

Phylogenetic Study of the Evolution of PEP-Carboxykinase

Sanjukta Aich and Louis T.J. Delbaere

Department of Biochemistry, University of Saskatchewan, Saskatoon, Saskatchewan, S7N 5E5, Canada.

Abstract: Phosphoenolpyruvate carboxykinase (PCK) is the key enzyme to initiate the gluconeogenic pathway in vertebrates, yeast, plants and most bacteria. Nucleotide specificity divided all PCKs into two groups. All the eukaryotic mammalian and most archaeal PCKs are GTP-specific. Bacterial and fungal PCKs can be ATP- or GTP-specific but all plant PCKs are ATP-specific. Amino acid sequence alignment of PCK enzymes shows that the nucleotide binding sites are somewhat conserved within each class with few exceptions that do not have any clear ATP- or GTP-specific binding motif. Although the active site residues are mostly conserved in all PCKs, not much significant sequence homology persists between ATP- and GTP-dependent PCK enzymes. There is only one planctomycetes PCK enzyme (from *Candidatus Kuenenia stuttgartiensis*) that shows sequence homology with both ATP- and GTP-dependent PCKs. Phylogenetic studies have been performed to understand the evolutionary relationship of various PCKs from different sources. Based on this study a flowchart of the evolution of PCK has been proposed.

Keywords: PEP-carboxykinase, evolution, phylogenetics

Introduction

Phosphoenolpyruvate carboxykinase (PCK), a carboxylase enzyme in nature, (EC 4.1.1.32 (GTP-dependent) or EC 1.1.49 (ATP-dependent)), is present in all known groups of living organisms. It catalyzes metal-nucleotide coupled reversible decarboxylation and phosphorylation between phosphoenolpyruvate (PEP) and oxaloacetate (OAA) depending on the system and the availability of the intermediates. In vertebrates, fungi, plants and in most bacteria, production of PEP from OAA by PCK is the key step for gluconeogenesis to produce glucose during fasting. In humans, increased gluconeogenesis is responsible for the high blood glucose level in non-insulin-dependent diabetes mellitus (NIDDM) patients. Otherwise, in healthy people cytosolic PCK enzyme is only present during glucose starvation; cytosolic PCK rapidly disappears on the resupplying of glucose due to hormonal control of the transcription of the cytosolic PCK-gene. In some bacteria such as *Anaerobiospirillum succiniciproducens* (Cotelesage et al. 2005), parasitic helminthes like *Ascaris suum* (Rohrer et al. 1986), nematodes such as *Haemonchus contortus* (Klein et al. 1992), PCK carry out the reverse reaction to produce OAA from PEP. In kinetoplastid parasites, such as *Trypanosoma cruzi* (Trapani et al. 2001) and all species of the genus *Leishmania*, this enzyme is very active even in the presence of high levels of carbohydrate, producing a mixture of CO₂, succinate and alanine as end products.

PCKs can be divided into two groups, based on its specificity towards the nucleotide substrate: ATP-dependent PCKs are mainly present in bacteria, yeast and plants, GTP-specific PCKs are mostly present in higher eukaryotes, most archaeons as well as in some bacteria (Fukuda et al. 2004). Fungal PCKs can be either ATP- or GTP-dependent. While there is significant sequence homology among each class, no statistically significant homology is found between the PCKs of the two classes. The structural studies from six different structures of PCKs, solved so far, demonstrated that the metal-binding and oxaloacetate binding active site residues are conserved in both ATP- and GTP-dependent PCKs (Holoak and Nowak, 2006; Cotelesage et al. 2005; Dunten et al. 2002; Trapani et al. 2001; Matte et al. 1996; Sudom et al. 2001; Leduc et al. 2005). The nucleotide binding motifs are also almost conserved within each category but are unique for each class with few exceptions. The most archaeon PCKs and *Giardia intestinalis* PCK have a unique GTP-binding region, compared to the GTP-binding motif present in other GTP-specific PCKs (Fukuda et al. 2004). It is interesting to know how and why two types of nucleotide specificities evolved.

Correspondence: Louis T.J. Delbaere, Department of Biochemistry, University of Saskatchewan, 107 Wiggins Road, Saskatoon, Saskatchewan, S7N 5E5, Canada. Tel: (306)966-4360; Fax: (306)966-4390; Email: louis.delbaere@usask.ca



Copyright in this article, its metadata, and any supplementary data is held by its author or authors. It is published under the Creative Commons Attribution By licence. For further information go to: <http://creativecommons.org/licenses/by/3.0/>.

The universal tree of evolution based on ribosomal RNA separates the three domains of archaea, bacteria and eukarya and places extreme thermophiles at the base of the bacteria (Brown et al. 2001). Another approach to make the universal tree based on protein sequence produces lots of mixing between domains. In spite of intermixing of the domains based on protein sequence homology, this method might be useful to understand the branching more accurately, as proteins control the cellular processes to maintain life. The three dimensional structure of a protein in its active form can provide important clues about how the protein performs its function. However structure determination of the protein is not always easy. Sequence alignment of similar proteins in distantly related organisms can also provide us with evolutionary relationships. Recent progress by genomic sequencing projects from a wide variety of species allows us to resolve more robust phylogenetic relationships among species.

