

# Research Journal of **Microbiology**

ISSN 1816-4935



Research Journal of Microbiology 6 (3): 297-303, 2011 ISSN 1816-4935 / DOI: 10.3923/jm.2011.297.303 © 2011 Academic Journals Inc.

# Cloning of Choline Dehydrogenase from *Escherichia coli*: Its Polynucleotide and Polypeptide Analysis

# L. AnbuRajan, S. Sindhuja, Y. Rajalakshmi and V. Umashankar

Department of Bioinformatics, School of Biosciences, SRM University, Ramapuram, Chennai, Tamil Nadu, India

Corresponding Author: L. AnbuRajan, Department of Bioinformatics, School of Biosciences, SRM University, Ramapuram, Chennai, Tamil Nadu, India

#### ABSTRACT

Choline dehydrogenase catalyzes the oxidation of choline to glycine betaine via betaine aldehyde in glycine betaine biosynthesis. Glycine betaine is a compatible solute, able to restore and maintain osmotic balance of living cells under stress. In this study, choline dehydrogenase (betA) gene encoding for glycine betaine biosynthesis in Escherichia coli isolated from salted shark (Scoliodon sp.) was cloned and sequenced. The betA gene sequence reported in this study contains several base substitutions with that of reported sequences in GenBank, resulting in the altered amino acid sequences of the translated proteins.

**Key words:** E. coli, compatible solutes, glycine betaine, salt stress, betA

## INTRODUCTION

Microorganisms are often exposed to the changes in environmentalparameters (Csonka, 1989; Lucht and Bremer, 1994). To cope with the environment stress, bacteria accumulate osmotically active compatible solutes, including: potassium, proline, glutamic acid, glutamine, α-aminobutyric acid, ectoine and betaine (Poolman and Glaasker, 1998). Among the compatible solutes, glycine betaine is a effective osmoprotectant (Wood et al., 2001). Glycinebetaine (N, N, N-trimethylglycine) is a quaternary ammonium compound that occurs naturally in a wide variety of plants, animals and microorganisms (Rhodes and Hanson, 1993). Numerous in vitro experiments have indicated that betaine acts as an osmoprotectant by stabilizing both the quaternary structure of proteins and cellular membrane against the adverse effects of high salinity and extreme temperatures (Gorham, 1995). Biosynthesis of betaine is catalyzed in a single step reaction by choline oxidase in soil bacterium, Arthrobacter globiformis (Ikuta et al., 1977). In higher plants, such as spinach (Rathinasabapathi et al., 1997), sugar beet and amaranth (Russell et al., 1998) betaine biosynthesis is catalyzed by choline monooxygenase in combination with betaine aldehyde dehydrogenase. In E. coli, the biosynthetic pathway for the production of glycine betaine from choline has been well characterized at the genetic level (Landfald and Strom, 1986). It has been shown that four genes encoding choline dehydrogenase (betA), betaine aldehyde dehydrogenase (betB), a putative regulator (betI) and a choline transporter (betT) are clustered in the bet operon (Andresen et al., 1988). The enzymes involved in the biosynthesis of betaine from bacteria and plants have been functionally characterized (Weretilnyk and Hanson, 1990; McAllister-Wilkins, 2003).

Recently, bacterial glycine betaine synthesizing enzymes have become a major target in developing stress tolerant crop plants of economic interest. Previous studies reports the resistance towards salinity and low temperature in transgenic tobacco expressing the two  $E.\ coli$  genes betA and betB (Holmstrom et al., 2000), signifying the practical applications of choline dehydrogenase. Choline dehydrogenase (betA) of  $E.\ coli$  catalyses the first step in glycine betaine biosynthesis, the oxidation of choline. However, this enzyme also catalyses the second step, the dehydrogenation of betaine aldehyde to betaine (Landfald and Strom, 1986). Choline dehydrogenase catalyzes the four electron oxidation of choline to glycine betaine via a betaine aldehyde intermediate (Tsuge et al., 1980). Only nominal studies in characterization of choline dehydrogenase from  $E.\ coli$  have been reported to date. In this study, we report the characterization and structural analysis of choline dehydrogenase in  $E.\ coli$ . Moreover, the sequence analysis of choline dehydrogenase from our isolate shows several base substitutions with that of reported sequences in GenBank, resulting in the altered amino acid sequences of the translated proteins.

# MATERIALS AND METHODS

Bacterial strains and plasmids: Escherichia coli was isolated from salted shark (Scoliodon sp.) procured from retail fish markets in Cochin, Kerala, India in 2004 September. Microbial identification and biochemical characterization of E. coli was carried out as per [U.S. Food and Drug Administration (USFDA)]. Escherichia coli JM109 and plasmid pTZ57R/T (MBI Fermentas, Hanover, Maryland, USA) were used as transformation host and cloning vector, respectively.

DNA extraction and gene amplification: Genomic DNA extraction from *E. coli* was performed following the method of Ausubel *et al.* (1994). Plasmid DNA was recovered from the transformed clones by alkali lysis method (Sambrook and Russell, 2001). A pair of primers bAF (5'-CGTATGCAATTTGACTACATCATT-3') and bAR (5'-GCATCATTTTTTCGCTCTCACCG-3') were designed to amplify the *betA* gene. PCR amplification was carried out with 30 cycles of 1 min at 94°C, 1.5 min at 50°C and 2 min at 72°C. Additional extension was carried out for 5 min at 72°C using high fidelity PCR enzyme mix (MBI Fermentas).

