Journal of Medical Genetics and Genomics Vol. 2(3), pp. 67-80, December 2010
Available online http://www.academicjournals.org/jmgg
ISSN 2141-2278 ©2010 Academic Journals

Full Length Research Paper

In silico analysis of the two forms of lysine decarboxylase

Gokulnath Krithika and Namasivayam Gautham*

Centre of Advanced Study in Crystallography and Biophysics, University of Madras, Guindy campus, Chennai- 600 025, India.

Accepted 18 July, 2010

Amino acid decarboxylases are expressed in response to acidic pH in bacteria. Among amino acid decarboxylases, lysine decarboxylase is expressed to levels constituting 2% of the total cell protein during acid stress. In bacteria, two forms of lysine decarboxylase exist, namely the inducible CadA and the constitutively expressed Ldc. The two proteins catalyse the conversion of lysine to cadaverine as an acid tolerance response. Here, we report the *in silico* analysis of these two sequences in an attempt to identify the differences between the two proteins and the reason for the existence of the two variants. The analysis showed that the two proteins could be differentiated on the basis of the amino acid composition, and the organisation of sequence motifs. The two forms of lysine decarboxylase were found to segregate into different branches in phylogenetic analysis. Sequence comparison with other amino acid decarboxylases showed that most of the motifs are conserved among pyridoxal phosphate binding amino acid decarboxylases.

Key words: Lysine decarboxylase, inducible form(cadA),constitutive form(Ldc),acid tolerance, motif analysis, phylogenetic analysis, virulence, conserved sequence motifs, pyridoxal phosphate binding enzyme, amino acid decarboxylase, gamma proteobacteria, PCA, cluster analysis, functional annotation.

INTRODUCTION

Pathogenic bacteria face a number of pH changes as they enter the human body. Though the mechanisms that operate in sensing and responding to the external pH have not been completely understood, a number of genes that are expressed in response to acidic pH have been identified [Slonczewski et al., 1987]. These genes are gaining a lot of importance in the development of drugs since acid tolerance is often a crucial factor for successful pathogenesis (Finlay and Falkow, 1997; Vazquez-Juarez et al., 2008).

Among the many genes induced at acidic pH, amino acid decarboxylases were identified the earliest (Gale and Epps, 1942). As their names suggest, these enzymes catalyse the decarboxylation of their associated amino acids, and the products of the reactions serve to neutralize the low pH. Most of these enzymes utilize pyridoxal phosphate as the cofactor. They all catalyse the same type of reaction, and differ only in their substrate specificity (Momany et al., 1995a). It is therefore interesting that there is a high degree of diversity in the sequences of these proteins [John, 1995].

Lysine decarboxylase catalyses the conversion of lysine to cadaverine. This polyamine is then pumped out of the cell and helps in raising the external pH (Gale, 1940; Meng and Bennet, 1992a). Lysine decarboxylase has multiple biological roles. The most significant role of lysine decarboxylase is its involvement in acid tolerance response. This feature is especially important in pathogenic bacteria. Lysine decarboxylase is known to be involved in pathogenecity (Maurelli et al., 1998; Day et al., 2001), and is necessary for the virulence of bacteria such as *Vibrio cholerae* (Merrell and Camilli, 1999) and *Salmonella typhimurium* (Lin et al., 1995; Portillo et al., 1993).

Two forms of amino acid decarboxylases exist. One is constitutively expressed and the other is induced to high levels of expression at acidic pH (Sabo et al., 1974a). The

*Corresponding author. E-mail: n_gautham@hotmail.com. Fax: +91 44 22352494.

inducible form of lysine decarboxylase is encoded by *cadA*. This has been extensively studied. CadA is the enzyme related to virulence and pathogenecity of bacteria [Lin et al., 1995; Portillo et al., 1993; Merrell and Camilli, 1999]. The constitutive form of the enzyme was discovered by constructing mutants that blocked all known polyamine synthesis (Tabor et al., 1980; Goldemberg, 1980). This form has not received much attention so far. The gene that codes for the constitutive form, *ldc* is located far away from *cadA* on the *E. coli* chromosome (Goldemberg, 1980; Lemmonier and Lane, 1998).

Meng and Bennet (1992a) studied the sequences of the three amino acid decarboxylases and showed that they are different from each other. Constitutive arginine decarboxylase does not share any homology with constitutive ornithine decarboxylase or inducible lysine decarboxylase. The amino acid sequence of constitutive ornithine decarboxylase, however, bears a strong resemblance to that of inducible lysine decarboxylase and provides further evidence that constitutive ornithine decarboxylase and inducible arginine, lysine and ornithine decarboxylases may share a common origin (Meng and Bennet, 1992a).

Molecular mechanisms behind expression of *CadA* have been extensively studied. However, studies on the basic differences between the two forms of lysine decarboxylase based on sequence analysis have not been undertaken. Here, we use the tools of bioinformatics to explore the differences between the two forms of lysine decarboxylase.

Amino acid decarboxylases are pyridoxal phosphate binding enzymes and have many common features in terms of sequence motifs and functions. Therefore sequences of lysine, ornithine and arginine decarboxylase of plants and bacteria were compared. This was in an attempt to find conserved motifs among these enzymes and to understand their phylogeny.

METHODS

Primary amino acid sequences for lysine decarboxylases of different origin were obtained by a keyword search from SWISS-PROT and TrEMBL databases [http://www.expasy.org/sprot/]. The preliminary search yielded a total of 241 sequences. Of these 8 were from SWISS-PROT and 233 were from TrEMBL. One sequence from SWISS-PROT and 186 sequences from TrEMBL were omitted as they were classified as one of the following: putative sequences, fragments, truncated proteins, and sequences that were not annotated specifically as lysine decarboxylase. The final dataset used for the present study has 54 sequences (Table.1).

Multiple sequence alignment was carried out using MultAlin [Corpet, 1988] and ClustalX1.81 [Thompson et al., 1994]. A search for homologous motifs was carried out in the Supercomputer Center, San Diego, USA, on the Multiple Em for Motif Elicitation website, [http://meme.sdsc.edu/meme/website/meme.html]. The motifs were identified using the program 'Discover' in the MEME website [Bailey, 1998]. This allows the user to specify a number of parameters. In our study the minimum width of the motif was set as 3, and maximum size was 15. A maximum of 20 motifs were requested. For the other parameters the default options were chosen.

The evolutionary distances between the sequences were computed using the program in the software package PHYLIP 3.6a2 [Felsenstein, 1996]. Phylogenetic trees were designed in accordance with the maximal probability, maximal parsimony, and long-distance methods of PHYLIP 3.6a2 [Felsenstein, 1996] (PROTPARS, PROTDIST, NEIGHBOR). For each tree, 100 bootstrapped alignments were generated using the SEQBOOT program, and the consensus trees were constructed using the CONSENSE program.

The amino acid composition of the dataset containing 54 sequences of lysine decarboxylase was calculated and subjected to Principal Component Analysis (PCA) [Davis, 1986]. PCA has many applications [Jolliffe, 2002]. In the present study it has been used for simple reduction of the data, consisting of the 20 (amino acids) X 54 (sequences) (or 20 (amino acids) X 18 (sequences)) matrix of values of the composition, to its most important components.

The PCA routine in PAST [http://folk.uio.no/ohammer/past/index.html] finds the eigenvalues and eigenvectors of the variance-covariance matrix or the correlation matrix. The program also gives the percentage of variance accounted for by these components. In a successful PCA, the first two components usually represent a considerable percentage of the variance. Functional classification was carried out using the web-based program, VICM Pred [http://www.imtech.res.in/raghava/vicmpred/index.html] [Saha and Raghava, 2006]. For functional annotation, the sequences of enterobacteria, a sub-class of gamma proteobacteria alone were considered. This is because it is in these bacteria that the two forms of the enzyme are clearly annotated as belonging to either the constitutive or the inducible form. The analysis was carried out to find if the two forms of lysine decarboxylase fall into different functional classes.