Alignment of all PCK enzymes from the NCBI data bank shows some interesting results. In this paper we made the phylogenetic trees based on number of PCK proteins and genes sequences available in NCBI until 2007. To simplify the figure we only present a few selective species from all three domains of life. Extensive sequence alignment shows that the active site amino acid residues and metal binding sites (kinase 1a and kinase 2 regions) are almost but not completely conserved in PCKs of any origin and some have different nucleotide-binding sequences, which do not have any obvious specific nucleotide-binding motif for ATP or GTP. Based on protein sequence similarity and nucleotide specificities we propose a very simple evolutionary flowchart for PCK.

Phylogenetic Analysis

Initial sequence alignment of all the species containing PCK proteins and genes were performed using ClustalW (Higgins et al. 1994). Prodist and Dnadist software programs under PHYLIP (Felsenstein, 1989) were used on the aligned sequences which produce a distance matrix file. Performing Neighbor-Joining and UGMA (unweighted mean) on the distance matrix file and lastly by employing Drawtree we produced an unrooted tree diagram for all PCKs to give the evolutionary relationship between the species (Felsenstein, 1989). The phylogenetic tree (Fig. 1) based on the enzyme

sequences of PCKs, nicely divided ATP- and GTP-dependent PCKs into two regions. In each region of the tree, the PHYLIP program also grouped different species in the same branch or in proximity with very few exceptions, which are labeled. The archaeal *Aeropyrum pernix* (A-per) is placed close to other ATP-dependent bacterial PCKs, although it does not contain the conserved ATP-binding pockets (RX₅TR) of bacterial or eukaryotic PCKs. At the NCBI data bank *Aeropyrum pernix* (A-per) (NCBI accession no: **Q9YG68**) is the only ATP-dependent PCK from archaea. A-per PCK shows less than 10% sequence homology with all other archaeal PCKs, which are GTP-specific but it shares more than 25% sequence homology (by bl2seq) with other ATP-specific PCKs from bacterial or eukaryotic origin (Table 1) (Tatusova and Madden, 1999). Another bacterial PCK from the group planctomycetes *Candidatus Kuenenia stuttgartiensis* (C-stu) (NCBI accession no: **CAJ75104**) is placed close to A-per PCK. C-stu PCK also does not have a conserved ATP-binding pocket. These two PCKs might have a conserved arginine in the ATP-binding pocket as shown in the active site alignment table (Table 2). Both of these PCKs from A-per and C-stu have 51% similar (positive) and 31% identical amino acid sequences. C-stu PCK has 23%–36% amino acid sequence identity (by bl2seq) with various ATP- and GTP-specific PCKs from bacterial and eukaryotic species (Table 1). C-stu PCK does not have sequence similarity with other archaeal PCKs (except A-per) and with a few eukaryotic PCKs. A PCK from the archaon *Thermoplasma volcanium* (T-vol) (NCBI accession no: **P58306**), which contains a well conserved PCK-binding pocket like other GTP-dependent bacterial and eukaryotic species, also shows sequence similarity (23% identity, and 40% similarity) with C-stu PCK. The PCK from T-vol shows sequence similarity and identity with all other archaeal GTP-PCKs as well as all other GTP-dependent PCKs (Table 1). Another two GTP-dependent PCKs such as *Drosophila melanogaster* (NCBI accession no: **P20007**) and *Chasmagnathus granulate* (NCBI accession no: **AAL78163**) also have enzyme sequence homology with few other GTP-specific PCK enzymes similar to that of T-vol (Table 1).

In archea there are two distinct types of GTP-binding pocket; in one group they have a conserved GTP-binding pocket like other bacterial and eukaryotic PCKs (F/YXXXXF/Y), while in the other

