corresponding amino acid sequence. J Biol Chem 261:11010–11014

Received January 15, 1990/Revised April 25, 1990/Accepted May 10, 1990