Secondary structure prediction and alignment was carried out using FUGUE. This is a program that compares a given sequence to a structure, utilizing environment-specific substitution tables and structure dependant gap penalties. The secondary structure of the sequence is then predicted after calculating the sequence-structure compatibility scores [Shi et al., 1993].

RESULTS AND DISCUSSION

The sequences of lysine decarboxylase from 54 organisms obtained from SWISS-PROT are tabulated in Table 1. The bacteria have been arranged according to the classification in the KEGG database. The deduced aminoacid sequences that have been annotated as belonging to either the inducible or the constitutive form are indicated. Though some bacteria were not annotated, two sets of sequences were taken wherever possible, to see if such annotation was possible through the present analysis. There is an obvious difference in the lengths of sequences. Proteins that belong to the class Proteobacteria (except *Legionella pneumophila subsp. pneumophila* and *Brucella melitensis*) have longer sequences compared to the other classes. The average length of the longer Proteobacterial sequences is 723. The other sequences belonging to the classes Firmicutes, Lactobacillales, Clostridia, Actinobacteria, Cyanobacteria and Bacteroides have an average length of 452. Most of the lysine decarboxylase sequences in

Table 1. The data set used for the analyses.

Bacterial Class (KEGG)	SwissProt /TrEMBL id	Organism	Mode of expression	Label	Sequence length
Proteobacteria/Gamma/Enterobacteria	P52095	*Escherichia coli*	Constitutive	1DCLZECOL	713
	P0A9H4	*Escherichia coli*	Inducible	2LDCIECO57	715
	Q0T9R0_ECOL5	*Escherichia coli (UPEC)*	Constitutive	3IECOL5	715
	Q0TLE7_ECOL5	*Escherichia coli (UPEC)*	Inducible	4IECOL5	712
	P0A1Z1	*Salmonella typhi*	Inducible	5LDCISALTI	714
	P0A1Z0	*Salmonella typhimurium*	Inducible	6LDCISALTY	714
	Q8Z998_SALTI	*Salmonella typhi*	Constitutive	7SALTI	713
	Q8ZRN7_SALTY	*Salmonella typhimurium*	Constitutive	8CSALTY	713
	Q57LF2_SALCH	*Salmonella choleraesuis*	Constitutive	9SALCH	714
	Q57T21_SALCH	*Salmonella choleraesuis*	Inducible	10SALCH	713
	Q5PD78_SALPA	*Salmonella paratyphi-a*	Inducible	11SALPA	713
	Q5PIH8_SALPA	*Salmonella paratyphi-a*	Constitutive	12SALPA	714
	Q3Z5H1_SHISS	*Shigella sonnei*	Constitutive	13SHISS	713
	Q83SL1_SHIFL	*Shigella flexneri*	Constitutive	14SHIFL	713
	P05033	*Hafnia alvei*	Not annotated	40DCLYHAFA	739
Proteobacteria/Gamma	Q5E344_VIBF1	*Vibrio fischeri*	Not annotated	15VIBF1	712
	Q76KS0_VIBPA	*Vibrio parahaemolyticus*	Not annotated	16VIBPA	724
	Q7MIY2_VIBVY	*Vibrio vulnificus*	Not annotated	17VIBVY	711
	Q9KV75_VIBCH	*Vibrio cholerae*	Inducible	18VIBCH	733
	Q5ZSL8_LEGPH	*Legionella pneumophila* sub sp. *neumophila*	Not annotated	42LEGPH	190
	Q5NHP1_FRATT	*Francisella tularensis*	Not annotated	36FRATT	713
	Q9FAD0_MORJA	*Moritella japonica*	Not annotated	43MORJA	712
Proteobacteria/Beta	Q399G8_BURS3	*Burkholderia sp*	Not annotated	26BURS3	779
	Q39DV9_BURS3	*Burkholderia sp*	Not annotated	27BURS3	759
	Q9Z4R7_EIKCO	*Eikenella corrodens*	Not annotated	34EIKCO	709
	Q474W2_RALEJ	*Ralstonia eutropha*	Not annotated	46RALEJ	756
	Q3SH30_THIDA	Thiobacillus denitrificans	Not annotated	53THIDA	747
Proteobacteria/alpha/ rhizobacteria	Q8YFM5_BRUME	*Brucella melitensis*	Not annotated	28BRUME	221
Firmicutes/Bacillales	Q4MTE9_BACCE	*Bacillus cereus*	Not annotated	20BACCE	493
	Q6HEK0_BACHK	*Bacillus thuringiensis*	Not annotated	21BACHK	335
	Q81W12_BACAN	*Bacillus anthracis*	Not annotated	22BACAN	473
	Q9K9A3_BACHD	*Bacillus halodurans*	Not annotated	23BACHD	190
	Q9KGM0_BACHD	*Bacillus halodurans*	Not annotated	24BACHD	482
	Q8EU54_OCEIH	*Oceanobacillus iheyensis*	Not annotated	44OCEIH	472
	Q5L130_GEOKA	*Geobacillus kaustophilus*	Not annotated	37GEOKA	490
	Q5L3X0_GEOKA	*Geobacillus kaustophilus*	Not annotated	38GEOKA	477
	Q8CMS4_STAES	*Staphylococcus epidermidis*	Not annotated	49STAES	445
	Q7DFY1_SELRU	*Selenomonas ruminantium*	Not annotated	47SELRU	473
	O50657	*Selenomonas ruminantium*	Not annotated	48DCLOSELR	393
	Q3AG13_CARHZ	*Carboxydothermus hydrogenoformans*	Not annotated	35CARHZ	442
Lactobacillales	Q8DQ73_STRR6	*Streptococcus pneumoniae*	Not annotated	50STRR6	491
Clostridia	Q97GM8_CLOAB	*Clostridium acetobutylicum*	Not annotated	30CLOAB	487

Table 1. Contd.

	Q97MA1_CLOAB	*Clostridium acetobutylicum*	Not annotated	31CLOAB	481
	Q8XHN0_CLOPE	*Clostridium perfringens*	Not annotated	32CLOPE	476
	Q8XMY9_CLOPE	*Clostridium perfringens*	Not annotated	33CLOPE	484
	Q2ZE76_CALSA	*Caldicellulosiruptor saccharolyticus*	Not annotated	29CALSA	461
Actinobacteria	Q414S2_KINRA	*Kineococcus radiotolerans*	Not annotated	41KINRA	486
Cyanobacteria	Q8YMP6_ANASP	*Anabaena sp. (strain PCC 7120)*	Not annotated	19ANASP	488
	Q7NFN7_GLOVI	*Gloeobacter violaceus*	Not annotated	39GLOVI	467
	Q7VBI0_PROMA	*Prochlorococcus marinus*	Not annotated	45PROMA	440
	Q5N3V7_SYNP6	*Synechococcus sp*	Not annotated	51SYNP6	489
	P72774_SYNY3	*Synechocystis sp.*	Not annotated	52SYNY3	483
Bacteroides	Q650A4_BACFR	*Bacteroides fragilis*	Not annotated	25BACFR	176
	Q8A2J8_BACTN	*Bacteroides thetaiotaomicron*	Not annotated	25BACTN	161

enterobacteria (a sub-class of gamma proteobacteria) have been annotated as belonging to either the inducible or constitutive form. The sequences of the two forms are of approximately the same length, with average values of 716 and 714 residues for the constitutive and inducible forms, respectively.

Sequence alignment using BLAST (Altshul, 1990) shows that the sequences of the constitutive *E. coli* lysine decarboxylase and the inducible form share 69% identity with an E value of 0. Thus, the two forms of lysine decarboxylase, which perform essentially the same function, and have almost the same sequence length, are only 69% identical and not higher.

The genes are located far away from each other in the *E. coli* chromosome. The *cadA* gene that codes the inducible lysine decarboxylase is mapped at 93.7 min and the *ldcC* gene that codes the constitutively expressed lysine decarboxylase is mapped at 4.7 min on the *E. coli* chromosome (Shi and Bennet, 1995).