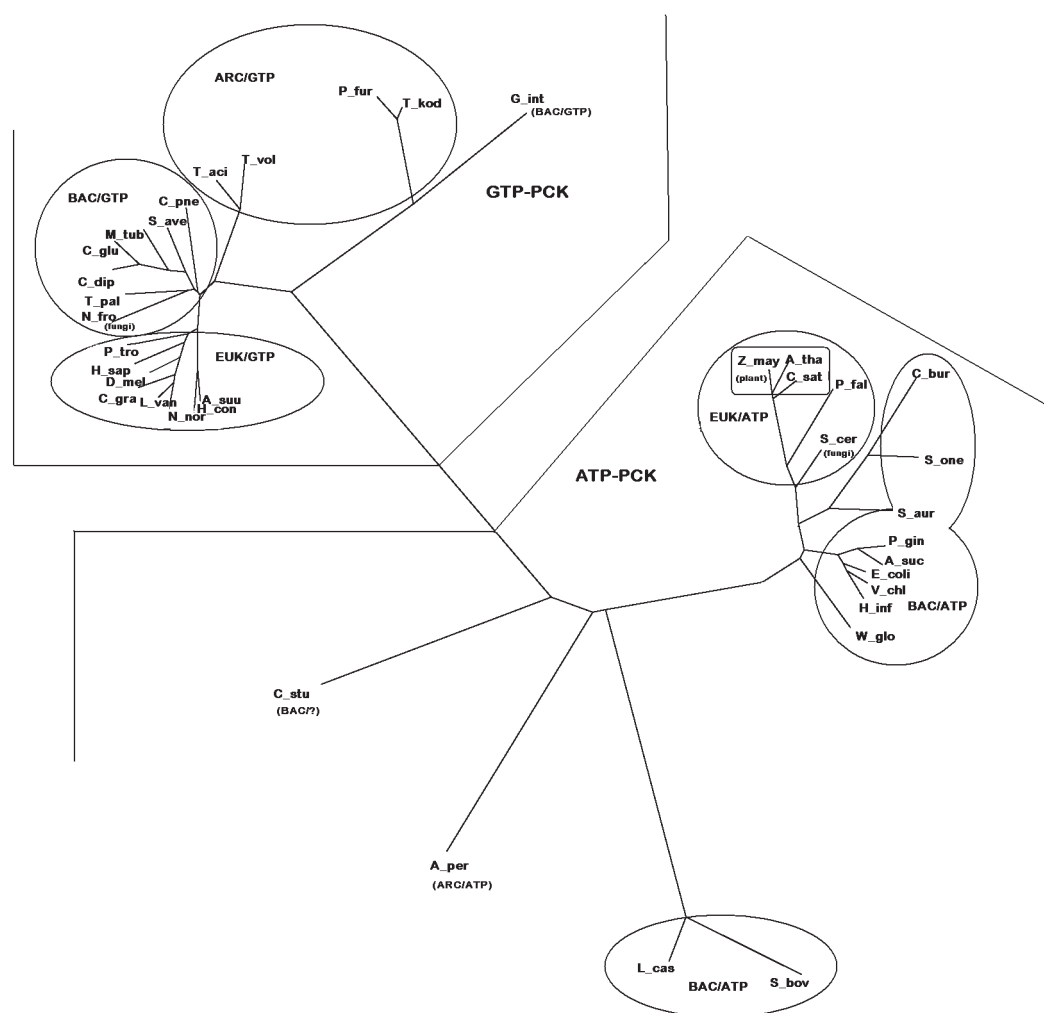


Figure 1. Phylogenetic tree based of amino acid sequences of few selected PCKs originated by the program PHYLIP, initially aligned with ClustalW.

few archaea and in *Giardia intestinalis* (NCBI accession no: **AAG47713**), PCKs have distinct GTP-binding residues (Table 2). These archaea and *Giardia intestinalis*, PCKs also show more than 23% sequence identity with T-vol PCK.

In other two PCKs from *Lactobacillus casei* (NCBI accession no: **ZIP_00386357**) and *Streptococcus bovis* (NCBI accession no: **BAE46992**) the amino acid sequence of the ATP-binding pocket is not totally conserved (Table 2).

PCKs from *Nephrops norvegicus* (NCBI accession no: **CAB65311**) and *Litopenaeus vanamei* (NCBI accession no: **CAB85964**) do not have the lysine for the kinase 1a/P-loop (XKT) which is conserved in all other PCKs, but the threonine is conserved (Table 2). Alignment (Table 2) shows that in *Nephrops norvegicus*, the conserved glycine is also absent. In *Lactobacillus casei* and

Streptococcus bovis, between lysine and threonine there is one serine residue. The conserved asparagine residue in this P-loop for GTP-specific PCKs is replaced by threonine in ATP-dependent PCKs. In the GTP-dependent PCK with a unique GTP-specific pocket, which is not conserved in other bacterial and eukaryotic PCKs, this asparagine is replaced by a serine residue. Similarly alignment studies also show that A-per PCK does not have the second lysine in its PCK-specific domain (XKK). The kinase 2 (XDD) domain is conserved in all PCKs where X in XDD is also glycine in most cases, with few exceptions where it is replaced by serine/histidine/glutamine. Initially we aligned all the PCK sequences available at NCBI and made the phylogenetic tree; later we chose a few from each domain group and the PCKs which have some importance (Table 2) to avoid overlap

Table 1. BL2seq shows sequence Identity (I) and homology (positive, P) between few selected PCK proteins from all domains of life.

