Functional Classification of Amino Acid Decarboxylases from the Alanine Racemase Structural Family by Phylogenetic Studies

Heidi Kidron, Susanna Repo, Mark S. Johnson, and Tiina A. Salminen

Department of Biochemistry and Pharmacy, Åbo Akademi University, Turku, Finland

Arginine decarboxylase (ADC) and ornithine decarboxylase (ODC) are involved in the biosynthesis of putrescine, which is the precursor of other polyamines in animals, plants, and bacteria. These pyridoxal-5′-phosphate–dependent decarboxylases belong to the alanine racemase (AR) structural family together with diaminopimelate decarboxylase (DapDC), which catalyzes the final step of lysine biosynthesis in bacteria. We have constructed a multiple-sequence alignment of decarboxylases in the AR structural family and, based on the alignment, inferred phylogenetic trees. The phylogenetic tree consists of 3 distinct clades formed by ADC, DapDC, and ODC that diverged from an ancestral decarboxylase. The ancestral decarboxylase probably was able to recognize several substrates, and in archaea and bacteria, ODC may have retained the ability to bind other amino acids. Previously, a parologue of ODC has been proposed to account for ADC activity detected in mammalian cells. According to our results, this appears unlikely, emphasizing the need for more caution in functional assignment made using sequence data and illustrating the continuing value of phylogenetic analysis in clarifying relationships and putative functions.

Introduction

Pyridoxal-5′-phosphate (PLP)–dependent enzymes are a large group of proteins that are mostly involved in amino acid metabolism. Their diverse reaction specificities, including α decarboxylation, β and γ elimination and replacement reactions, transamination, racemization, as well as aldol cleavage, have evolved before their substrate specificities (Alexander et al. 1994). The PLP-dependent enzymes can be divided into 5 families based on sequence similarity (Grishin et al. 1995) and structural comparisons (Jansonius 1998). The 5 families are not homologues of each other, but surprisingly the different folds form a structurally similar PLP-binding core, where phosphate is bound in a similar way whereas the pyridoxal ring is not (Denesiouk et al. 1999; Denesyuk et al. 2003). Jansonius (1998) named the families after the first enzyme whose 3-dimensional structure had been determined in each family: the aspartate aminotransferase family, the d-amino acid transferase family, the glycogen phosphorylase family, the tryptophan synthase β family, and the alanine racemase (AR) family. In addition to alanine racemase, the AR family contains prokaryotic AR, prokaryotic biosynthetic arginine decarboxylase (ADC), diaminopimelate decarboxylase (DapDC), and eukaryotic ornithine decarboxylase (ODC). ADC, DapDC, and ODC have low sequence identity, only 20–30%, but the sequence similarities had been determined in each family: the aspartate aminotransferase family, the d-amino acid transferase family, the glycogen phosphorylase family, the tryptophan synthase β family, and the alanine racemase (AR) family. In addition to alanine racemase, the AR family contains prokaryotic AR, prokaryotic biosynthetic arginine decarboxylase (ADC), diaminopimelate decarboxylase (DapDC), and eukaryotic ornithine decarboxylase (ODC). ADC, DapDC, and ODC have low sequence identity, only 20–30%, but the sequence similarity between the decarboxylases and AR is very low.

The crystal structures of ARs (Shaw et al. 1997; Fenn et al. 2003), DapDCs (Ray et al. 2002; Gokulan et al. 2003), and ODCs (Kern et al. 1999; Almrud et al. 2000; Jackson et al. 2003) with different ligands have been solved, but no experimentally determined structure of an ADC has been reported to date. AR, DapDC, and ODC are dimeric proteins where each monomer consists of 2 domains, the N-terminal α/β-barrel domain binding PLP and the C-terminal β-sandwich domain, which are slightly rotated in respect to each other in the different enzymes (Shaw et al. 1997; Kern et al. 1999; Ray et al. 2002). The 2 active sites, where PLP is bound, are formed by residues from both the N-terminal domain and the C-terminal domain of the different monomers of the decarboxylases. Compared with DapDCs and ODCs, the ADC sequences contain an insertion of approximately 85 amino acids that is thought to form an additional domain, introduced between the N- and C-terminal domains (Jantaro et al. 2006).