The sequences of lysine decarboxylase were submitted to the MEME server for motif analysis (Bailey, 1998). The analysis was carried out to check if the two forms of the enzymes have identifiable differences in the motifs or their occurrence. A total of 20 different motifs were identified in 48 of the 54 sequences (Table 2).

Some of the motifs, which have been shown to have functional significance (highlighted in bold face in Table 2) in case of ornithine decarboxylase (Momany et al., 1995b), and those which help in differentiating between the two forms of the enzyme, are shown in Figure 1. The figure shows that the pattern in which the motifs are organised is different in the two forms of the enzyme. For example, in the constitutive form two copies of motif 14 occur alternating with motif 10. There are also two copies of motifs 9, 7 and 11. The inducible form has only one copy each of motifs14, 9, 7 and 11.

The differences in sequence and motif organisation between the two forms of lysine decarboxylase prompted us to analyse their phylogeny. The results are shown in Figure 2 in the form of a dendrogram. As mentioned earlier, the sequences from enterobacteria have been annotated as belonging to either constitutive or inducible form. The figure shows that these sequences branch into two separate groups. The sequences of the two forms of lysine decarboxylase are distantly related. The constitutive form of the enzyme from *Shigella* and *E. coli* is the most distant relative of all lysine decarboxylase sequences.

Table 3 gives the average values and standard deviations of the amino acid compositions of the 20 amino acids in the 18 sequences of constitutive and inducible forms of lysine decarboxylase from gamma proteobacteria, the class for which most of the sequences are annotated. A principal component analysis was carried out to determine the particular amino acids that could perhaps distinguish the sequences. Table 4 gives the eigenvectors and eigenvalues of the first three principal components. These are sufficient to capture 84% of the variance in the data. The first two components alone account for 70%. The largest contributions to these two components come from the following amino acids – Q, L, A and I, indicating that the compositions of these amino acids present the largest variation in the data set. The assortment of the two forms of the protein into two clusters urged us to find out the amino acids that contribute the most to such difference between them. For this, the sequences of the class gamma proteobacteria alone were taken. It is clear from Figure 3 that the two forms of lysine decarboxylase separate into two clusters. The plot represents 70.5% of the variance in the data.

Table 2. Motifs identified from the 54 sequences of lysine decarboxylase.

Motif number	Length of motif	Sequence of motif
1	10	DRNCHKSLTH
2	10	EAFNQASMMH
3	10	KYTFCTPGHM
4	10	ICPYPPGVPL
5	10	GSVVDHTGPY
6	10	TSPSYPIVAS
7	10	IHFDSAWVPY
8	10	DGWFFDVWQP
9	10	YETQSTHKML
10	10	NSTYDGLCYN
11	10	YPGFETDIHG
12	8	LLMLCSIG
13	10	FYKNMRIQEL
14	10	WDQYSLDLCS
15	10	PFTKALFKYV
16	10	PTRNAYGILG
17	10	TNFHPIYQGK
18	10	FYDFFGPNTM
19	10	EAEEYIARTF
20	10	GTAFQKSPVG

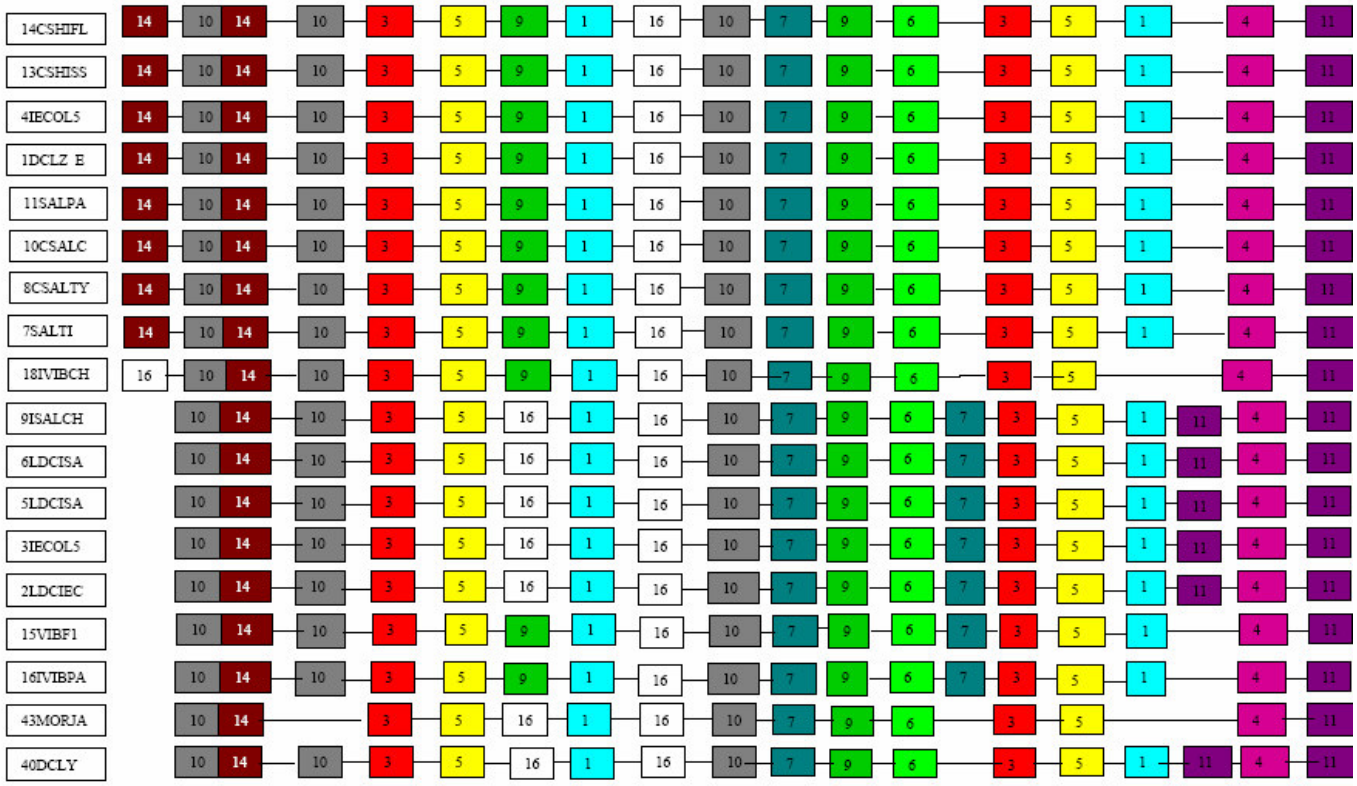

Figure 1. Motifs identified by the MEME server are listed. Only selected motifs are shown. The alignment has been adjusted to compare occurrence pattern of motifs and do not reflect actual sequence lengths. The constitutive enzymes are labelled as C (following their number eg.14CSHIFL) and the inducible forms have been labelled as I (eg. 9ISALCH).