No	Organism	<i>D_mel</i>		<i>C_gra</i>		<i>T_vol</i>		<i>A_per</i>		<i>C_stu</i>	
		I	P	I	P	I	P	I	P	I	P
1	<i>C_stu</i>	30	43	29	42	23	40	31	51	100	100
2	<i>D_mel</i> (GTP)	100	100	64	77	42	60	—	—	30	43
3	<i>C_gra</i> (GTP)	64	77	100	100	39	56	—	—	29	42
4	<i>T_vol</i> (GTP)	42	60	39	56	100	100	—	—	23	40
5	<i>A_per</i>	—	—	—	—	—	—	100	100	31	51
6	<i>A_suu</i> (GTP)	56	73	53	70	43	61	—	—	—	—
7	<i>H_con</i> (GTP)	55	71	53	71	41	61	—	—	—	—
8	<i>N_fro</i> (GTP)	50	68	45	64	40	58	—	—	—	—
9	<i>G_inf</i>	29	47	27	44	23	45	—	—	—	—
10	<i>P_fur</i>	29	48	30	48	31	49	—	—	—	—
11	<i>T_kod</i>	31	47	30	46	32	49	—	—	—	—
12	<i>S_ave</i> (GTP)	51	65	48	64	46	64	—	—	—	—
13	<i>T_aci</i> (GTP)	43	60	39	56	72	86	—	—	—	—
14	<i>P_tro</i> (GTP)	63	78	67	86	36	53	—	—	—	—
15	<i>H_sap</i> (GTP)	64	78	65	78	41	59	—	—	28	41
16	<i>L_van</i> (GTP)	67	78	75	84	40	57	—	—	28	41
17	<i>C_glu</i> (GTP)	48	65	44	62	44	61	—	—	28	43
18	<i>C_pne</i> (GTP)	51	66	47	65	46	62	—	—	28	45
19	<i>T_pal</i> (GTP)	52	67	47	62	45	63	—	—	30	41
20	<i>E_coli</i> (ATP)	—	—	—	—	—	—	33	44	30	43
21	<i>C_bur</i> (ATP)	—	—	—	—	—	—	27	41	26	41
22	<i>A_tha</i> (ATP)	—	—	—	—	—	—	27	44	32	41
23	<i>A_suc</i> (ATP)	—	—	—	—	—	—	22	44	24	39
24	<i>V_chl</i> (ATP)	—	—	—	—	—	—	25	46	25	49
25	<i>Z_may</i> (ATP)	—	—	—	—	—	—	25	44	29	45
26	<i>S_one</i> (ATP)	—	—	—	—	—	—	25	42	25	53
27	<i>H_inf</i> (ATP)	—	—	—	—	—	—	29	47	32	44
28	<i>S_aur</i> (ATP)	—	—	—	—	—	—	26	43	26	38
29	<i>W_glo</i> (ATP)	—	—	—	—	—	—	25	44	36	48
30	<i>S_cer</i> (ATP)	—	—	—	—	—	—	27	43	29	44
31	<i>P_gin</i> (ATP)	—	—	—	—	—	—	25	45	29	43
32	<i>C_sat</i> (ATP)	—	—	—	—	—	—	24	40	30	38
33	<i>L_cas</i>	—	—	—	—	—	—	25	43	24	40
34	<i>S_bov</i>	—	—	—	—	—	—	25	40	24	43

in the figure. The reactive cysteine residue (Cys 288 for human) which is conserved in most of the GTP-specific PCKs is replaced by serine in T-vol and *Thermoplasma acidophilum* (T-aci) PCKs (NCBI accession no: **Q9HLV2**) PCK.

The phylogenetic tree based on PCK genes (Fig. 2) is partly similar to the tree obtained from PCK enzymes with few exceptions. A-per and C-stu PCKs are in the same branch. But D-mel, a GTP-dependent PCK is placed more closely to ATP-dependent PCKs. Also L-cas and S-bov, two ATP-dependent PCKs, were nicely placed in the same branch but more closely to the GTP-specific PCKs. Otherwise the program grouped all other similar PCKs closer just like other phylogenetic trees based on PCK enzymes.

The score resulted from ClustalW (Table 3) for PCK genes, representing the identity between two, is very inconclusive. The value is 10 between A-per and C-stu otherwise all values are less than 10 for these two PCK genes compared to other PCK genes. For D-mel PCK, the gene score value is also very low for all other PCKs. Other two GTP-dependent PCKs C-gra and T-vol show some high scores only with other GTP-dependent PCKs.

Discussion Based on Phylogenetic Results and Sequence Alignments

The sequence alignment of all the PCKs demonstrate the presence of a PCK-specific domain (XKK), kinase 1a/P-loop (XKT), kinase 2 region

Table 2. Alignment of the active site residues including the PCK-specific region, Kinase 1a/P-loop, kinase-2 region and nucleotide binding sites of few selected PCKs. Active site residues are shown in red, mismatches are in blue.