The reaction mechanism of ODC has been studied most, but the enzymatic mechanism of ADC, DapDC, and ODC is assumed to be similar (Brooks and Phillips 1997; Jackson et al. 2000). The reaction begins when the PLP-binding Lys69 (numbering according to human ODC) is displaced by L-ornithine, which is then decarboxylated. PLP functions as an electron sink to stabilize the decarboxylated substrate. The protonation of the decarboxylated substrate is controlled by Cys360 that rotates into the active site (Jackson et al. 2000). The release of the product is the rate-limiting step in the reaction (Brooks and Phillips 1997).

ADC and ODC catalyze reactions in alternative pathways of putrescine biosynthesis. Putrescine is the precursor of all polyamines, which are involved in growth and developmental processes of many organisms. In animals, there is only 1 pathway for putrescine biosynthesis: the direct synthesis of putrescine from L-ornithine by ODC (Coleman et al. 2004). In some species, ODC activity is regulated by an antizyme that binds to ODC and targets it for degradation (Murakami et al. 1992). In addition to ODC, to date 2 other types of ODC-like protein sequences have been found in mammals—one is an antizyme inhibitor (AzI) (Murakami et al. 1996) and the other an ODC parologue (ODCp) that has been proposed to be either an AzI or an ADC (Pitkänen et al. 2001; Zhu et al. 2004). In plants and bacteria, there is an additional pathway leading to the synthesis of putrescine, where L-arginine is first decarboxylated to agmatine by ADC. In plants, agmatine is then modified in a 2-step process to putrescine, but in bacteria, agmatine is directly converted to putrescine. The plant ADC is located mainly in chloroplasts and has been suggested to be of cyanobacterial origin, whereas the plant, fungal, and animal ODCs are found mainly in the cytoplasm and probably originate from other forms of bacteria (Borrell et al. 1995; Illingworth et al. 2003; Bortolotti et al. 2004).

DapDC is expressed in bacteria and in plants, where it catalyzes the final step in lysine biosynthesis (Velasco et al.
Kidron et al. 2002; Hudson et al. 2005). Lysine is needed for bacterial survival, and the DapDC gene, LysA, is essential for bacterial vitality (Pavelka and Jacobs 1996). DapDC has previously been recognized to be evolutionary related to eukaryotic ODCs and prokaryotic biosynthetic ADCs (Martin et al. 1988; Sandmeier et al. 1994). However, the evolutionary and functional relationships of the 3 decarboxylases and other highly similar sequences, like the AzIs, have not been studied in detail. We have now collected 130 protein sequences of amino acid decarboxylases from the AR structural family, and based on their multiple-sequence alignment, we have inferred phylogenetic trees. In addition, we have constructed structural models of the human AzI and ODCp proteins in order to compare their PLP-binding properties with ODC and to evaluate the suggested ADC activity (Zhu et al. 2004) of ODCp. Here, phylogenetic analyses led to a more detailed understanding of the evolutionary relationships between the enzymes and assisted in the classification of newly identified decarboxylases whose functions were uncertain.

Methods
Sequence Data
The sequences of ADC, AzI, DapDC, ODC, and ODCp, ranging from ~400 to ~700 amino acid residues, were obtained from the UniProt Knowledgebase (Bairoch et al. 2005). In order to identify potential decarboxylase sequences that are not yet identified by UniProt, we performed searches of sequences obtained from bacterial and archaeal genome projects using the human ODC protein sequence as the query and the genomic version of Blast (Cummings et al. 2002) provided by National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/BLAST/). Additional searches using Blast and the DapDC sequence of Escherichia coli as the query sequence identified putative DapDC sequences from 2 plants, Arabidopsis thaliana and Oryza sativa. All sequences analyzed in this study are listed in supplementary table 1, Supplementary Material online. Herein, the name of each sequence consists of 2 parts: the first part identifies the protein (ADC, AzI = AZIN, DapDC = DCDA, ODC, and ODCp), and the second part identifies the organism according to codes used by UniProt (Bairoch et al. 2005; http://www.ebi.uniprot.org/support/documents.shtml). In those cases where sequences were not available from UniProt, the second part of the sequence name consists of the first 3 letters of the genus followed by the first 2 letters of the species.

The ODC sequence from Leishmania donovani contains a large N-terminal extension not present in the ODC sequences from other organisms (Hanson et al. 1992). This 200-residue long extension was removed from the L. donovani sequence, and only residues 201–707 were used in the analysis.