Cloning and sequencing: The betA gene amplicon was purified using Perfectprep Gel Cleanup Kit (Eppendorf, Germany) and cloned into pTZ57R/T according to the manufacture's instructions. The pTZ57R/T-betA construct was transformed into E. coli JM109 (recA1, endA1, gyrA96, thi-1, hsdR17 (rK-mk+), e14-(mcrA-), supE44, relA1, Ä(lac-proAB)/F' [traD36, proAB+, lac Iq, lacZÄM15]. Positive transformants were selected for PCR amplification with vector primers M13f-M13r (MBI Fermentas) and the clones with the correct insert as judged by size were sequenced on ABI PRISM 377 genetic analyzer (Applied Biosystems Inc., Foster City, CA, USA).

In silico sequence analysis: The nucleotide sequences obtained were analysed with the available database sequences by BLAST analysis using the NCBI (http://www.ncbi.nlm.nih.gov). The sequences were aligned and clustered using CLUSTAL-X version 1.81 (Thompson et al., 1997). The output alignments were imported into the GeneDoc program (http://www.psc.edu/biomed/genedoc/) and BioEdit version 7.05 program (www.mbio.ncsu.edu/BioEdit/) to calculate the percent identities among the nucleotide and amino acid sequences. The molecular masses and the theoretical pl values ofthe polypeptides were predicted using the ProtParam tool (http://www.expasy.org/tools/protparam.html). The secondary structure prediction was performed using the PROTEAN program (DNASTAR, Inc., Madison, USA).

#### RESULTS AND DISCUSSION

Cloning and sequence analysis: Based on the sequence analysis, it was previously reported that the betA gene of E. coli codes for choline dehydrogenase (Lamark et al., 1991). To date, only minimum reports on the characterization of choline dehydrogenase has been reported (Tsuge et al., 1980; Gadda and McAllister-Wilkins, 2003). As the first step towards the functional characterization of choline dehydrogenase, in this study we cloned and characterized the betA gene from E. coli cells under the control of inducible promoter.

The choline dehydrogenase encodes the polypeptides comprised of 556 amino acids with the calculated molecular masses of 61848 Da., based on the *in silico* estimates. The *betA* gene coded for choline dehydrogenase was PCR amplified and is found to have the polynucleotides of 1671 bp length (Fig. 1). The *betA* amplicon was purified from agarose gel, ligated in pTZ57R/T cloning vector and transformed into *E. coli* JM109. Plasmid pTZ57R/T with *betA* gene insert was confirmed by nucleotide sequencing. The nucleotide sequence of *betA* gene was submitted to GenBank and have been given accession no. FJ823260.

The  $in\ silico$  sequence analysis of betA gene revealed a total of thirty-one base substitutions at the nucleotide level (Fig. 2) with that of the sequences deposited in the GenBank (accession nos. X52905 and M77738). However, only two of these changes translated into change of amino acids. The differences were observed at positions 133 and 452, which resulted in the amino acid substitution of valine with methionine and of threonine with valine residue, respectively (Fig. 3). No internal stop codons were observed due to the base substitution in the gene. Upon BLAST analysis it was found that the deduced amino acid sequence of betA was highly homologous to choline dehydrogenase of reported strains; 99% identity with E. coli (accession no. X52905) and 98% identity with E. coli (accession no. M77738). To authenticate the nucleotide variation in the betA sequence of our isolate, the gene was yet again PCR amplified from the genomic DNA with Taq DNA polymerase (Dynazyme II, Espoo, Finland). The underlying principle for this strategy was that feasible misincorporations of nucleotides in gene amplification might occur at different positions by using different DNA polymerases and PCR protocols. The PCR product was cloned in pDrive vector (Qiagen, Germany) and sequenced. The nucleotide sequence of the betA gene ligated in pDrive was the same as dogged using pTZ57R/T-betA cassette. This result confirms that the base divergence was conserved in the genome and not due to the external parameters.

Evaluation of the deduced amino acid sequence of betA gene with reported choline dehydrogenase sequences in the database revealed a maximum similarity. However, the sequence

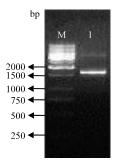


Fig. 1: Agarose gel electrophoresis of betA gene. Lane 1: betA amplicon; Lane M: DNA molecular weight marker