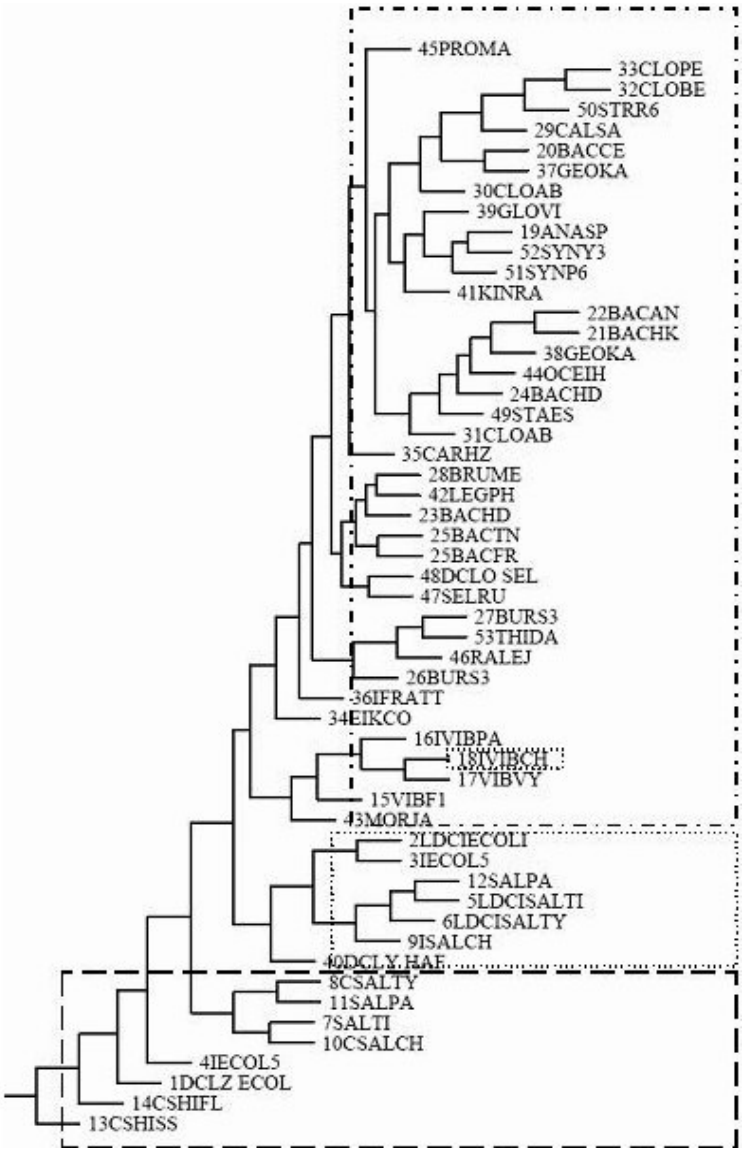

Figure 2. The dendrogram of 54 sequences of lysine decarboxylase. The branches within the dotted line represent the sequences annotated as inducible. The sequences within the dashed line have been annotated as constitutive. The branches within the dashed-dotted line are those that have not been annotated.13CSHISS (constitutive enzyme from *Shigella sonnei)*, 14CSHIFL (constitutive enzyme from *Shigella flexneri*) and the constitutive enzymes from *E. coli* (1DCLZECOL and 4ECOL5) are the most distant relatives of all the lysine decarboxylases.

The first component contributes 44.1% and the second, 26.4%. The contribution of each of the amino acids to the first component is represented in Figure 4a. Q and L contribute the largest magnitudes to this component. In order to identify the precise set of amino acids that contributes to this difference in the clustering behaviour, those that had a coefficient of < 0.2 in the first two principal component vectors were deleted, and the vectors corresponding to the other amino acids alone (i.e. Q, L, A, I, K, E and S) were plotted (Figure 4b). Such deletion did not affect the plot, and that the amino acids were able to represent the same clustering pattern as obtained when all the 20 amino acids were plotted. After progressively deleting each of these amino acids Q, L, A and I alone were found sufficient to represent the clustering pattern (Figure 4c). The amino acids that contribute maximally to the difference between the two forms are thus alanine, isoleucine, leucine and glutamine. The PCA was also carried out on all 54 sequences given in Table 1. Figure 5 is a scatter plot of the 54 sequences

Table 3. Difference in amino acid composition of constitutive and inducible gamma proteobacteria (standard deviation in parentheses). Those that contribute maximally to the overall difference are highlighted in bold.

Amino acid	Average composition (%) in the inducible lysine decarboxylase in gamma proteobacteria	Average composition (%) in the constitutive lysine decarboxylase in gamma proteobacteria
P	6.5 (0.6)	6.8 (0.2)
A	6.8 (0.4)	7.0 (0.1)
G	4.2 (0.6)	4.6 (0.1)
R	5.1 (0.2)	5.8 (0.2)
W	1.0 (0.2)	1.7 (0.0)
V	5.4 (0.7)	5.6 (0.2)
S	5.6 (0.6)	5.2 (0.1)
T	6.2 (0.7)	6.0 (0.0)
H	3.1 (0.4)	3.4 (0.2)
D	5.4 (0.5)	5.8 (0.2)
E	7.2 (0.8)	6.7 (0.1)
L	8.7 (0.3)	9.7 (0.2)
Q	2.6 (1.0)	3.8 (0.1)
C	1.1 (0.3)	1.1 (0.1)
M	4.9 (0.5)	4.6 (0.0)
F	4.7 (0.4)	3.9 (0.1)
Y	4.1 (0.7)	3.9 (0.1)
N	6.6 (0.5)	6.4 (0.1)
K	4.6 (0.9)	3.7 (0.1)
I	6.0 (0.5)	4.3 (0.2)

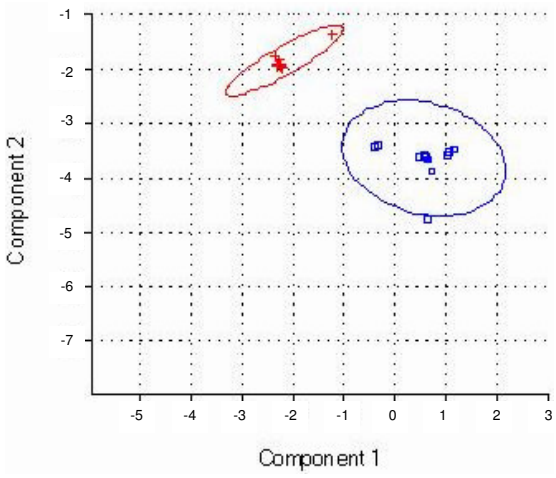

Figure 3. Cluster analysis based on PCA. The sequences of gamma proteobacteria separate into two distinct groups - the inducible and constitutive forms respectively. The (+) represent the inducible form and the (□) represent the constitutive forms.

along the first two principal components. This represents 57.1% of the variance in the data. The first component contributes 43.6% and the second 13.5%. The green crosses represent the sequences that were not annotated. The red crosses represent the inducible form of lysine decarboxylase from gamma proteobacteria and the blue squares represent the constitutive enzyme from the same taxonomic class. It is clear from this plot that the bacteria do not show any separation into distinct clusters. None of the unannotated sequences can be identified from this diagram as either inducible or constitutive. The sequences of the constitutive and

Table 4. Variance for the first three principal components and the coefficients for each amino acid. The amino acids in bold show maximum variance when the first two components are considered.

% Variance	PC 1	PC 2	PC 3
	44	26	14
Amino acid	**Coefficients of the first three principal components**		
P	-0.17	0.00	0.12
A	0.22	0.51	0.11
G	-0.14	0.17	-0.40
R	-0.12	0.04	0.04
W	-0.09	0.19	-0.02
V	0.01	0.21	0.14
S	0.22	-0.07	0.23
T	0.13	0.02	-0.17
H	-0.08	0.13	-0.24
D	-0.14	-0.01	0.40
E	0.13	-0.22	-0.01
L	-0.54	-0.34	-0.19
Q	-0.47	0.12	0.22
C	0.06	0.16	0.12
M	0.10	-0.01	0.33
F	0.13	-0.20	-0.15
Y	0.05	-0.16	-0.06
N	0.02	-0.07	-0.37
K	0.38	0.10	-0.30
I	0.29	-0.56	0.19

inducible forms of lysine decarboxylase were submitted to the VICM Pred server [http://www.imtech.res.in/raghava/vicmpred/index.html]. The sequences of gamma proteobacteria alone were used for this purpose, since it is in these organisms that the annotation into constitutive and inducible forms is clearly defined. In all the cases, the server identified the inducible CadA as a virulence factor, whereas the constitutive enzyme Ldc was identified as a metabolism molecule (Table 5). In each case, the score for the other possible functions was significantly lower than for the one identified, indicating that the protein is unlikely to perform other functions according to the predictions of this server.