Phylogeny Reconstruction
Automatic multiple-sequence alignments were performed with MALIGN (Johnson and Overington 1993; Johnson et al. 1996). Due to the putative inserted domain present in ADCs, a previously published alignment of 2 ADC, 2 DapDC, and 2 ODC sequences (Jantaro et al. 2006) was used as the initial alignment in this study. The initial alignment was separately aligned with the DapDC, ODC (together with the AzI and ODCp sequences), and the ADC sequences. Finally, the 3 separate alignments were combined into 1 multiple-sequence alignment of 130 sequences (see Supplementary Material online) in the Alignment Editor of Bodil (Lehtonen et al. 2004).

Several programs in the PHYLIP package (version 3.65; Felsenstein 1989) were used to reconstruct phylogenies. The Jones–Taylor–Thornton model implemented in PROTDIST was used to calculate a distance matrix for the multiple-sequence alignment. In order to evaluate the confidence of the inferred tree, a bootstrap set of 1,000 pseudoalignments was generated by SEQBOOT. Thereafter, distance matrices for each pseudoalignment were calculated with PROTDIST. The program NEIGHBOR was used to compute a Neighbor-Joining (NJ) tree from each distance matrix, and finally a consensus tree was computed with CONSENSE. In order to combine the topology of the consensus tree with the distances from the original matrix, the program FITCH was used. Finally, a picture of the inferred tree was generated with the programs DRAWGRAM and DRAWTREE. A maximum likelihood (ML) method for phylogeny estimation, PROML, was also used, again with a bootstrap set of 1,000 pseudoalignments, and the consensus tree was computed with CONSENSE.

Modeling
Structural models of human AzI and ODCp were made with MODELLER 8.1 (Sali and Blundell 1993) based on the multiple-sequence alignment, and the human ODC crystal structure was used as the structural template (Protein Data Bank code 1D7K; Almrud et al. 2000). After visual examination, the model with the lowest objective function derived by MODELLER was chosen for further study in each case. The quality of each model was evaluated with PROSAII (Sippl 1993).

Visualization
The Bodil modeling environment was used for the visualization and superimposition of crystal structures (Lehtonen et al. 2004). Figures 1–4 and supplementary figures 1–3, Supplementary Material online were finalized using Gimp 2.2 and Inkscape 0.43. Figures 5A and 6 were produced with Pymol (version 0.99, DeLano 2002), and the labels were added with Gimp 2.2. The figure containing the multiple-sequence alignment (Supplementary Material online) was produced using Ascript version 2.0 (Barton 1993). Due to limitations of space, some trees are not reported in the text or Supplementary Material online, but they can be obtained by contacting the authors.

Results and Discussion
We have collected sequences of amino acid decarboxylases from the AR structural family representing all levels of the tree of life, which includes 33 ADC, 47 DapDC, 32 ODC, 2 AzI, and 3 ODCp sequences. In addition, 10 bacterial sequences and 3 archaeal sequences, which are clearly homologous to ODC according to the Blast searches, were included in the study, even though they had not yet been annotated (supplementary table 1, Supplementary Material
online). We decided to limit the total number of sequences in the study to 130 in order to avoid redundancies in the trees caused by having very similar sequences from closely related species. The program hmmer (http://www.wustl.edu/) and 2 hmmer profiles constructed from the original alignment of the 130 decarboxylase sequences identified 1,000 hits from GenBank (release153, Benson et al. 2006) each with $e$ values between $7 \times 10^{-242}$ and 10. In each case, these 130 sequences distribute quite evenly among the 1,000 hmmer hits, and we can thus conclude that our set of 130 sequences represents well the decarboxylases and that the set includes both closely and more distantly related enzymes. The bootstrap method of Felsenstein (1985) was used to evaluate the confidence of the reconstructed trees;

FIG. 1.—Unrooted consensus tree of amino acid decarboxylases from the AR structural family. The branches are scaled according to evolutionary distance. There are 3 main clades that are formed by ODC (A), ADC (B), and DapDC (C) sequences.
From the sequence alignment. In contrast, no distance matrix is used in the ML method; instead, the starting point is a random tree, which is modified until the highest likelihood of producing the original multiple-sequence alignment is achieved. The same set of 1,000 pseudoalignments, derived from the multiple-sequence alignment of the 130 amino acid decarboxylases of the AR structural family, was used to assess the confidence of both methods. The overall topology of the NJ tree is congruent with the topology of the ML tree; however, within the main branches, there are few minor differences. We will first discuss the topology of the NJ tree and then compare the topological differences between the NJ and ML trees.