# Res. J. Microbiol., 6 (3): 297-303, 2011

	10	20	30	40	50	50	70	80	90	100
				 				II		
EU0325	35AT GCAATTTGA CTA C	ATCATTATTO	GTGCCGGCT	RGCCGGCRAC	CGTTCTCGCT	CCCGTCTGAC	T GRAGATOC	MATRCCTCC	GT GCT GCT GCT	
X52905 H77738	T									
	110									200
								l l	I I	11
	35 C G G G C G G C C G G A C T									
					.TC	A	<b>T</b>			G
	210 	220	230	240	250	250	270	280	290	300
EU0325	35GCCTGAACCGTTTAT	GRATARCCG	CGCATGGAG	receencece:	TARAGGCCT	GGTGGATCGT	CGCTGATCA	CGGCATGTG	CTACATCCGT	GGCRAT
	A									
H77738	A	320	330	340	T 350	360	370	380	390	400
	310    35GCGCTGGATCTCGAT		I I	l	I I			I I	1	11
EU0325	35 G C G C T G GAT C T C GAT	AACT GGGCG(	MAGAACCCG	FT CT GGA GAA	TGGAGTTAT( C C	TCGACTGCCT	GCCCTACTA(	CGCARGGCC	GRORCTCGCG	ATGTGG
H77738					C C.					A
	410 	420	430	440	450	450	470	480	490	500
EU0325	 35GCGAGAACGACTACC	ACGGTGGCŒ	TGGCCCGGT	PAGCGT CACCE	LCCTCCAAAC	CGGCGTCAAT	CCGCTGTTT(	PARGCGATGA	TTGAAGCGGG	CGTGCA
X52905										
H77738		520	530	T. 540	5.50	560	570	580	590	600
	510 		I I	l	I I			I I	1	11
	35 GGCAGGCTAC C CGC G									
M77738	G		T			•				C
	510 	520	630	540	550	660	670	580	590	700
EU0325		GAT CA GG CC	AATCGCGCC	TARCCT GAC	ATT CGTACT	ACGCCATGAC	CGATCACAT(	ATTTTTGAC	GGCARAC GCG	CGGTGG
X52905	T C		<b>T</b> .			<b>T</b>				
H77738	TC	720	T	740	7.50	750	220	780	790	800
	710    35GCGTCGAATGGCTGG		I	II	I I			IĬ	1	11
EU0325	35GCGTCGAATGGCTGG	AAGGCGACA(	GACCATCCC	ACCCGCGCAI	CGGCCAACA	AGAAGT GCTG	TTAT GT GCA	GC GC GATT G	CCTCACCGCA	GATCCT
H77738										
	810    35GCRACGCTCCGGCGT	820	830	840	850	850	870	880	890	900
EU0325	35GCAACGCTCCGGCGT	CGGCAACGC	GAACTGCTG	 SCGGAGTTTGF	 ATATT CC GCT(	GTGCATGAAT	TACCCGGCG	  CGGCGRARA	TCTTCAGGAT	CATCTG
X52905										
H77738										
	910	920	930	940	9.50	950	970	980	990	1000
	910 	920 	930 	940 	950 	950	970 	980 	990 	1000 
	35 GA GAT GTATCT G CAA	TAT GA GT GC	laagaa cogg:	TTCCCTCTA(	CCTGCCCTG	agregregaa	CCAGCCGAA	AT CGGT GC G	GAGT GGCT GT	1000    TTGGCG
EU0325 X52905 H77738	35 GA GAT GTATCT G CAA	TAT GA GT GC	ARGAR CC GG	FTT C CCT CTA(	CCTGCCTG	AGT GGT GGAA	CCAGCCGAAL	ATCGGTGCG	GAGT GGCT GT	rtggcg
X52905	35 GA GAT GTATCT G CAA	TAT GA GT GC	ARGAR CC GG	FTT C CCT CTA(	CCTGCCTG	AGT GGT GGAA	CCAGCCGAAL	ATCGGTGCG	GAGT GGCT GT	rtggcg
X52905 H77738 EU0325	35 GR GRT GTATCT GCRA 1010    35 GCRCT GGC GTT GGT G	TAT GA GT GCP	LAAGAACCGG	TTTCCCTCTA(	1050	RGT GGT GGPR  1050  1  TGR GGRATTT	1070	RTCGGTGCG	GAGT GGCT GT:	TTGGCG  1100    GCCAGT
X52905 H77738 EU0325 X52905	35 GA GAT GTATCT GCRA 10 10    35 GCACT GGC GTT GGT G	TRT GR GT GCF  1020	LAGAACCGG 1030    LCTTTGAAGCI	TTT C GCT CTA (	1050	RGTGGTGGAR 1050     TGRGGRATTT	1070	PAT CGGT GCG	GAGT GGCT GT:	TTGGCG
X52905 H77738 EU0325	35 GA GAT GTATCT GCAA 10 10 	TAT GA GT GC 1020 	LAGRACCGG 1030    LCTTTGRAGC	ITT C GCT CTA ( 10 40    AGGT GGATTTA	1050	RGT GGT GGPR 1050     YT GRGGRATTT	CCAGCCGAAI 1070 	AT CGGT GCG 1080 	GAGT GGCT GT	TTGGCG
X52905 H77738 EU0325 X52905 H77738	1010 1010 135 GCACT GGC GTT GGT G	1020 1020 CCAGCNACC	LAGARCCGG	1040	1050	LOSO  LOSO  TORGERATIT  LISO	1070 	1080    	1090 LL ACCATTTCCT	TTGGCG
X52905 H77738 EU0325 X52905 H77738	35 GR GRT GTATCT GCRA  1010 35 GCRCT GGC GTT GGT G  1110 35AGCGRTTARCTATAR	1020 	1030 	1040    1140	1050    1150	LGTGGTGGRA 1050 L   TGRGGRATTT 1150 LLE0 LLE0 GTCGGCTCRA	1070 	1080    ATRTTCRGT	LOGO GET	1100    GCCAGT    1200    ATTRAR
X52905 H77738 EU0325 X52905 H77738	DOID  DOID  SECRET GRACE GGG  1110  1110  SAGGGRITARCTATAR	TAT GAGT GCI	LARGAR CC GG:	1040 	1050 LITTGGCAGCCC	LGT GGT GGLA	1070  - GCGTGGCCG	ATCGGTGCG  1080  L ATRITCAGT:  1180  L  RAGCCGTGG	LOSO LOSO LOSO LOSO LOSO LOSO LOSO LOSO	TTGGCG  1100    GCCRGT  1200    RTTRAR
