

# Eukaryotic protein kinases and choline kinases share a common ancestor related to ancient glutaminyl-tRNA synthetase



KINEXUS

Shenshen Lai<sup>1</sup> and Steven Pelech<sup>1, 2</sup>

<sup>1</sup> Department of Medicine + Brain Research Centre, University of British Columbia, Vancouver, BC, Canada, <sup>2</sup> Kinexus Bioinformatics, Vancouver, BC, Canada

# Abstract

The universal existence of reversible phosphorylation and dephosphorylation of proteins catalyzed by protein kinases and protein phosphatases in eukaryotes supports a crucial role of eukaryotic protein kinases (ePKs) in the emergence of nucleated cells in the third superkingdom of life. Choline kinases (ChKs) could also be important in the early evolution of eukaryotes, because of their function in the biosynthesis of phosphatidylcholine, which is unique to eukarvotic membranes. However, the genomic origins of ePKs and ChKs are unclear. The high degeneracy of protein sequences and the broad expansion of protein kinase families in eukaryotic organisms have made this fundamental question difficult to answer. To explore the ancestry of these important protein families, we defined a consensus amino acid sequence for the catalytic domain of human protein kinases and used it for genome-wide BLAST searches from multiple species. Our results revealed a surprisingly close relationship in the primary sequences of ePKs and ChKs with glutaminyl-tRNA synthetase (GInRS). The hypothesis that ePKs and ChKs share a common ancestor related to GInRS was supported by further sequence analysis and 3D structure comparisons. This discovery provides considerable insight into the evolution of ePKs and ChKs in ancient eukarvotes. The molecular paleontology approach undertaken in this study also provides a novel and broadly applicable method to generally investigate the origins of large gene families.

# Introduction

In eukaryotic organisms, reversible protein phosphorylation provides an efficient means to regulate almost every aspect of physiological activities including metabolism, transcription, DNA repair, cell growth and division, and apoptosis. Eukaryotic protein kinases (ePKs) comprise a broadly expanded family of proteins that feature a conserved catalytic domain. Within the catalytic domain, extremely conserved subdomains involved in ATP binding and the phosphotransfer reaction have been well characterized. In humans, 516 protein kinases genes have been identified, and they account for ~1.7% of the 23,000 genes encoded by the human genome. Choline kinases (ChKs) catalyze the phosphorylation of choline, which is the first step for synthesis of phosphatidylcholine (PC), the predominant phospholipid species of eukaryotic membranes. PC can also be cleaved by phospholipases to generate important second messengers such as fatty acids and diacylglycerol, which are activators for protein kinases in signal transduction pathways. The threedimensional structure of ChK has been reported to share strikingly high similarity with ePKs, with little primary sequence similarity identified. The important functions of ePKs and ChKs implicate fundamental roles for these enzymes in the evolution of eukaryotic life. The ancestry of these proteins would provide deeper understanding in the development of eukaryotic intracellular signal transduction systems

# **Results & Discussion**

## 1. Consensus Sequence of Ser/Thr Kinase Catalytic Domain

The amino acid sequences of 393 human serine/threonine kinase (STK) catalytic domains were collected and precisely aligned (Table 1). The frequency of each of the 20 common amino acids at each position was calculated to define an optimal consensus sequence of 247 amino acids. This STK sequence contains 12 catalytic subdomains that includes 30 highly conserved amino acids, and 10 gaps that provide variable regions for further specialization of individual kinases.

## Table 1. Alignment of human STK catalytic domains (partial)

## 2. Proteins Most Closely Related to Protein Kinase

The STK consensus sequence was employed in BLAST queries conducted with six well studied organisms. We were able to identify 11 candidates that were present among the top hits (non-protein kinases with E-values less than 1.000 and alignment length longer than 35 amino acids) in more than one species (Table 2). ChK received the highest average score of all the candidate proteins Surprisingly, two of the three class-Laminoacyl-tRNA synthetases were identified with high similarities with the STK consensus sequence. One of them, the glutaminyl-tRNA synthetase (GInRS), showed conservation in most of the well characterized kinase catalytic subdomains, which was not seen in any other candidate besides ChK. This result revealed a close evolutionary relationship between GInRS\_ePK and ChK

#### Table 2. Candidate proteins from BLAST search results

	Arabidopsis	E. coli	Yeast	C. elegans	Drosophila	Human	Average Score
Choline/ethanolamine kinase				18.7		26.4	22.6
Phenylalanine-4-hydroxylase				21.9	21.5		21.7
Dynein heavy chain				21.2		20.5	20.9
Glutaminyl-tRNA synthetase		21.5	19.8				20.7
Actin-related protein	20.5		20.8				20.7
Peptidyl-tRNA hydrolase		21.5		19.4			20.5
Alanyl-tRNA synthetase	22.2	18.4					20.3
Cullin protein				20.8	19.4	20.8	20.3
3',5'-cyclic phosphodiesterase				19.8	19.1	21.5	20.1
DNA topoisomerase		18.7			20.5		19.6
Nucleolar GTP-binding protein				19.1	19.1		19.1

# 3. Amino Acid Sequence Alignment of GInRS, ePK and ChK

To further investigate the sequence similarity of GInRS, ePK and ChK, we took a similar strategy and generated consensus sequences for both GInRS and ChK using data from various species. The three consensus sequences were manually aligned with each other (Table 3). From the alignment, we noticed the GInRS and STK consensus sequences shared the highest identity of 24%. including 18 of the 30 key residues for kinase catalytic activity, followed by a 20% identity between GInRS and ChK consensus sequences. The identity of STK and ChK consensus sequences was lower but comparable to that of two random human protein kinases with each other.