The phylogenetic tree of the AR structural family of amino acid decarboxylases contains 3 main branches (fig. 1). As expected, these 3 main groups are formed by ADCs, DapDCs, and ODCs. In figure 1, the branch lengths have been scaled to indicate the relative evolutionary distance between the decarboxylases. Comparison of the distances among the 3 clades implies that the DapDC and ODC are more similar to each other than either group is to the ADC group. The distance between ADCs and DapDCs or ODCs is more than double compared with the distance between DapDCs and ODCs.

Members of the ADC subfamily contain an inserted domain of ~85 residues, which is found neither in the DapDCs nor in the ODCs. We have previously published an alignment (Jantaro et al. 2006), where the inserted domain of the ADCs aligns as a separate domain. In the current study, that alignment was used as the basis for the multiple-sequence alignment of the 130 sequences as described in the Methods. In order to assess the effect of the inserted domain on the phylogeny reconstruction, we removed the inserted domain from the ADC sequences in the alignment and computed the phylogenetic tree. As expected, the topology of this tree is identical to the tree derived from the whole alignment (data not shown). In addition, the evolutionary distances are equally long in both trees. This suggests that the inserted domain does not affect the evolutionary relationships among the ADCs, DapDCs, and ODCs.

In order to evaluate whether the use of the initial alignment (see Methods) affected the topology of the tree, we aligned a subset of sequences using a different method. First, 44 sequences from the collected set of 130 decarboxylases were randomly chosen and aligned. The alignment included 14 DapDCs, 14 ODCs, 1 ODCp, and 1 AzI sequence, together with 14 ADC sequences from which the inserted domain had been removed. A phylogenetic tree was inferred from the alignment using the NJ method. The topology of the tree and the evolutionary distances (data not shown) are very similar to the tree depicted in figure 1. There are 3 clearly distinguishable branches, which are formed by the ADCs, DapDCs, and ODCs. Thus, we can conclude that in our case, the use of a previously aligned set of sequences to seed a larger alignment did not bias the phylogeny reconstruction.

Topological differences between the NJ and ML trees imply that the DapDC and ODC are more similar to each other than either group is to the ADC group. The distance between ADCs and DapDCs or ODCs is more than double compared with the distance between DapDCs and ODCs.

In order to evaluate whether the use of the initial alignment (see Methods) affected the topology of the tree, we aligned a subset of sequences using a different method. First, 44 sequences from the collected set of 130 decarboxylases were randomly chosen and aligned. The alignment included 14 DapDCs, 14 ODCs, 1 ODCp, and 1 AzI sequence, together with 14 ADC sequences from which the inserted domain had been removed. A phylogenetic tree was inferred from the alignment using the NJ method. The topology of the tree and the evolutionary distances (data not shown) are very similar to the tree depicted in figure 1. There are 3 clearly distinguishable branches, which are formed by the ADCs, DapDCs, and ODCs. Thus, we can conclude that in our case, the use of a previously aligned set of sequences to seed a larger alignment did not bias the phylogeny reconstruction.

Topological differences between the NJ and ML trees imply that the DapDC and ODC are more similar to each other than either group is to the ADC group. The distance between ADCs and DapDCs or ODCs is more than double compared with the distance between DapDCs and ODCs.

In order to evaluate whether the use of the initial alignment (see Methods) affected the topology of the tree, we aligned a subset of sequences using a different method. First, 44 sequences from the collected set of 130 decarboxylases were randomly chosen and aligned. The alignment included 14 DapDCs, 14 ODCs, 1 ODCp, and 1 AzI sequence, together with 14 ADC sequences from which the inserted domain had been removed. A phylogenetic tree was inferred from the alignment using the NJ method. The topology of the tree and the evolutionary distances (data not shown) are very similar to the tree depicted in figure 1. There are 3 clearly distinguishable branches, which are formed by the ADCs, DapDCs, and ODCs. Thus, we can conclude that in our case, the use of a previously aligned set of sequences to seed a larger alignment did not bias the phylogeny reconstruction.

Topological differences between the NJ and ML trees imply that the DapDC and ODC are more similar to each other than either group is to the ADC group. The distance between ADCs and DapDCs or ODCs is more than double compared with the distance between DapDCs and ODCs.
whereas the other branch contains the bacterial ADC sequences. The bacterial ADCs form 2 branches with 53% bootstrap support, in which the cyanobacterial ADCs cluster together with bacteroidetes, planctomycetes, and Deinococcus-Thermus ADCs, whereas the beta- and gamma-proteobacterial ADCs form the other branch.