X52905 H77738 BU0325 X52905 H77738 BU0325 X52905	DOID  DOID  SECRET GRACE GGG  1110  1110  SAGGGRITARCTATAR	TAT GAGT GCI	LARGAR CC GG:	1040 	1050 LITTGGCAGCCC	LGT GGT GGLA	1070  - GCGTGGCCG	ATCGGTGCG  1080  L ATRITCAGT:  1180  L  RAGCCGTGG	LOSO LOSO LOSO LOSO LOSO LOSO LOSO LOSO	TTGGCG  1100    GCCRGT  1200    RTTRAR
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	1010 35 GCA GAT GTATCT GCAN 1010 35 GCACT GGC GTT GGT G 1110 35 AGCGATTAACTATAR 1210 1210 135 TCCCGGCGACCCGCAC	TAT GAGT GCT  1020  CCR GCARCCT  1120  CGGCT CGRAT  1220  CRGCATCC CGRAT	LANGRACC GG*  1030  LCTTT GRAGCI  1130  GCRGT GRAR  1230	10 40  LOGGERTTI  LL40  L240  LL40  LL40  LL40  LL40	1050 LITT CGCAGCCC 1150 LTT CGCAGCCCAGCCCAGCCCAGCCCAGCCCAGCCCAGCC	LEGT GGT GGRA  1060     HT GR G GRATTI  1160     GTC GGCT CRA  1260     LEGGACT GGCR	1070 	1080 	1090	1100    GCCRGT    1200    RTTRAR    1300    GCGRGR
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	1010 35 GCACT GCACT 35 GCACT GGC GTT GGT G 1110 35 AGCGATTARCTATAR 1210 35 AGCGATTARCTATAR 35 AGCGATTARCTATAR	1020 1020 1120 1220 1220	1030 LCTTT GRAGCI 1130 LCGCGT GRARI	1040  1  1040  1  AGGT GGATTI  1140  1  AGCAC GGTT  1240  1  ITARCTACAT	1050 LLATTCGCAGCCC 1150 CCAGTGCCAC	LAGT GGT GGAA  1050	1070 	1080 1080 11 ATATT CAGT. 1180 11 RAGCCGTGG	1090 1 ACCATTTCCT  1190 1 GCRTGTGGGG	1100      1200      1200      17TARR
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	1010 35 GCACT GCAT GTATCT GCAT 1010 35 GCACT GGC GTT GGT G 1110 35 AGCGATTARCTATAR 1210 35 TCCCGGCGACCC GCAC	1020 CCRGCRACC  1120 CCGGCT CGRRI  1220 CCGGCTTCCGC	LANGRACC GG:  1030  COTTT GRAGCI  1130  GCRGT GRARA  1230  1230	1040  L	1050 1050 1050 1050 1150 1150 1050 1250 1250 1250	LEGT GGT GGRA  1050  1050  TGR GGRATTT  1150  1250  1250  1250  LEGGRATT GGRA	1070	1080	LOSO LOSO LOSO LOSO LOSO LOSO LOSO LOSO	1100     12
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 K77738	1010 35 GCACT GCAT GTATCT GCAN 1010 35 GCACT GGC GTT GGT G 1110 35AGCGATTARCTATAR 1210 35T CCCGGGACCC GCAC	1020 1020 CCRGCRACC 1120 1220 CRGCTCCRAI	LARGERACC GG:  10:30  LCTTT GRAGCI  11:30  LCTT GRARACCI  12:30  LCTT GTT GTT	10 40  LO 40	1050 LISO LISO LISO LISO LISO LISO LISO LISO	LAGT GGT GGRA  1050  1150  1150  1250  1250  1250  1250	1070 1170 1170 1170 1170 1270 1270	LOSO LOSO LOSO LOSO LOSO LOSO LOSO LOSO	1090 1 ACCATTTCCT 1190 1 GCRTGTGCGG1 1290 1 1290 1 1290	TTGGCG
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 K72905 H77738	1010 35 GCACT GCCTT GGT G 1110 35 GCACT GGC GTT GGT G 1110 35 AGCGATTAACTATAA 1210 1210 1210 1210	1020 1020 1020 1120 1220 1220 1320 1320	LARGERACC GG:  10:30  LCTTT GRAGCI  11:30  12:30  LCGATT GT GT.	10 40  1	1050 1050 1150 1150 1250 1250 1250 1250	LEGT GGT GGRA  1050  1150  1150  1250  1260  1260  1360  1360	1070	ATCGGTGCG  1080  1 ATTATT CAGT: 1180  1280  1280  1380  1380  1280	1090 1	1100
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	JOIN JOSE GALET GEAT GEAT GEAT GEAT GEAT GEAT GEAT GE	1020 1020 1120 1120 1220 1220 1320 1320	LARGERA CC GG:  10:30  LCTTT GRA GC:  11:30  LCTT GRA GC:  12:30  LCTT GTT GTT GTT GTT GTT GTT GGT GGG	10 40  L	1050 1050 1050 1150 1150 1150 1250 1250	LAGT GGT GGRA  1050  1150  1150  1250  1250  1250  1250  1250  1250  1250  1250  1250  1250  1250  1250  1250  1250  1250  1250	1070	LOSO LOSO LOSO LOSO LOSO LOSO LOSO LOSO	1090 1 RCCATTTCCT 1190 1 GCRTGTGCGGG	1100   1  GCCRGT   1200   1  RTTRRR   1300   1  GCGRGR   1  GCGRGR   1
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	1010 35 GCACT GGC GTT GGT GGT GGT GGT GGT GGT GGT G	1020 1020 1120 1220 1220 1320 1320 1320	1030 1130 1230 1230 1230 1330 1430	10 40  1	1050 1 1150 1 1250 1 1250 1 1250 1 1250 1 1250 1 1250 1 1250 1 1250 1 1250	LEGIGGIGGE  1060  1160  1260  1260  1360  1360  1460	1070 1070 1170 1170 1270 1270 1370 1370 1470	1080 1080 1180 1180 1280 1280 1380 1480	1090	1100   1  GCCRGT   1200   1  RTTRRR   1300   1  GCGRGR   1400   1  TRACCR   1500