## Table 3. Pairwise alignment of consensus sequences of GInRS, STK and ChK

оная 180-16-4 У У У У Раз С С Р Г Р В П С Т Т Р В Т С Т Т В С В О Р Г Т Х Г А С А О Р К С С В С С В С В С В С В С В С В С В С	
OK OLOPKLYNYFPNORLEEFIPERTLETEDL OPEISXIAEKMAOFHX.LKMPLSKEP	
DHE GVGVVVTPE.GIEE	
GHRS TESEMEDERGEALKF HKPGENYK TOGYVYTPNTME ELKONLEITGGYHT	
	DLEPENT
OK YLELYKELXFXREA-KIKKLLKILSYDLKEEIDN-LHKLLDXLXSPVVFCHN	DLOEON
Ques TERLWERL HOEALETHEPOENVETOUVVVTPN. THELL KOHLEITOUVHTH	FPPEPNO
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
U LUI U UU U U UU U U U U U U U U U U U	FEPYKVT
STA LLDEDXONINEADFOLANGLEXXXXLTTFCOTFEYMAPEVLLOXXQYSKAVOINSL	VILVELL
DH LLEED	svidge.
OHR ILHIGHAKAININFOLAKAN - OGICYLAFDOTNPE KEEEEYITSIKONY	EWLOFEP
GHE VARDYFDOLVEWAVELIKKOLAVVCHOTPEELK = EOKATLENKL = FAPHH	
STN TOKPPPKOESELELLXKILXXPXLEIPXXLSPEAKOLLAKLIXKOPXKR	

GHAS YRYTYABDYPESLLLPEDMRKGRFARGKATIMKLDMESDAT

Ibit
Bayes
Bayes

0000
00000
00000

0000
00000
00000

0000
00000
00000

0000
00000
00000

0000
00000
00000

0000
00000
00000

0000
00000
00000

0000
00000
00000

0000
00000
00000

0000
00000
00000

0000
00000
00000

0000
00000
00000

00000
00000
00000

00000
00000
00000

00000
00000
00000

00000
00000
00000

00000
00000
00000

00000
00000
00000

00000
00000
00000

00000
00000
00000

00000
00000
00000

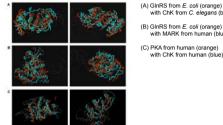
00000
00000
000000

000000

#### 4. Three-dimensional Structures of GInRS, ePK and ChK

Three-dimensional structure comparison tools from the RCSB Protein DataBank (http://www.pdb.org) were employed to align the tertiary structures of GInRS, ePK and ChK. There was no available 3D structure of GInRS from eukaryotes. The absence of N-terminal domain in E. coli GInRS resulted in a slightly lower alignment score from GInRS and STK structures. However, similar structures like the alpha helix important for ATP binding and phosphate transfer in ePKs and ChKs were observed in GlnRS (Figure 1). The structural comparison score of GInRS and ChK was similar with that of ePK and ChK (Figure 2).

#### Figure 1. 3D structure comparison of GInRS, ePK and ChK



with ChK from C. elegans (blue); (B) GInRS from E. coli (orange) with MARK from human (blue);

(C) PKA from human (orange) with ChK from human (blue)

Figure 2. Summary of sequence and structure comparison of GInRS, ePK and ChK



# 5. The Most Ancient Human Protein Kinases

We expected that the most ancient human protein kinases should be more closely related to certain PKs in primary structure, and also conserved across species. The catalytic domains of all the human protein kinases were aligned with GInRS consensus sequence by BLAST. Seven of the top 20 hits including AMPKs and some RSK family members, also appear among the top 100 hits in the BLAST search against STK consensus sequence (Table 4). Among these seven candidates, AMPKs showed the highest conservation scores, indicating the possibility that AMP-dependent protein kinases are most closely related to the ancient protein kinases.

#### Table 4. Most probable ancient protein kinases

Name	Uniprot ID	Identity (%) to PK consensus sequence	Conservation score	
AMPK2	P54646	42.25	715	
p70S6K1	P23443	45.90	570	
AMPK1	Q13131	42.25	527	
KAMK2D	Q13557	42.21	523	
MAST2	Q6P0Q8	37.68	385	
RSK4	Q9UK32	45.08	347	
p70S6K2	Q9UBS0	44.40	264	

# Conclusion

Eukaryotic protein kinases and choline kinases share a common ancestor related to glutaminvl-tRNA synthetase, which is found in bacteria. Our findings indicate that duplications and mutations of the GInRS gene to create ePKs and ChKs may have been a key factor in the successful of emergence of ancient eukarvotic cells from bacterial colonies.