The DapDC sequences segregate into 4 main branches (fig. 3), which appear to have diverged over a relatively short evolutionary time frame. Under such a short time frame, the number of substitutions supporting each node and the data set as a whole remain small and is particularly sensitive to random fluctuations in observed mutations and small differences in the mutation rate, which ultimately leads to low bootstrap values (Soltis DE and Soltis PS 2003). The first cluster of DapDC sequences is formed by 4 epsilon-proteobacteria, whereas the second branch comprises a clearly defined cluster of cyanobacteria, firmicutes, and actinobacteria sequences. The third and fourth branches of the DapDC sequences diverge with low bootstrap support (10%). The branches consist of the alpha-, beta-, delta-, and gamma-proteobacteria DapDCs, together with the aquificae, chlorobi, and thermotogae sequences, and the 2 viridiplantae sequences identified by a Blast search against the genomes of *A. thaliana* and *O. sativa*. The bootstrap support, however, is very low (11%) for the viridiplantae branch. The 5 archaeal DapDC sequences form 2 branches with relatively low bootstrap values (6–33%). In addition, the distances among the archaea sequences are very long within the branches. The DapDC of *Deinococcus radiodurans*, which diverges from the rest of the DapDC sequences with a bootstrap support of 99%, lies completely outside the cluster of DapDC sequences. The phylogenetic tree of Hudson et al. (2005), which consists of 33 DapDC sequences but where the significance and robustness of the branching order of the tree were not evaluated, resembles our tree both in the general topology and the short internal branch lengths associated with the earlier diverging members of the DapDC sequences.

With the DapDCs, 1 of the archaea branches is formed by sequences that were identified in Blast searches of the genomes of *Methanococcoides burtonii* and *Ferroplasma acidarmanus*. The classification of the *M. burtonii* and *F. acidarmanus* decarboxylases as DapDCs is supported by the presence of 2 conserved residues, glutamate and tyrosine, which are found in all DapDCs, whereas the corresponding residues in the human ODC are aspartate (Asp361) and phenylalanine (Phe397). The histidine that in the known decarboxylase structures stacks with the pyridoxal ring of PLP and is conserved across the decarboxylases is, however, not conserved in the *M. burtonii* and *F. acidarmanus* DapDCs, where the corresponding residue

---

**Fig. 3.**—The consensus tree of DapDC sequences. The root indicates the point where the 3 families of decarboxylases diverge from each other, see figure 1. The branches are scaled according to distance. The percentage of bootstrap support for each node is described as in figure 2.
is a methionine, which is also capable of interacting favorably with the ring surface. Even if the catalytic and the other residues that interact with PLP are conserved, the histidine to methionine mutation at the active site raises the question as to whether the DapDCs of *M. burtonii* and *F. acidarmanus* are indeed functional PLP-binding decarboxylases. If the function of histidine is limited to interacting with the PLP ring, as opposed to a direct role in catalysis, then methionine should be able to take on that role with little effect on catalysis.

The other putative decarboxylase (putative DC) sequences, which were identified in bacteria and archaea by the Blast searches, cluster within the branch containing the ODC sequences (fig. 4). Nonetheless, these sequences are clearly segregated from the eukaryotic ODCs with a bootstrap support of 98%. The ODC of *Selenomonas ruminantium*, which has both lysine and ODC activity (Takatsuka et al. 2000), is separated from the other prokaryotic ODC sequences and forms a distinct branch together with ODC from *Rhodopirellula baltica*. According to the GenBank annotation for the *R. baltica* ODC, the protein is a putative lysine/ornithine decarboxylase.

The eukaryotic plant, fungal, and animal ODC sequences separate into 4 distinct branches (fig. 4). The plant ODCs segregate with a bootstrap value of 57%, and the branch that includes the arthropod and nematode ODCs segregate with a bootstrap support of 96%. The fungal branch segregates from the chordates with a bootstrap support of 77%. Interestingly, the 2 protist ODC sequences do not cluster together. The ODC of *L. donovani* forms a separate branch, and the ODC of *Trypanosoma brucei* segregates from the chordata branch of the ODCs, respectively, with bootstrap support of 99% and 52%. The surprising similarity of the ODC of *T. brucei* with the chordate ODCs has been noted previously (Salzman et al. 2000). The chordate ODCs cluster together into 1 branch except for the ODC2 from *Xenopus laevis*, which forms a branch together with the ODCp sequences (with 87% bootstrap support). This suggests that the ODC2 of *X. laevis* may be an ODCp, which has also been suggested by others (pitkanen et al. 2001). The AzI sequences from human and zebrafish segregate from the ODCps with a bootstrap value of 83%.