The G+C content of the two forms of lysine decarboxylase of gut bacteria was calculated. The constitutive form had G+C content 0.50 and the content in the inducible form was 0.45. The statistical significance of such difference in the G+C content was calculated using two-tail paired T test. The P value for the significance of the difference is 0.04, showing that the two proteins were indeed different from each other.

Since lysine decarboxylase is a pyridoxal phosphate binding amino acid decarboxylase (Momany et al., 1995), it is possible that the functional significance of some of the motifs in its sequence may be understood by comparing it with other amino acid decarboxylases with which it shares similarity. As already mentioned above, the sequences of the *E. coli* lysine decarboxylase share about 34% identity with inducible arginine decarboxylase and constitutive ornithine decarboxylase. Lysine decarboxylase is very similar to ornithine decarboxylase in the manner in which the functional domains are organised. The major difference between the two proteins lies in the N-terminal domain.

The dataset in Table 1 was expanded to include sequences of arginine decarboxylase and ornithine decarboxylase (Table 6). The sequences of these two enzymes from bacterial and plant sources were retrieved from SWISS-PROT using a keyword search, as before. Of the bacterial enzymes, those which had two entries clearly annotated as constitutive and inducible were chosen for the present study. Also chosen were sequences of *Vibrio cholerae*, *Lactobacillus sp*. and *Salmonella paratyphi,* which had only one entry in SWISS-PROT. *Vibrio cholerae* and *Salmonella paratyphi* are pathogenic bacteria. The three-dimensional structure of ornithine decarboxylase from *Lactobacillus* is known (PDB ID -1ORD).

The sequences of lysine decarboxylase from gamma proteobacteria, given in Table 1 were aligned with sequences of bacterial arginine and ornithine decarboxylase in Table 6 using the multiple sequence alignment tools, ClustalX (Thompson et al., 1994) and MultAlin (Corpet, 1988). The alignments show that regions that are highly conserved among the other amino acid decarboxylases are present also in lysine decarboxylase. However, as shown in Figure 6, a stretch of about 16 residues (residues from 333 to 348) was strikingly different between the two forms, constitutive and inducible, of the enzyme in the all three amino acid decarboxylases. A BLAST search with each subsequence specifically picked up only that form and type of enzyme from which it was taken. For example the sequence “NEEYLRQQIRDVAPEK” picks up only constitutive ornithine decarboxylases from various organisms. These sequences therefore differentiate the amino acid decarboxylases from one another, as well as the two forms, inducible or constitutive.

The sequences of *E. coli* inducible lysine decarboxylase (CadA) and constitutive lysine decarboxylase (Ldc) were compared with the sequence of 1ORD. The *E. coli* constitutive form shares 29.3% identity and the inducible form shares 28.2% identity with the sequence of this protein. The crystal structure of ornithine decarboxylase is known. Figure 7 (obtained using the FUGUE server http://www-cryst.bioc.cam.ac.uk/fugue/) shows the possible common secondary structures in the sequences. The secondary structures of the sequences are very similar. There are two active sites per dimer in case of ornithine decarboxylase [John, 1995]. The active site residues, (residue numbers 1130 to 1140, Figure 7) are conserved in the all amino acid decarboxylases.

The other functionally important motifs of ornithine

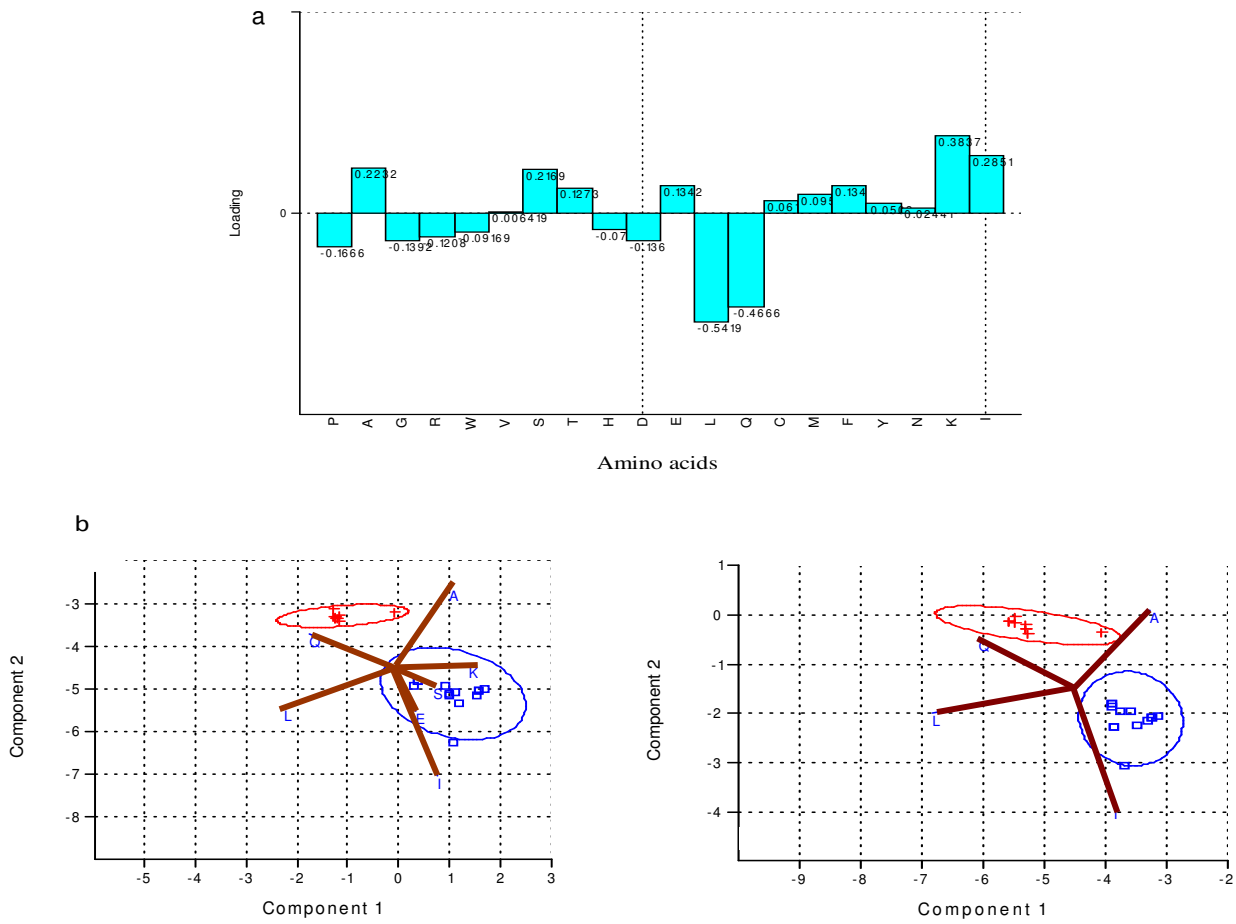

Figure 4. The contributions of the amino acids to the first principal component.

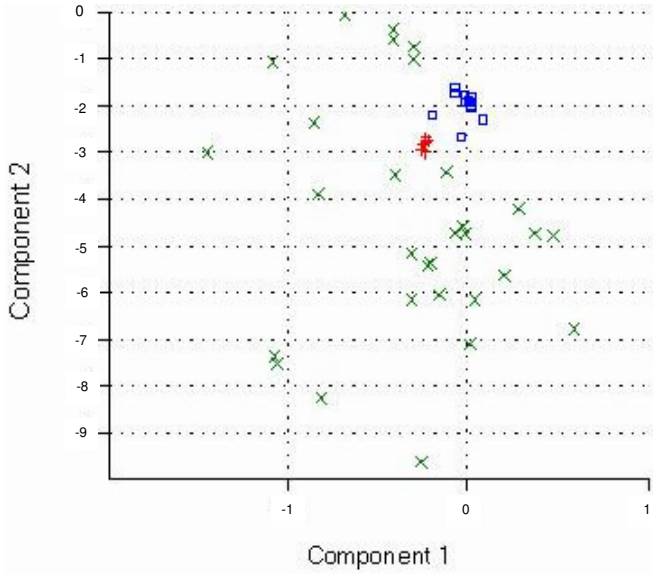

Figure 5. Scatter plot of the 54 sequences plotted along the first two principal components. X represent unannotated sequences. + represent the inducible forms and □ represent the constitutive forms.