Species	PCK-specific domain	Kinase-1a (P-loop)	Kinase-2	Nucleotide binding region
<i>A_suum</i>	YGGNSLLGKK	SACGKTN	VIGDD	WFRQSA-DHKFLWPGY
<i>H_con</i>	YGGNSLLGKK	SACGKTN	CVGDD	WFRDA-NNKFLWPGY
<i>N_nor</i>	YGGNSLLGKK	SAC-ATN	CVGDD	WFRKVK-KGRFIWPGF
<i>L_van</i>	YGGNTLLGKK	SACGSTN	CVGDD	WFRKDE-KARFIWPGF
<i>C_gra</i>	YGGNSLLGKK	SACRKTN	CVGDD	WFRKNE-KGRFKGPGF
<i>D_mel</i>	YGGNSLLGKK	SACGKTN	CVGDD	WFRKSA-EGKFMWPGY
<i>H_sap</i>	YGGNSLLGKK	SACGKTN	CVGDD	WFRKSA-EGKFMWPGY
<i>P_tro</i>	YGGNSLLGKK	SACGKTN	CVGDD	WFRDE-AGHFLWPGF
<i>C_dip</i>	YGGNAILAKK	SACGKTN	VVGDD	WFRRGD-DGRFLWPGF
<i>C_glu</i>	YGGNAILAKK	SACGKTN	VVGDD	WFRRGD-DGRFLWPGF
<i>M_tub</i>	YGGNALLGKK	SACGKTN	TLGDD	WFRRGD-DGRFLWPGF
<i>S_ave</i>	YGGNALLGKK	SACGKTN	TIGDD	WFRKND-EGKFVWPGF
<i>N_fro</i>	YGGNALLGKK	SACGKTN	CVGDD	WFRKD--NGRFLWPGY
<i>T_pal</i>	YGGNALLGKK	SACGKTN	TVGDD	WFRKDA-EGNFLWPGY
<i>C_pne</i>	YGGNALLGKK	SACGKTN	CIGDD	WFRKNN-QGEFLWPGF
<i>T_aci</i>	YGGNALLSKK	SASGKTN	LLSDD	WFRRRQ-DGSFIWPGF
<i>T_vol</i>	YGGNALLSKK	SASGKTN	LISDD	WFRRA-DGTFIWPGF
<i>P_fur</i>	YGGNVIGLKK	SMCGKTS	IVGDD	YFLRE--NGQWLNEKLD
<i>T_kod</i>	YGGNTIGLKK	SMCGKTS	IVGDD	YFLRE--NGVWLNHKLD
<i>G_int</i>	YAGNALACKK	SACGKTS	IIGDD	YFLKDPPTGEYLSKLD
<i>Z_may</i>	YAGE---MKK	SGTGKTT	LIGDD	YGVG--KRIRLPYTR
<i>A_tha</i>	YAGE---MKK	SGTGKTT	LIGDD	YGTG--SRIKLAYTR
<i>C_sat</i>	YAGE---MKK	SGTGKTT	LIGDD	YGS--NRIKLAYTR
<i>S_cer</i>	YAGE---MKK	SGTGKTT	LIGDD	YVSGG-KRCPLKYTR
<i>B_mar</i>	YAGE---MKK	SGTGKTT	LIGDD	YGVG--SRFKLYTR
<i>P_fal</i>	YAGE---MKK	SGTGKTT	LIGDD	YGSNGIRIPLKYTR
<i>T_cru</i>	YAGE---MKK	SGTGKTT	LIGDD	RADRGAKRMPLRVTR
<i>E_coli</i>	YGGE---MKK	SGTGKTT	LIGDD	WNGTG-KRISIKDTR
<i>V_chl</i>	YGGE---MKK	SGTGKTT	LIGDD	WNGSG-KRISIKDTR
<i>H_inf</i>	YGGE---MKK	SGTGKTT	LIGDD	WNGTG-KRISIKDTR
<i>P_gin</i>	YGGE---MKK	SGTGKTT	LIGDD	WNGTG-KRISIKDTR
<i>W_glo</i>	YGGE---IKK	SGTGKTT	LIGDD	WNGKR-ERYSLEYTR
<i>C_bur</i>	YAGE---MKK	SGTGKTT	LIGDD	WTGGAERFSIPTTR
<i>S_one</i>	YAGE---MKK	SGTGKTT	LIGDD	WTGGIGKRFDIPTTR
<i>S_ace</i>	YAGE---MKK	SGTGKTT	LIGDD	WTGGKYRRISLHYTR
<i>S_bov</i>	YFGE---LKK	SGSGKST	VLHDD	YLG---QSIPKEVTL
<i>L_cas</i>	YFGE---LKK	SGSGKST	VLHDD	YNG---QNVKPADTL
<i>C_stu</i>	YYGE---SKK	SGTGKTT	ILQDD	GKKVTK-RKVKRVEI
<i>A_per</i>	YYGE---LKM	SGTGKTT	VMQDD	HAKD---RKIPPELS

(GDD) and nucleotide-specific pockets (Table 2). The catalytic site consists of the PCK-specific domain, the kinase 1a and kinase 2 regions which are almost conserved in ATP- and GTP-dependent PCKs from all organisms with very few exceptions as shown in the sequence alignment (Table 2). Sequence alignment of all PCKs available at NCBI shows us a few interesting features of some PCKs. The structural analysis from X-ray diffraction studies shows that the second lysine in the PCK-specific domain binds with divalent metal ion in the second

metal binding site. The absence of the conserved second lysine residue in A-per PCK might eliminate the second metal ion binding site.