The presence of ADC activity in mammalian cells has been reported (Regunathan and Reis 2000; Horyn et al. 2005).
but the results have been questioned by others (Coleman et al. 2004). Moreover, human ODCp has been suggested to function as the mammalian ADC (Zhu et al. 2004). The phylogenetic tree presented here, however, does not support this hypothesis. In our tree, the human, mouse, and rat ODCp sequences are clearly clustered within the branch of chordate ODCs indicating that the ODCp sequences are very distantly related to the ADCs of plants and prokaryotes. When the human ODCp was first identified, it was proposed to function as a tissue-specific antizyme inhibitor (Pitkänen et al. 2001). This hypothesis is supported by our tree, where the ODCp sequences and the frog ODC2 sequence form 1 branch, which diverges from the branch containing the AzI sequences. Furthermore, the ODCp and AzI sequences segregate from the chordate ODCs, which suggests that the ODCp and AzI genes are a result of a gene duplication event that occurred early in the evolution of the chordate ODC sequences.

Comparison of the Topologies of Trees Produced by the NJ and ML Methods

Although the overall topologies of the NJ and the ML trees are identical, inside the main branches, a few minor changes can be observed in the tree topology. Similarly to the NJ tree, the ADC branch of the ML tree diverges into 2 branches but with only 52% bootstrap support (supplementary fig. 1, Supplementary Material online). In the ML tree, however, the plant and cyanobacterial ADCs are located within the same branch, with a bootstrap support of 72%, whereas in the NJ tree the cyanobacterial ADCs cluster together with other bacterial ADCs with 90% bootstrap support. Previous studies (Illingworth et al. 2003) support the topology of the ML tree where the plant ADCs are derived from the cyanobacterial ADCs because, in plants, ADC has been shown to be expressed in the chloroplasts (Borrell et al. 1995).
In comparison with the 4 DapDC branches found in the NJ tree, there are only 3 main branches containing the DapDC sequences in the ML tree (supplementary fig. 2, Supplementary Material online). The 4 epsilon-proteobacteria sequences, which formed a separate branch in the NJ tree, form a branch together with the alpha-, beta-, delta-, and gamma-proteobacteria in the ML tree. The DapDCs from archaea form 3 separate branches, and the DapDC of D. radiodurans segregates from a branch that contains the archaea and the beta- and gamma-proteobacteria DapDCs. In the ML tree, the ODC of T. brucei is separated from the human ODC branch; however, the bootstrap support for this node is 67% (supplementary fig. 3, Supplementary Material online).

The bootstrap values of the ML tree are lower compared with the bootstrap values in the NJ tree, with an average bootstrap support for the ML tree of 74% (calculated as an average across all of the nodes), whereas the average support for the NJ tree is 82%. Overall, there are 42 nodes having bootstrap support below 60% in the ML tree but only 25 in the NJ tree. Although it appears that the NJ method works better in the current case, producing trees with higher overall bootstrap confidence compared with the ML method, the ML method may be more accurate for nodes where the evolutionary time frame is relatively short, for example, the short internal branch lengths associated with the DapDC sequences and where the diversification of ADCs occurs.

Substrate Specificity and Active-Site Structure

The decarboxylases are highly specific for their corresponding substrates. The specificity is obtained through alterations at the region of the active site that recognizes the nonreactive part of the substrate. In contrast, the PLP-binding region of the active site accommodating the reactive part of the substrate is well conserved. An amino group of the substrate forms a Schiff base with PLP, and depending on the stereochemistry of the substrate, the carboxylate group of the substrate binds to either the si or the re face of PLP. The carboxylate group of the d-stereocenter in diaminopimelate may form hydrogen bonds with the side chains of His214A and Ser364B (Ray et al. 2002). In contrast, the carboxylate group of the l-stereocenter of ornithine binds to a hydrophobic and negatively charged pocket formed by the side chains of Phe397, Tyr389, Asp361, and the methylene groups of Lys69 (Jackson et al. 2003).