Table 5. Scores for functional classes obtained from VICM Pred. (*). The protein is assigned to the class that returns the maximum score.

lysine decarboxylase sequence	Scores in each functional class*			
	Cellular processes	Information	Metabolism	Virulence
P0A9H4 - *E. coli* -Inducible	-0.119	-2.67	0.60	1.159
P52095 - *E. coli* - Constitutive	0.059	-5.27	0.94	0.79
Q0T9R0_ECOL5 - *E. coli* (UPEC) - Inducible	-0.31	-3.09	1.03	1.08
Q0TLE7_ECOL5 - *E. coli* (UPEC) - Constitutive	-0.06	-5.45	1.27	0.66
P0A1Z1 - *Salmonella typhi* - Inducible	-0.13	-4.82	0.55	1.16
Q8Z998_SALTI - *S.typhi* - Constitutive	-0.43	-6.68	1.79	-0.17
P0A1Z0- *Salmonella typhimurium* - Inducible	-0.13	-4.82	0.55	1.16
Q8ZRN7_SALTY- *S. typhimurium* - Constitutive	-0.43	-6.68	1.79	-0.17
Q57LF2_SALCH - *Salmonella choleroesius* - Inducible	0.168	-4.18	-0.008	1.04
Q57T21_SALCH -*S. choleroesius* - Constitutive	-0.15	-6.33	1.58	0.22
Q3Z5H1_SHISS - *Shigella sonnei* - Constitutive	0.156	-5.20	1.02	0.75
Q83SL1_SHIFL - *Shigella flexneri* – Constitutive	-0.34	-5.90	0.81	0.77

Table 6. The sequences of arginine and ornithine decarboxylase used along with table 2.1 for phylogenetic analysis. The 3667337* and 3668731* are NCBI-Gene numbers.

Amino acid decarboxylases	Taxonomic class	Swiss-prot id	Mode of expression	Organism	Label
Arginine decarboxylase	Bacteria/ Proteobacteria/ Gamma	P21170	Inducible	*Escherichia coli.*	1ADIA_ECOLI
		P28629	Constitutive	*Escherichia coli.*	2SPEA_ECOLI
		P60658	Inducible	*Salmonella typhi*	7SPEA_SALTI
		Q8Z1P1	Constitutive	*Salmonella typhi*	8ArgSALTI
		P60658	Inducible	*Salmonella typhimurium*	9SPEA_SALTY
		Q8Z1P1	Constitutive	*Salmonella typhimurium*	10argSALTY
		Q7CH88	Constitutive	*Yersinis pestis*	11agrconYE
		Q8ZHG8	Constitutive	*Yersinis pestis*	12argconYE
		Q83Q93	Constitutive	*Shigella flexneri*	13argconSHIF
		Q83PA4	Inducible	*Shigella flexneri*	14argSHIFLind
		3667337*	Constitutive	*Shigella sonnei*	15argconSHIS
		3668731*	Inducible	*Shigella sonnei*	16argindSHISS
		A0A546	Constitutive	*Salmonella paratyphi*	18argSALPA
		Q9KLD1	Constitutive	*Vibrio cholerae.*	17SPEA_VIBCH
	Cyanobacteria	P72587	Inducible	*Synechocystis sp*	5SPEA1_SYNY3
		P74576	Constitutive	*Synechocystis sp*	6SPEA2_SYNY3
	Plants/Dicotyledons/Mustard family	Q9SI64	Inducible	*Arabidopsis thaliana*	3SPE1_ARATH
		O23141	Constitutive	*Arabidopsis thaliana*	4SPE2_ARATH
Ornithine decarboxylase	Bacteria/ proteobacteria/ gamma	P21169ECOLI	Constitutive	*Escherichia coli*	19OconECOLI
		P24169_ECOLI	Inducible	*Escherichia coli*	20OindECOLI
		Q8ZQW6	Constitutive	*Salmonella typhimurium*	21OconsSALTY
		Q8ZM37	Inducible	*Salmonella typhimurium*	22OindSALTY
		Q83Q82	Constitutive	*Shigella flexneri*	23ornSHIFL
		Q9KKN9	Constitutive	*Vibrio cholerae*	24ornVIBCH
	Bacteria/firmicutes/ lactobacillales	Q673H3	Inducible	*Lactobacillus acidophilus*	25ornLACAC

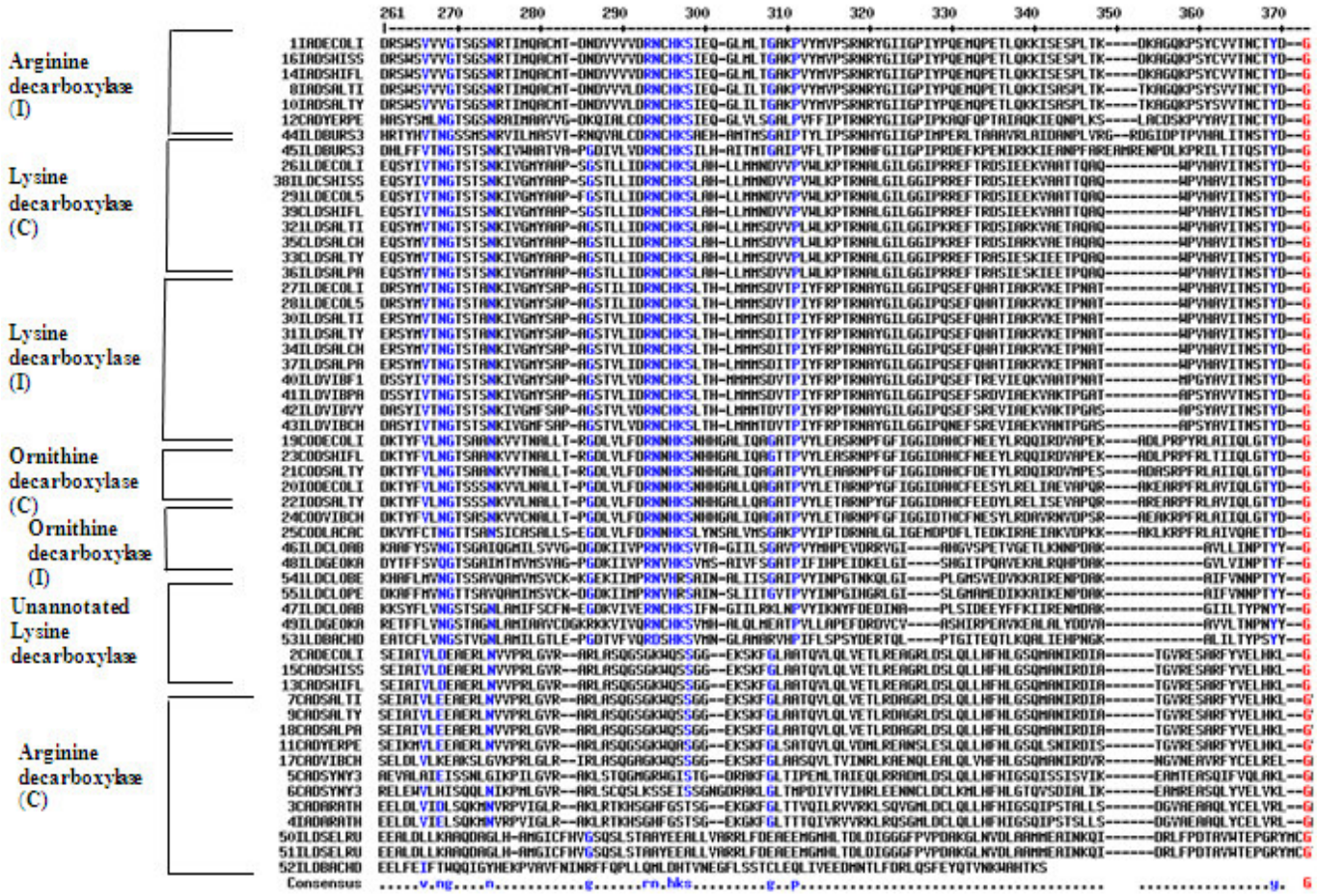

Figure 6. Multiple sequence alignment of amino acid decarboxylases. The sequence between 330-345 shows sequences identified to be highly specific for the form (inducible and constitutive form) and type of the enzyme (based on substrate specificity) (C) represents constitutive form and (I) represents inducible form.