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	JOINO SEGRATICATOTECAN JOINO SEGRATICATORE JULIO SEGRATIANA JULIO SETCATORE JU	1020 1020 1020 1120 1120 1220 1220 1220	LARGERACC GG:  10:00  LCTTT GRAGCI  11:00  LCTT GRAGCI  11:00  LCTT GRAGCI  12:00  LCTT GGT GGT GGT GGT GGT GGT GGT GGT GGT	1140 1240 1240 1240 1240 1240 1240 1240	1050 1 2TTCGCRGCCCC 1150 1 1250 1 1250 1 2TCGCRCGCRCCR 1250 1	LEGT GGT GGPA  1050   HTGRGGRATTT  1150   1250   LEGGRATG GGRA  1250	1070	ATCGGTGCG  1080  1 ATTATT CRGT.  1180  1 EAGCCGTGG  1280  1280  1380  1480  1480	1090 1 ACCATTT CT 1190 1 GCAT GTGCGG1 1290 1 1290	1100   1
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	1010 35 GCACT GGC GTT GGT G 1110 35 GCACT GGC GTT GGT G 1110 35 AGCGATTAACTATAA 1210 1210 1210 1210 35 T CCCG GCACC GGCAC 1410 35 CCCG GRAACC GCCT 1410 35 CGCG GRAACC GCCTT	1020 1020 1120 1220 1220 1320 1320 1320	1030  LCTTT GRAGCI  1130  1230  LCTAT CF GG:  1430  LCTAT CF GG:	1040  1	1050 1050 1150 1150 1250 1250 1250 1250	LEGIGGIGGER  1050  1150  1150  1150  1250  1250  1250  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350	1070 GCGTGGCCGARI  1170 TGCGCTCGCC  1270 1370 ACGGATTGARI	1180 1180 1280 1280 1280 1280 1480 1	1090 1 1190 1 GCRTTTCCT 1290 1 CGCRTCACCC 1390 1 1390 1 RGTTCGTGCG	1100   1    1200
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	JOIN JOSE GALEGE GRANGE GALE JOIN JOSE GALEGE GEORGE GALEGE JULIO JOSE GALEGE GALEGE GALEGE JULIO JOSE GALEGE GALEGE GALEGE JULIO JOSE CAT GALEGE GALEGE GALEGE JULIO JOSE CAT GALEGE GALEGE GALEGE	1020 1020 1120 1120 1220 1220 1320 1320	LARGERACCIGGE  1090  LOTTT GRAGCI  1190  LOTT GRAGCI  1290  LOTT GT GT  1390  LOTT CGT GG	1140 1 1240 1 1240 1 1240 1 1240 1 1240 1 1440 1 1440	1050 1 ATT GGCRGCCA 1150 1 CCRGTGCCAC 1250 1 1250 1 TGGCCCGGCAC TGG 1450 1 1450	LAGT GGT GGRA  1050  1150  1150  1250	1070	1080 1 ATATT CAGT. ATATT CAGT. 1180 1 PARGCC GTGG. 1280 1 1280 1 1280 1 1280 1 1280	1090 1 RCCATTT CCT 1190 1 GCRT GTGCGGG 1 1290 1 1290 1 1290 1 1390 1 1390 1 1490 1 1490	1100   1    GCCRGT   1200   1    RTTRRR   1300   1    GCGRGR   1    I    TRACCR   1    I
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	JOIN JOSE GRAT GRATCT GCAN JOIN JOSE GCACT GGC GTT GGT G JOIN JOSE GCACT GGC GTT GGT G JOIN JOSE GCACT GCC GCAC JOIN JOSE GCACT GCAT CAACCC G JOIN JOSE GCAC GAACCC GCAC JOIN JOSE GCAC GAACCC GCAC JOIN JOSE GCAC GAACCC GC	1020 1020 1120 1120 1220 1220 1320 1320	LARGERA CC GG:  1030  LUTTT GRAGCI  1130  CGCRGT GRARA:  1230  LGTRT CGT GG:  1430  LGTRT CGT GG:  1430  LGTRT CGT GG:  1430	1140 1140 1140 1140 1140 1140 1140 1140	1050 1050 1150 1150 1250 1250 1250 1250	LAGT GGT GGRA  1050  1150  1150  1250	1070 GCGTGGCCGA1 1170 TGCGCTCGCC 1270 1270 1270 1270 1270 1270 1270 1270	LOSO LOSO LOSO LOSO LOSO LOSO LOSO LOSO	1090 11 ACCATTT CCT 1190 11 GCRT GTGCGG1 1.290 11 CGCRT CR CCC 1390 11 AGTT CGT GCGCT 1490 11 CGGGTT R GRA.	1100   1100
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	JOINO JOIN JOIN	1020 1020 1020 1120 1120 1220 1220 1320 1420 1420 1520 1520 1520	LARGERACC GG:  10:00  LCTTT GRAGCI  11:00  LCTTT GRAGCI  11:00  LCTTT GRAGCI  12:00  LCTT GRAGCI  12:00  LCTT GGT  12:00  LCTT GGT  13:00  LCTT GGT  13:00  LCTT GGT  LCTT GGT  15:00  LCTT GGT  LCT	1140 1240 1240 1240 1240 1240 1240 1240	1050 1 2TTCGCRGCCCC 1150 1 1250	LEGT GGT GGPA  1050  1150  1150  1250	1070	ATTCGGTGCG  1080  1 ATTTTCRGT.  1180  1180  1280  1280  1280  1380  1480  1480  1580  1580  ARARATEGG	1090 1	1100   1    GCCRGT   1    STTRRR   1    1    GCGRGL   1    1