According to the available crystal structures of decarboxylases complexed with ligands, the nonreactive part of the substrate is accommodated next to a substructure (strand–loop–helix) formed by Met315–His333 (in human ODC) that seems to have a large impact on the active-site structure. Sequence variation within this substrate-defining substructure appears to be critical for substrate selection. Mutational studies, together with the crystal structures of DapDC and ODC in complex with ligands, have revealed that residues from this substructure participate in the binding of the nonreactive part of the corresponding substrate (Grishin et al. 1995; Ray et al. 2002). No experimentally determined structure of an ADC is currently available, but according to the structural modeling of Synechocystis ADC (Jantaro et al. 2006), 1 of the 3 aspartates suggested to bind the guanidium group of arginine comes from the same substructure-defining substructure. Ray et al. (2002) have suggested that the active sites of the decarboxylases have evolved to contain “molecular rulers” that measure the length of the substrate. Ornithine, the substrate of ODC is composed of a 4-carbon chain, whereas diaminopimelate contains a 5-carbon chain, and the substrate of ADC, arginine, is the longest molecule with a chain of 1 nitrogen and 5 carbon atoms. In ODC, the distance between the Asp332A from the substrate-defining substructure and the PLP is \( \sim7 \) Å, whereas the distance between the corresponding Tyr337A and PLP in DapDC is \( \sim10 \) Å. In the structural model of ADC, the distance between Asp521A from substrate-defining substructure and PLP is \( \sim13 \) Å. These lengths agree well with the lengths of the corresponding substrates and thus support the theory that the active site of the decarboxylases has evolved away to measure the length of the substrate, helping to ensure specificity.

In order to gain a deeper understanding of the relationship between the putative DC sequences and the DapDCs and ODCs, we examined the multiple-sequence alignment. Despite the close similarity between the ODCs and the putative DCs, according to the alignment, there are several sites where the sequences of the putative DCs are more similar to the DapDCs. For example, the residue on the C-terminal side of the PLP-binding lysine is a cysteine in ODC (Cys70) but an alanine in the putative DCs and DapDCs. In another region, a conserved tryptophan (Trp536 in human ODC) is substituted by glycine or alanine in the sequences of the putative DCs, whereas in the DapDCs, the corresponding residue is alanine, valine, or glycine. When this site is in the crystal structures of ODC and DapDC was compared, we noticed that the structural impact of this change is not as severe as one might have expected: In the crystal structure of Methanococcus jannaschii DapDC (Ray et al. 2002), the side chain of Trp321 from a parallel strand elsewhere in the fold occupies almost the same space as Trp356 in human ODC. Trp321 is not conserved in the eukaryotic ODCs, but in DapDCs and in the putative DCs, there is always an aromatic residue (tryptophan or tyrosine). This suggests that the last common ancestor from which the DapDCs and ODCs evolved may have had an alanine on the C-terminal side of the PLP-binding lysine and an aromatic residue equivalent to Trp321 in M. jannaschii DapDC.

One of the putative DCs, S. ruminantium, has previously been characterized as a lysine/ornithine decarboxylase (Takatsu et al. 2000). In order to look into the possible broader substrate specificity of the putative DCs, we analyzed the multiple-sequence alignment and studied the available crystal structures of the decarboxylases. Close to the active site of the eukaryotic ODCs, there is a conserved glycine (Gly201 in human ODC) (fig. 5A), which is substituted by a glutamine in the putative DC sequences. In the 2 known DapDC X-ray structures, this residue is also glutamine. According to the sequence alignment, glutamine is also conserved in the ADCs at this location but may be glutamine or glycine in other DapDCs. Significant differences in this region are observed when the crystal structures of DapDCs and ODCs are compared. The glutamine in DapDC interacts with the main chain of a nearby loop, affecting the 3-dimensional structure of the loop, whereas the corresponding interaction is not possible in ODC due to the presence of glycine instead of glutamine. Opposite to this residue, on the
other side of the active site, there is the substrate-defining substructure (Met315–His333) that contributes to the substrate specificity of the decarboxylases (fig. 5A). The residues in this substructure are well conserved within the eukaryotic ODCs but are clearly different in the putative DCs (fig. 5B). Taken together, the dissimilarities between the putative DCs and eukaryotic ODCs suggest that the 3-dimensional structure of the active site in the putative DCs differs from that of the eukaryotic ODC structures. The distinct active-site structure could enable the putative DCs to recognize and bind other substrates in addition to ornithine, for example, lysine in the case of the decarboxylase of \textit{S. ruminantium} (Takatsuka et al. 2000). Due to the large differences in the substrate-defining substructure of the putative DCs in comparison to the ODCs, we have not attempted to model the putative DCs because it is unlikely that a reliable model can be constructed for their active-site structure.