Table 7. The motifs identified in lysine decarboxylase which are highly conserved among amino acid decarboxylases and their functions in ornithine decarboxylase.

Signature	Functional or structural role in Ornithine decarboxylase (1ORD)	Functional residue and its position in Ornithine decarboxylase (1ORD)	Residue position in *E.coli* inducible lysine decarboxylase
(SXHK)	Schiff's base with pyridoxal –P	Lys355	Lys387
	H-bond to OP, possible proton donor in decarboxylases	His354	His 386
(FDSAW)	H-bond to OH of pyridoxal –P	Trp319	Trp351
	Salt bridge /H-bond to N1 of pyridoxal –P	Asp316	Asp348
(DRNCHKS)	Bend, H-bond to Asp316 (/Asp348)	Asn221	Asn254
	Cofactor binding, possible proton donor	His223	His256
(GTSTSNK)	Positive dipole formation during catalysis	Ser198	Ser 221
(GIXXEK)	Substrate binding	Glu532	Glu526

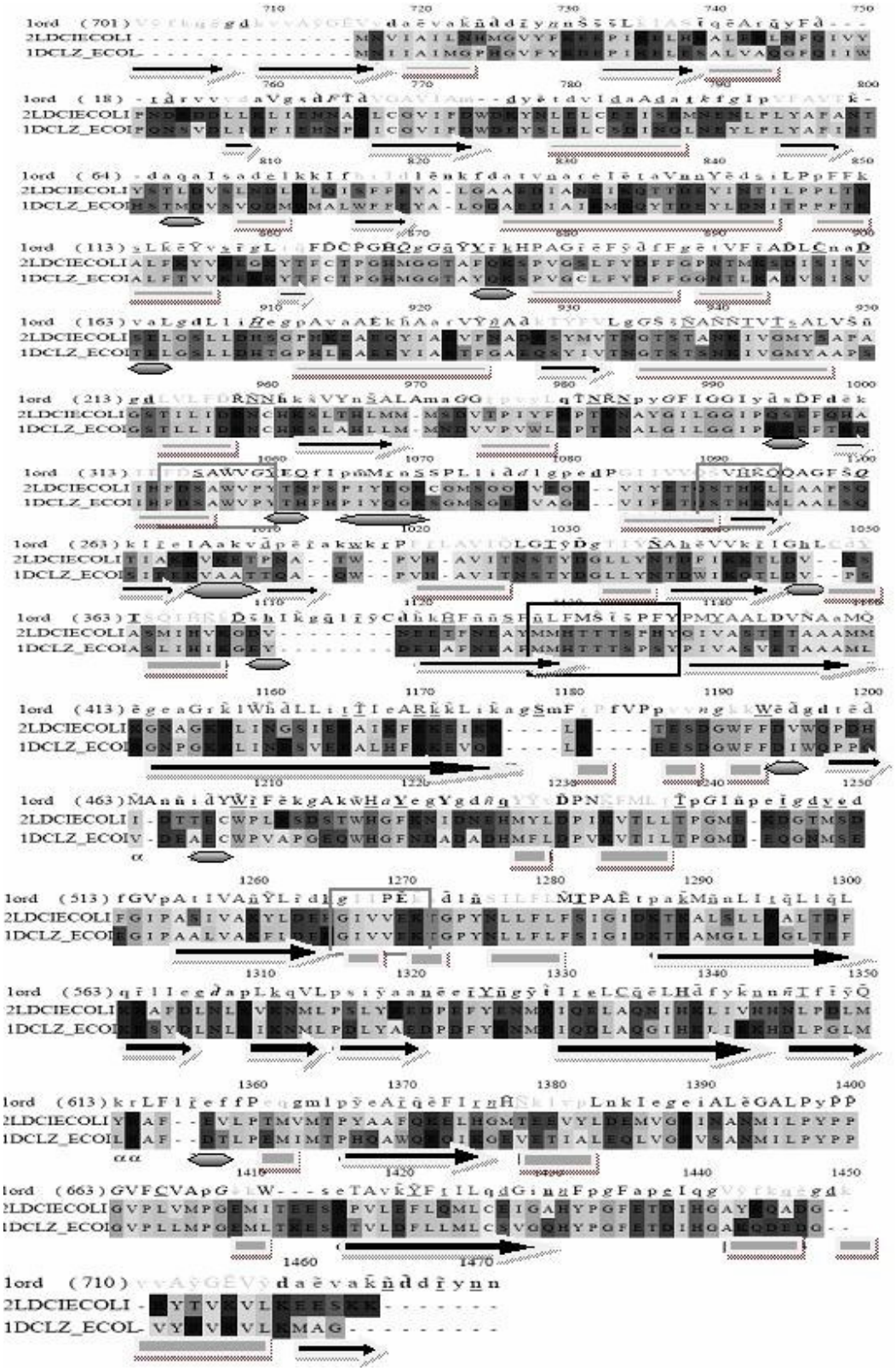

Figure 7. Predicted secondary structures. The secondary structures of the sequences of constitutive (1DCLZ_ECOL) and inducible (2LDCIECOL) lysine decarboxylase of *E. coli* were predicted and aligned with ornithine decarboxylase of *Lactobacillus sp*. The PDB id of ornithine decarboxylase is 1ORD. The first line shows the secondary structure as seen in the PDB structure, the last line shows the predicted secondary structure. The conserved motifs are highlighted by boxes.