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	JOIN	1020 CCRGCRACC  1120 CCGGCTCGRAI  1220 CCGCTCGCATCGC  1320 CCGCTCGCTCGC  1420 CCCTCCGTCGCTCCGTCCCTCCCTCCCTCCCTCCCTCCC	LARGERACC GG:  1090  LOTTT GRAGCI  1190  LOTT GRAGCI  1290  LOTT GGT  1290  LOTAT CGT GG:  1420  LOTAT CGT GG:  1520  LOTAT CGT GG:  1520	1140 1 1240	1050 1 ATT GGCAGGCCA 1150 1 CCCAGTGCCAC 1250 1 ATGCGCAGGCAGC 1250 1 TGG 1.450 1 AGGCCCGGCAGC 1450 1 TGG 1 AGGCCGAGAGC 1 TGG 1 AGGCCGAGAGC 1 TGG 1 AGGCCGAGAGC 1 TGG 1 AGGCCGAGAGC 1 AGGCCGAGAGC 1 AGGCCGAGAGC 1 AGGCCGAGAGC 1 AGGCCGAGAGC 1 AGGCCGAGAGC 1 AGGCCGAGAGCCGAGACC 1 AGGCCGAGAGCCGAGACC 1 AGGCCGAGAGCCGAGACC 1 AGGCCGAGAGCCGAGACC 1 AGGCCGAGAGCCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCCGAGACC 1 AGGCCGAGACCCGAGACC 1 AGGCCGAGACCCGAGACC 1 AGGCCGAGACCCGAGACC 1 AGGCCGAGACCCCGAGACC 1 AGGCCGAGACCCCGAGACC 1 AGGCCGAGACCCCGAGACC 1 AGGCCGAGACCCCGAGACC 1 AGGCCCCGAGACCCCCGAGCCCCCCCCCCCCCCCCC	LRGT GGT GGRA  1050  1150  1150  1250  1250  1250  1250  1250  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350	1070  GCGTGGCCGA1  1170  TGCGCTCGCC  1270  1270  1270  1470  ACGGGTGARGA  1570  GRATTGGCGA	1180 1280 1280 1280 1280 1480 1480 1480 1480 1480 1480	1090 1	1100
X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	JOIN	1020 CCRGCRACC  1120 CCGGCTCGRAI  1220 CCGCTCGCATCGC  1320 CCGCTCGCTCGC  1420 CCCTCCGTCGCTCCGTCCCTCCCTCCCTCCCTCCCTCCC	LARGERACC GG:  1090  LOTTT GRAGCI  1190  LOTT GRAGCI  1290  LOTT GGT  1290  LOTAT CGT GG:  1420  LOTAT CGT GG:  1520  LOTAT CGT GG:  1520	1140 1 1240	1050 1 ATT GGCAGGCCA 1150 1 CCCAGTGCCAC 1250 1 ATGCGCAGGCAGC 1250 1 TGG 1.450 1 AGGCCCGGCAGC 1450 1 TGG 1 AGGCCGAGAGC 1 TGG 1 AGGCCGAGAGC 1 TGG 1 AGGCCGAGAGC 1 TGG 1 AGGCCGAGAGC 1 AGGCCGAGAGC 1 AGGCCGAGAGC 1 AGGCCGAGAGC 1 AGGCCGAGAGC 1 AGGCCGAGAGC 1 AGGCCGAGAGCCGAGACC 1 AGGCCGAGAGCCGAGACC 1 AGGCCGAGAGCCGAGACC 1 AGGCCGAGAGCCGAGACC 1 AGGCCGAGAGCCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCGAGACC 1 AGGCCGAGACCCGAGACC 1 AGGCCGAGACCCGAGACC 1 AGGCCGAGACCCGAGACC 1 AGGCCGAGACCCGAGACC 1 AGGCCGAGACCCCGAGACC 1 AGGCCGAGACCCCGAGACC 1 AGGCCGAGACCCCGAGACC 1 AGGCCGAGACCCCGAGACC 1 AGGCCCCGAGACCCCCGAGCCCCCCCCCCCCCCCCC	LRGT GGT GGRA  1050  1150  1150  1250  1250  1250  1250  1250  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350  1350	1070  GCGTGGCCGA1  1170  TGCGCTCGCC  1270  1270  1270  1470  ACGGGTGARGA  1570  GRATTGGCGA	1180 1280 1280 1280 1280 1480 1480 1480 1480 1480 1480	1090 1	1100
EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	1010 35 GCACT GGC GTT GGT GGT GGT GGT GGT GGT GGT G	1020 1020 1120 1120 1220 1220 1220 1320 13	LARGRACC GG:  1030  LCTTT GRAGCI  1130  1230  LCGRATT CF GT:  1330  LGTRT CF GG:  1420  LGGRACCT GCI  1530  LGGRACCT GCI  1630	1140 1140 1140 1140 1140 1140 1140 1140	1050 1050 1150 1150 1250 1250 1250 1250	LAGT GGT GGRA  1060	1070  CCRECCERI  1170  TGCGTGGCCG1  1270	1180 1280 1280 1280 1280 1480 1480 1480 1480 1480 1480	1090 1	1100
EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738 EU0325 X52905 H77738	1010 35 GCACT GGC GTT GGT G 1110 35 GCACT GGC GTT GGT G 1110 35 AGCGATTAACTATAA 1210 1210 1310 1410 1410 1510 35 CGCGCARACCGCCTT 1510 35 CGCGTGT GGT GGAT GGAT GGAT GGAT GGAT GG	1020 1020 1020 1120 1120 1220 1220 1320 13	1130 1230 1230 1430 1430 1430 1530 16GERGEGERTER 1630 1630	1040  LOGGERATTI  1140  LOGGERATTI  1240  LOGGERATE  1340  LOGGERATE  1440  LOGGERATE  L	1050 1.150	1050 1050 1050 1150 1150 1250 1250 1260 1260 1260 1260 1260 1260 1260 126	1070  GCGTGGCCGARJ  1170  1270  1270  1370  1470  1470  1570  CRITGGCARGCCARGCCARGCCARGCCARGCCARGCCARGCCA	1180 1280 1280 1280 1280 1480 1480 1480 1480 1480 1480	1090 1	1100