Structural Models of AzI and ODCp

The sequence alignment reveals that there are important differences between the residues that interact with PLP and the residues essential for the decarboxylase reaction in ODC, ODCp, and AzI. In order to study the effects of the substitutions on the protein structure, we have made homology models of the human ODCp and AzI dimers. In ODC, the phosphate of PLP is coordinated by several hydrogen bonds as part of the phosphate-binding cup (Denesyuk et al. 2003). The network is formed by the main-chain nitrogen of 3 residues, Gly237, Gly276, and Arg277 (numbering from human ODC), together with a conserved water molecule and the side-chain nitrogen (NH$_2$) of Arg277 and the side-chain oxygen (OH) of Tyr389 (fig. 6A) (Kern et al. 1999). Arg277 has been mutated to alanine and shown to reduce the PLP-binding affinity 270-fold in ODC (Osterman et al. 1992) and have no intrinsic enzymatic activity (Murakami et al. 1996).

In contrast to the AzIs, the residues that bind PLP in the ODCps are conserved, and the ODCps probably bind PLP in a manner similar to the ODCs. The ODCps, however, appear to be incapable of catalyzing the decarboxylation reaction because several key residues directly involved in catalysis are not conserved. Most notably Cys360, which is essential for the enzymatic reaction in ODCs, is substituted by a valine in human (Val361), rat, and mouse (Val360) ODCp. Previ-ously, it was shown that in ODC, the mutation of Cys360 to alanine or serine reduces the $k_{cat}$ value by 50–1,000-fold and changes the catalyzed reaction into a transamination reaction, which results in the formation of pyridoxamine-5-phosphate and gamma-aminobutyraldehyde (Jackson et al. 2000). The function of ODCp is not clear, but based on our modeling results and the phylogenetic analysis, it seems unlikely that ODCp accounts for the reported ADC activity present in mammalian cells but instead could function in transamination.

Conclusion

The existence of ODC-like bacterial and archaeal decarboxylases, called putative decarboxylases in this study, supports the theory that the reaction specificity of the PLP-dependent decarboxylases evolved before the substrate specificity. Thus, the ancestral decarboxylase is envisioned as having promiscuous substrate specificity, catalyzing the decarboxylation of arginine, diaminopimelate, lysine, and ornithine and possibly other amino acids too. The ancestral decarboxylase probably was a dimeric enzyme; however, from our results, it is not possible to conclude whether the monomer contained 2 or 3 domains. ADCs could have gained the ~85-amino acid insertion, or this insertion could have been lost from the other decarboxylases at the time of their divergence. The DapDCs and ODCs are more closely related to each other than to ADC, and as evidence of this, the putative decarboxylases have retained several features found in DapDCs that have subsequently been lost in the ODCs. Sequence and structure comparisons suggest that the putative decarboxylases may have retained the ability of the ancestral decarboxylase to recognize several substrates.

The rapid pace of identifying new protein sequences from genome projects is increasing; however, the annotation process is seriously lagging behind (Ouzonis and Karp 2003). With Blast searches, one can identify homologous sequences, but sequence matching on its own is often not sufficient to reliably identify the likely function of the protein coded by a newly sequenced gene. Phylogenetic analysis is an alternative method to classify protein sequences in cases where the studied protein shares sequence homology with a superfamilly made up of subfamilies spanning a wide range of similarity (Eisen 1998; Eisen and Fraser 2003). In this study, the decarboxylases from \textit{M. burtonii} and \textit{F. acidarmanus} were identified as DapDCs, whereas the other putative DCs clustered with the ODCs. Whether all the putative DCs are capable of using both lysine and ornithine as substrates, similarly to \textit{S. ruminantium} ODC, cannot be determined based on this study but should be experimentally verified. However, we consider phylogenetic analysis to be a suitable complementary method for annotating new sequences of yet unknown function.

Supplementary Material

Supplementary figures 1–3 and table 1 are available at \textit{Molecular Biology and Evolution} online (http://www.mbe.oxfordjournals.org/).
**Note Added in Proof**

A recently published paper reports that the mouse ODCp has antizyme inhibitor activity and no ornithine or arginine decarboxylating activity (Lopez-Contreras et al. 2006), which is consistent with our results indicating that ODCp cannot function as a decarboxylase.

**Acknowledgments**

We thank M.Sc. Mikko Huhtala for help with the PHYLIP package. This work was supported by the Academy of Finland, the Technology Development Center of Finland, National Graduate School in Informational and Structural Biology, The Sigrid Juselius Foundation, and the ÅA Center of Excellence in Cell Stress.

**Literature Cited**


**Felsenstein J. 1989. PHYLIP—phylogeny inference package (version 3.2). Cladistics. 5:164–166.**


Michele Vendruscolo, Associate Editor

Accepted September 18, 2006