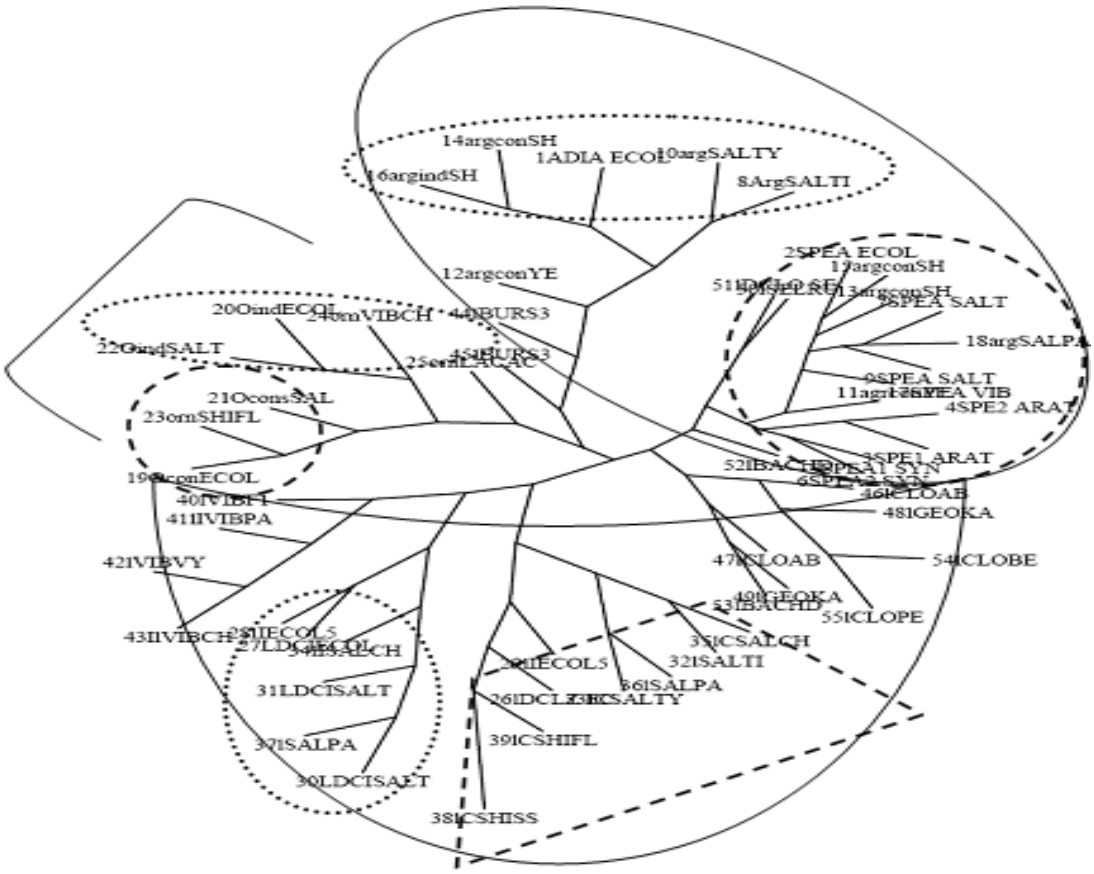

Figure 8. Phylogenetic tree of amino acid decarboxylases. The sequences within the dashed line are constitutive lysine decarboxylases. Those marked with dotted lines are the inducible lysine decarboxylases. The sequences within the ellipse are arginine decarboxylase. The sequences within the arc are ornithine decarboxylases. The sequences within the crescent shaped line represent lysine decarboxylases.

decarbo-xylase also align at corresponding positions in lysine decarboxylase (Table 7).

To determine the phylogenetic relationship between lysine decarboxylase, ornithine decarboxylase, and arginine decarboxylase, their sequences were subjected to phylogenetic analysis. Figure 8 shows that the sequences of lysine decarboxylase from gamma proteo-bacteria clearly form a separate branch. The sequences of inducible arginine decarboxylase and the sequences of ornithine decarboxylase have a common ancestor. The sequences of unannotated lysine decarboxylase are more closely related to constitutive arginine decarbo-xylase. The sequences of Vibrio sp. fall into a separate branch. Among them, one is annotated as belonging to the inducible form.

The above analyses suggest that the two forms of lysine decarboxylase are different from each other. The differences could have been due to their phylogenetic distance. Lysine decarboxylase is similar to arginine and ornithine decarboxylase and the functional motifs are also conserved among these enzymes. It is possible that cadA was integrated from a pathogen through homologous recombination. CadA has been associated with sequences of pathogenicity islands and virulence factors in many analyses.

ACKNOWLEDGEMENTS

We thank Council for Scientific and Industrial Research, India. Krithika Gokulnath was a CSIR-SRF. We gratefully acknowledge the help rendered by Dr. J. Arunachalam and Dr. D. Bharanidharan.

REFERENCES

Altshul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol., 215: 403-410.

Bailey TL, M Gribskov (1998). Combining evidence using p-values: application to sequence homology searches. Bioinformatics, 14: 48-54.

Corpet F (1988). Multiple sequence alignment with hierarchical clustering. Nucl. Acids Res., 16: 10881-10890.

Day WA, Fernandez RE, Maurelli AT (2001). Pathoadaptive mutations that enhance virulence: Genetic organization of the *cadA* regions o

Shigella spp. Infect. Immun., 69: 7471-7480.
Felsenstein J (1996). Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Meth. Enzymol., 26: 418-427.
Finlay BB, Falkow S (1997). Common themes in microbial pathogenicity revisited. Microbiol. Mol. Biol. Rev., 61: 136-169.
Gale EF, Epps HMR (1942). The effect of the pH of the medium during growth on the enzymic activities of bacteria (*Escherichia coli* and *Micrococcus lysodiekticus*) and the biological significance of the changes produced. Biochem. J., 36: 600-619.
Goldemberg SH (1980). Lysine decarboxylase mutants of *Escherichia coli*: Evidence for two enzymes. J. Bacteriol., 141: 1428-1431.
Shi J, Blundell TL, Mizuguchi K (2001). FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure- dependent gap penalties. J. Mol. Biol., 310: 243-257.
John RA (1995). Pyridoxal phosphate-dependant enzymes. Biochim. Biophy. Acta, 1248: 81- 96.
Lemmonier M, Lane D (1998). Expression of the second lysine decarboxylase gene of *Escherichia coli*. Microbiology, 144: 751-760.
Lin J, Lee IS, Frey J, Slonczewski JL, Foster JW (1995). Comparative analysis of extreme acid survival in *Salmonella typhimurium, Shigella flexneri* and *Escherichia coli*. J. Bacteriol., 177: 4097-4104.
Maurelli AT, Fernandez RE, Bloch CA, Rode CK, Fasano A (1998). Black holes and bacterial pathogenicity: A large genomic deletion that enhances the virulence of *Shigella* spp. and enteroinvasive *Escherichia coli*. Proc. Natl. Acad. Sci. USA, 95: 3943-3948.
Meng S, Bennet GN (1992a). Nucleotide sequence of the *Escherichia coli cad* operon: a system for neutralization of low extracellular pH. J. Bacteriol., 174: 2659-2669.
Meng S, Bennet GN (1992b). Regulation of the *Escherichia coli cad* operon: Location of a site required for acid induction. J. Bacteriol., 174: 2670-2678.
Merrell DS, Camilli A (1999). The *cadA* gene of *Vibrio cholerae* is induced during infection and plays a role in acid tolerance. Mol. Microbiol., 34: 836-849.
Momany C, Ghosh R, Hackert ML (1995). Structural motifs for pyridoxal-5'- phosphate binding in decarboxylases: An analysis based on the crystal structure of the *Lactobacillus* 30a Ornithine decarboxylase. Prot. Sci., 4: 849-854.
Sabo DL, Boeker EA, Byers B, Waron H, Fischer EH (1974a). Purification and physical properties of inducible *Escherichia coli* lysine decarboxylase. Biochem., 13: 662-670.
Momany C, Ernst S, Ghosh R, Chang N, Hackert ML (1995). Crystallographic Structure of a PLP-dependant Ornithine decarboxylase from *Lactobacillus* 30a to 3.0 A^{o} resolution. J. Mol. Biol., 252: 643-655.
Saha S, Raghava GPS (2006). VICMpred: SVM-based method for the prediction of functional proteins of gram-negative bacteria using amino acid patterns and composition. Genomics Proteomics Bioinformat., 4: 42-47.
Shi X, Waasdorp BC, Bennet GN (1993). Modulation of Acid-induced amino acid decarboxylase gene expression by *hns* in *Escherichia coli*. J. Bacteriol., 175: 1182-1186.
Shi X, Bennet GN (1995). Effects of multicopy LeuO on the expression of the acid-inducible lysine decarboxylase gene in *Escherichia coli*. J. Bacteriol., 177: 810-814.
Slonczewski JL, Gonzalez TN, Bartholomew FM, Holt NJ (1987). Mu d directed lacZ fusions regulated by low pH in *Escherichia coli*. J. Bacteriol., 169: 3001-3006.
Tabor H, Hafner EW, Tabor CW (1980). Construction of an *Escherichia coli* strain unable to synthesize putrescine, spermidine, or cadaverine: Characterization of two genes controlling lysine decarboxylase. J. Bacteriol., 144: 952-956.
Thompson JD, Higgins DG, Gibson TJ (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.
Vazquez-Juarez RC, Kuriakose, JA, Rasko DA, Ritchie JM, Kendall,MM, Slater TM, Sinha,M, Luxon,B.A, Popov,V.L, Waldor MK, Sperandio V, Torres AG (2008). CadA Negatively Regulates *Escherichia coli* O157:H7. Adherence Intestinal Col., 76(11): 5072-5081.