Fig. 2: BioEdit analysis of nucleotide substitutions of choline dehydrogenase (betA) in  $E.\ coli$  [FJ823260] (this study), with betA genes from  $E.\ coli$  strains [X52905 and M77738]

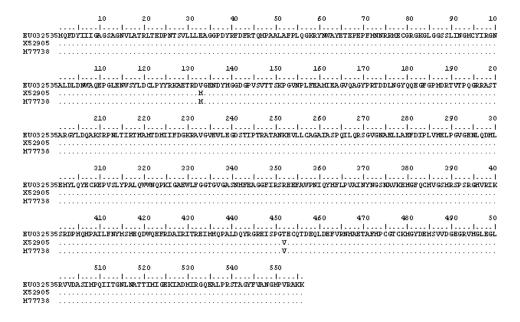


Fig. 3: BioEdit analysis of amino acid substitutions of choline dehydrogenase (betA) in E. coli [FJ823260] (this study), with betA genes from E. coli strains [X52905 and M77738]

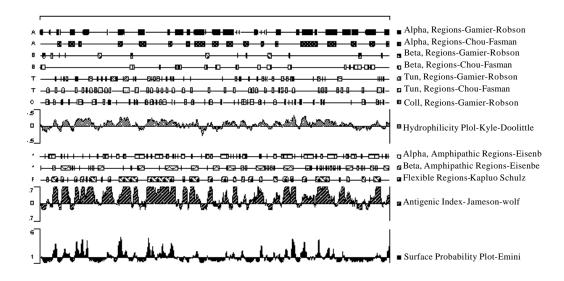


Fig. 4: Secondary structure prediction of choline dehydrogenase using PROTEAN

analysis of betA of our isolate showed several base substitutions with that of reported sequences, resulting in the altered amino acid sequences of the translated proteins.

Secondary structure prediction analysis: The secondary structure of choline dehydrogenase was predicted to have the alpha-helical structure with maximum hydrophilic molecules. The prediction analysis also revealed the presence of much acidic amino acids, regions with high antigenicity and very high backbone chain flexibility (Fig. 4). Upon analysis of *betA* protein, the

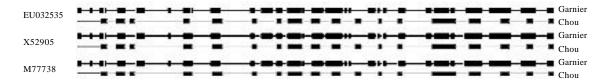


Fig. 5: Secondary structure analysis using PROTEAN. The analysis was performed using choline dehydrogenase from *E. coli* [FJ823260] (this study) and from *E. coli* strains [X52905 and M77738]

predicted charge at pH 7.0 was 15.45 with the isoelectric point of 5.52. Common amino acids include 58% glycine, 45% alanine, 41% glutamic acid, 40% leucine, 37% arginine, 33% praline and 30% each of isoleucine and aspartic acid.