The lysine residue in GKT/P-loop (Lys 254 in *E. coli*) is interacting with both the β - and γ -phosphoryl groups of ATP. The positive charge of lysine helps in stabilizing the phosphoryl group during transfer (Matte et al. 2001). The absence of this lysine in N-nor and L-van might affect the reactivity of these enzymes. In this P-loop there is a reactive cysteine which is mostly conserved in

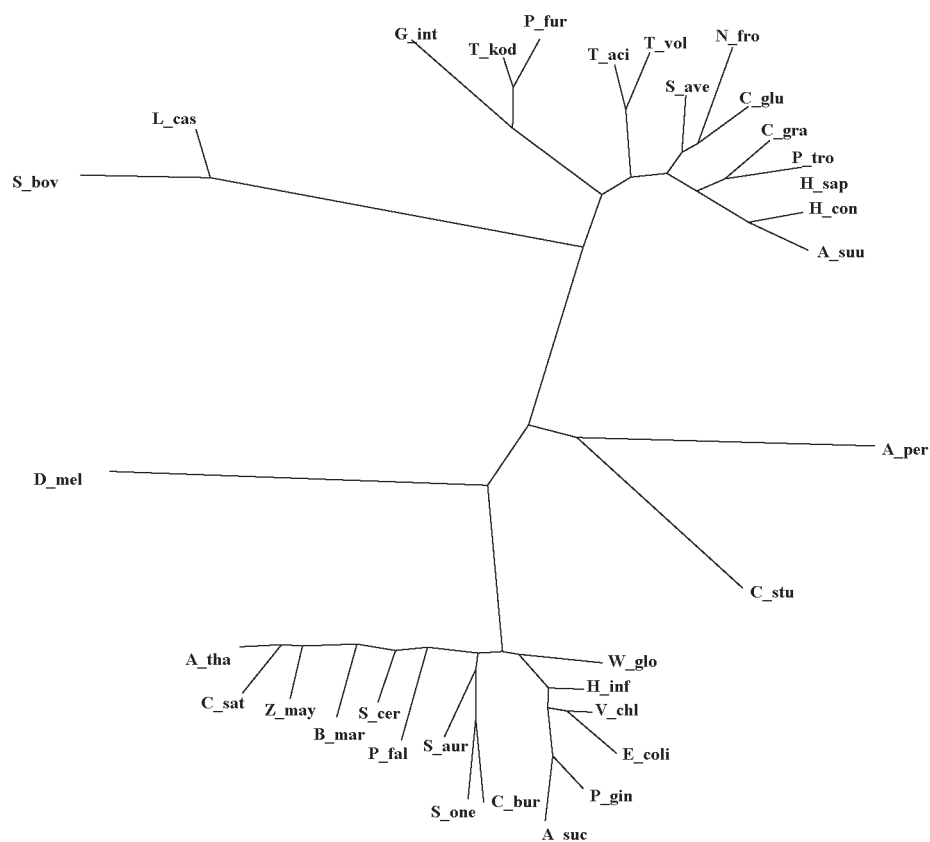


Figure 2. Phylogenetic tree based on DNA sequences of few selected PCKs originated by the program PHYLIP, initially aligned with ClustalW.

all GTP-dependent PCKs. Reactive cystine, however, is replaced by serine in T-vol and T-aci which are two archeal PCKs with well defined GTP binding motif. It might be interesting to know how the absence of this cysteine can affect the stability of the P-loop in the GTP-dependent PCKs in T-vol and T-aci.

Although the phylogenetic tree (Fig. 1) constructed from selected important PCK enzymes is very similar to the tree made by Fukuda et al. where the emphasis was given on the position of archeal PCKs (Fukuda et al. 2004), the current phylogenetic tree is more extensive and elaborate. We have included few more PCKs from other species to explain the evolutionary origin of PCK. The tree reported here also shows that ATP- and GTP-dependent PCKs are present in all three major domains of life. All mammalian PCKs are GTP dependent, plant PCKs are ATP dependent and bacterial and fungal PCKs can be ATP- or GTP-dependent. It is interesting to note that parasitic nematodes, like *Ascaris suum* and *Haemonchus contortus* where PCKs preferentially carboxylate

PEP to OAA, have been placed in a separate branch in the tree.