The secondary structure prediction results also exhibited considerable similarity with the reported choline dehydrogenases from *E. coli* strains. Upon structural analysis, both Chou-Fasman and Garnier-Robson prediction of alpha helix structures showed a maximum similarity. These results suggested that the active domains of the enzyme from our isolate have the considerable identity with the database reports (Fig. 5).

In conclusion, this study represents the first instance in which the choline dehydrogenase in  $E.\ coli$  isolated from salted fish has been cloned and characterized in detail. Moreover, the determination of protein structure modification due to the nucleotide substitutions will certainly provide the basis for performing site-directed mutagenesis to improve the production and configuration of the osmolytes of biotechnological interest. This, in turn, has great potential in biotechnological applications aimed at stress tolerance in crop plants of economic interest.

#### ACKNOWLEDGMENTS

Authors are grateful to the Pro-Vice Chancellor, SRM University, Ramapuram, Chennai for providing the necessary facilities to carry out this research work. The research concept was developed by L.A.R and research experiments were performed by S.S. and Y.R; V.U. organized for and provided the necessary facilities.

### REFERENCES

Andresen, P.A., I. Kaasen, O.B. Styrvold, G. Boulnois and A.R. Strom, 1988. Molecular cloning, physical mapping and expression of the bet genes governing the osmoregulatory choline-glycine betaine pathway of *Escherichia coli*. J. Gen. Microbiol., 134: 1737-1746.

Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl, 1994. Current Protocols in Molecular Biology. John Wiley and Sons, New York.

Csonka, L.N., 1989. Physiological and genetic responses of bacteria to osmotic stress. Microbiol. Rev., 53: 121-147.

Gadda, G. and E.E. McAllister-Wilkins, 2003. Cloning, expression and purification of choline dehydrogenase from the moderate halophile *Halomonas elongata*. Applied Environ. Microbiol., 69: 2126-2132.

Gorham, J., 1995. Betaines in Higher Plants-Biosynthesis and Role in Stress Metabolism. In: Amino Acids and their Derivatives in Higher Plants, Wallsgrove, R.M. (Ed.). Cambridge University Press, Cambridge, pp. 171-203.

- Holmstrom, K.O., S. Somersalo, A. Mandal, T.E. Palva and B. Welin, 2000. Improved tolerance to salinity and low temperature in transgenic tobacco producing glycine betaine. J. Exp. Bot., 51: 177-185.
- Ikuta, S., S. Mamura, H. Misaki and Y. Horiuti, 1977. Purification and characterization of choline oxidase from *Arthrobacter globiformis*. J. Biochem., 82: 1741-1749.
- Lamark, T., I. Kaasen, M.W. Eshoo, P. Falkenberg, J. McDougall and A.R. Strom, 1991. DNA sequence and analysis of the bet genes encoding the osmoregulatory choline-glycine betaine pathway of *Escherichia coli*. J. Mol. Microbiol., 5: 1049-1064.
- Landfald, B. and A.R. Strom, 1986. Choline-glycine betaine pathway confers a high level of osmotic tolerance in *Escherichia coli*. J. Bacteriol., 165: 849-855.
- Lucht, J.M. and E. Bremer, 1994. Adaptation of *Escherichia coli* to high osmolarity environments: Osmoregulation of the high-affinity glycine betaine transport system proU. FEMS Microbiol. Rev., 14: 3-20.
- Poolman, B. and E. Glaasker, 1998. Regulation of compatible solute accumulation in bacteria. Mol. Microbiol., 29: 397-407.
- Rathinasabapathi, B., M. Burnet, B.L. Russel, D.A. Gage and P.C. Liao *et al.*, 1997. Choline monooxygenase, an unusual iron-sulphur enzyme catalyzing the first step of glycine betaine synthesis in plants: Prosthetic group characterization and cDNA cloning. Proc. Natl. Acad. Sci. USA., 94: 3454-3458.
- Rhodes, D. and A.D. Hanson, 1993. Quaternary ammonium and tertiary sulfonium compounds in higher-plants. Ann. Rev. Plant Physiol. Plant Mol. Biol., 44: 357-384.
- Russell, B.L., B. Rathinasabapathi and A.D. Hanson, 1998. Osmotic stress induces expression of choline monooxygenase in sugar beet and Amaranth. Plant Physiol., 116: 859-865.
- Sambrook, J. and D.W. Russell, 2001. Molecular Cloning: A Laboratory Manual. 3rd Edn., Cold Spring Harbor Laboratory Press, New York, ISBN: 9780879695774, pp: 1-99.
- Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins, 1997. The clustal X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acids Res., 25: 4876-4882.
- Tsuge, H., Y. Nakano, H. Onishi, Y. Futamura and K. Ohashi, 1980. A novel purification and some properties of rat liver mitochondrial choline dehydrogenase. Biochim. Biophys. Acta, 614: 274-284.
- Weretilnyk, E.A. and A.D. Hanson, 1990. Molecular cloning of a plant betaine-aldehyde dehydrogenase, an enzyme implicated in adaptation to salinity and drought. Proc. Natl. Acad. Sci. USA., 87: 2745-2749.
- Wood, J.M., E. Bremer and N. Laszlo, 2001. Osmosensing and osmoregulatory compatible solute accumulation by bacteria comparative. Biochem. Physiol., 130: 437-460.