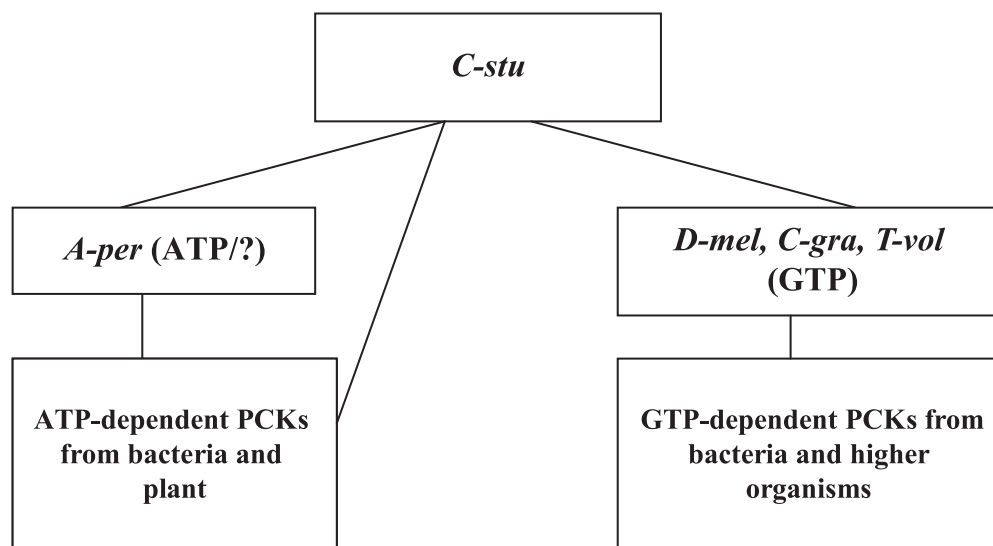
Many scientists hypothesize that the Archaea are the closest modern relatives of earth's first living cells. They are called universal ancestors from which all other life is believed to have evolved (Woese, 2000). All archaeal PCKs are GTP-dependent except for *Aeropyrum pernix* (A-per). From amino acid sequence similarity, A-per PCK is closer to the ATP-binding PCKs (Table 1). It does not have any sequence similarity with other archaeal PCKs, which justifies its position in the phylogenetic tree (Fig. 1).

The most interesting PCK is bacterial planctomycetes *Candidatus Kuenenia stuttgartiensis* (C-stu) which placed close to A-per PCK. ATP- and GTP-specific PCKs have less than 10% sequence similarities, although C-stu PCK shows a very unique feature (Table 1). C-stu PCK has 23% – 36% amino acid sequence identity (by blast2) with various ATP- and GTP-specific PCKs from bacterial and eukaryotic origin (Table 1). This is the only PCK, so far, which has sequence similarity

Table 3. Score values from ClustalW between PCK genes from all domains of life.

Organism	<i>D_mel</i>	<i>C_gra</i>	<i>T_vol</i>	<i>A_per</i>	<i>C_stu</i>
<i>D_mel</i>	100	0	3	1	3
<i>C_gra</i>	0	100	30	1	1
<i>T_vol</i>	3	30	100	2	1
<i>A_per</i>	1	1	2	100	10
<i>C_stu</i>	3	1	1	10	100
<i>A_suu</i>	0	46	30	1	1
<i>H_con</i>	1	48	30	1	1
<i>N_fro</i>	2	37	33	1	2
<i>H_sap</i>	1	57	30	2	1
<i>P_fal</i>	2	1	3	1	1
<i>G_inf</i>	1	3	7	1	1
<i>P_fur</i>	1	4	34	4	2
<i>E_coli</i>	2	1	1	2	1
<i>T_kod</i>	1	30	6	1	1
<i>C_bur</i>	1	1	1	3	3
<i>A_tha</i>	2	1	0	1	1
<i>C_glu</i>	1	44	37	1	1
<i>A_suc</i>	2	1	1	1	3
<i>V_chl</i>	1	1	0	2	3
<i>Z_may</i>	1	1	1	2	2
<i>S_one</i>	1	1	1	2	1
<i>H_inf</i>	4	1	1	1	3
<i>S_aur</i>	5	1	2	2	1
<i>S_ave</i>	3	58	35	1	1
<i>T_aci</i>	1	32	69	1	2
<i>W_glo</i>	1	1	2	1	1
<i>S_cer</i>	1	1	0	2	2
<i>P_tro</i>	0	51	12	1	1
<i>B_mar</i>	0	2	1	1	5
<i>P_gin</i>	1	1	2	1	3
<i>C_sat</i>	0	1	1	3	4
<i>L_cas</i>	1	1	1	1	1
<i>S_bov</i>	1	1	1	1	4

with both ATP- or GTP-binding PCKs but does not have any well defined nucleotide-binding motif. But still there is another valid point. C-stu PCK does not have sequence similarity with most other archaeal and few other GTP-dependent PCKs with only one exception (T-vol). The sequence homology between the archaeal PCK from T-vol and C-stu PCK might gave us the missing link. Sequence homology analysis also shows that other two GTP-specific PCKs, from D-mel and C-gra, have similar sequence homology with other GTP-specific PCKs and also with C-stu PCK but with any ATP-dependent PCKs. We can place C-stu PCK, which has sequence similarity with both ATP- and GTP-dependent PCKs) and does not have a very well recognized nucleotide-binding pocket, at the root (Table 2). Does it mean initially that PCKs did not have any ATP- or GTP-specificities, which have been acquired gradually with evolution? The position of plantomycetes is also very controversial in the phylogenetic tree. One recent analysis placed planctomycetes at the deep branching position near the root (Fuerst, 1995). Based on this concept and our homology studies we propose a very simplified evolution flowchart for PCK (Table 4). It would be interesting to know the active site residues and the tertiary structures of C-stu, T-vol and A-per PCKs to complement this evolutionary perspective.

Table 4. PCK evolution flowchart.

Acknowledgments

LTJD is a Canada Research Chair in Structural Biochemistry. This research was funded by an operating grant from the Canadian Institutes of Health Research to LTJD.

References

- Brown, J.R., Douady, C.J., Italia, M.J., Marshall, W.E., Stanhope, M.J. et al. 2001. Universal trees based large combined protein sequence data sets. *Nature Genetics*, 28:281–5.
- Cotelesage, J.J., Prasad, L., Zeikus, J.G., Laivenieks, M., Delbaere, L.T. et al. 2005. Crystal structure of *Anaerobiospirillum succiniciproducens* PEP carboxykinase reveals an important active site loop. *Int. J. Biochem. Cell Biol.*, 37:1829–37.
- Dunten, P., Belunis, C., Crowther, R., Hollfelder, K., Kammlott, U., Levin, W., Michel, H., Ramsey, G.B., Swain, A., Weber, D., Wertheimer, S.J. et al. 2002. Crystal structure of human cytosolic phosphoenolpyruvate carboxykinase reveals a new GTP-binding site. *J. Mol. Biol.*, 316:257–64.
- Felsenstein, J. 1989. PHYLIP-Phylogeny Inference Package (Version 3.2). *Cladistics*, 5:164–6.
- Fuerst, J.A. 1995. The Planctomycetes: Emerging Models for Microbial Ecology; Evolution and Cell Biology. *Microbiology*, 141:1493–1506.
- Fukuda, W., Fukui, T., Atomi, H., Imanaka, T. et al. 2004. First characterization of an archaeal GTP-dependent phosphoenolpyruvate carboxykinase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bact.*, 186:4620–7.
- Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G., Gibson, T.J. et al. 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.*, 22:4673–80.
- Holyoak, T., Sullivan, S.M., Nowak, T. et al. 2006. Structural insights into the mechanism of PEPCK catalysis. *Biochemistry*, 45:8254–63.
- Klein, R.D., Winterrowd, C.A., Hatzenbuehler, N.T., Shea, M.H., Favreau, M.A., Nulf, Sc., Geary, T.G. et al. 1992. Cloning of a cDNA encoding phosphoenolpyruvate carboxykinase from *Haemonchus contortus*. *Mol. Biochem. Para.*, 50:285–94.
- Leduc, Y.A., Prasad, L., Laivenieks, M., Zeikus, J.G., Delbaere, L.T. et al. 2005. Structure of PEP carboxykinase from the succinate-producing *Actinobacillus succinogenes*: a new conserved active-site motif. *Acta Crystallogr D Biol. Crystallogr.*, 61:903–12.
- Matte, A., Goldie, H., Sweet, R.M., Delbaere, L.T. et al. 1996. Crystal structure of *Escherichia coli* phosphoenolpyruvate carboxykinase: a new structural family with the P-loop nucleoside triphosphate hydrolase fold. *J. Mol. Biol.*, 256:126–143.
- Matte, A., Tari, L.W., Goldie, H., Delbaere, L.T. et al. 1997. Structure and mechanism of Phosphoenolpyruvate Carboxykinase. *J. Biol. Chem.*, 272:8105–8.
- Rohrer, S.P., Saz, H.J. and Nowak, T. 1986. Purification and characterization of Phosphoenolpyruvate carboxykinase from the parasitic helminth *Ascaris suum*. *J. Biol. Chem.*, 261:13049–55.
- Sudom, A.M., Prasad, L., Goldie, H., Delbaere, L.T. et al. 2001. The phosphoryl-transfer mechanism of *Escherichia coli* phosphoenolpyruvate carboxykinase from the use of AIF(3). *J. Mol. Biol.*, 314:83–92.
- Tatusova, T.A. and Madden, T.L. 1999. Blast 2 sequences—a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett.*, 174:247–50.
- Trapani, S., Linss, J., Goldenberg, S., Fischer, H., Craievich, A.F. and Oliva, G. 2001. Crystal structure of the dimeric phosphoenolpyruvate carboxykinase (PEPCK) from *Trypanosoma cruzi* at 2 Å resolution. *J. Mol. Biol.*, 313:1059–72.
- Woese, C.R. 2000. Interpreting the universal phylogenetic tree. *Proc. Natl. Acad. Sci. U.S.A.*, 97:8392–6.