The Origin and Evolution of Eukaryotic Protein Kinases

Asako Koike1,2  kenta@ims.u-tokyo.ac.jp
akoike@ims.u-tokyo.ac.jp

Kenta Nakai1  knakai@ims.u-tokyo.ac.jp

Toshihisa Takagi1  takagi@ims.u-tokyo.ac.jp

1 Institute of Medical Science, University of Tokyo, 4-6-1, Shirokane-dai, Minato-ku
Tokyo108-8639, Japan
2 Central Research Laboratory, Hitachi, Ltd., Higashi-koigakubo, Kokubunji 185-8601, Japan

Keywords: protein kinase, evolution, functional conservation, cell signaling

1 Introduction

Cellular signaling pathways in eukaryotes are indispensable to the process of tissue growth, cell differentiation, and rapid response to environmental changes. Many of signals are transmitted by the phosphorylation and dephosphorylation of proteins through the mediation of protein kinases and phosphatases [1]. The active sites of protein kinases are well conserved as protein kinase domains. Protein kinases of eukaryotes are subdivided into a unique eukaryotic protein kinase (hereafter, abbreviated as "eukaryotic protein kinase" or "ePK"), a histidine protein kinase, and atypical protein kinase. Most protein kinases in eukaryotes are ePKs. Because ePKs play a crucial role in cell signaling pathways in eukaryotes, it is important to know the origin and evolution of ePKs to clarify the evolution of eukaryotic cell signaling. As ePKs, the Pkn family has been identified in bacteria. Because Pkn protein kinases exist only in simple life-cycle bacteria and do not exist in archaea and complex life-cycle bacteria, horizontal gene transfer from eukaryotes to bacteria has been inferred. In this study, an evolutionary model of ePKs was developed by using their sequence similarity and protein structural information. Since the evolutionary distance between eukaryotes and prokaryotes is too far to enable clarifying ePK expansion based only on the ePK sequence similarity, the similarity between proteins coded in the neighborhood of ePK genes on the genomes was also analyzed. Further, by extracting all ePKs in major organisms and comparing them, the functional conservation between major organisms and evolution of signaling pathways in eukaryotes were discussed.

2 Methods and Materials

The amino-acid sequences and/or genome sequences of prokaryotes, S. pombe, S. cerevisiae, C. elegans, D. melanogaster, H. sapiens, and A. thaliana were obtained from NCBI ftp site, NCBI RefSeq, nr-aa, Ensemble, WormBase, Flybase, and TAIR database. The ePKs of Giardia lamblia, Trichomonas vaginalis, and Euglena gracilis were found by BLAST from a non-redundant database at NCBI by using the ePK domain sequences of S. cerevisiae as queries. The prokaryotes used in this study were 9 species in Archaea and 22 species in Bacteria. The ePK domains of eukaryotes were extracted by Pfam. The distant homologs of the ePKs in the prokaryotes were identified by PSI-BLAST from a non-redundant database at NCBI by using the ePK domain sequences of S. cerevisiae as queries. The prokaryotes used in this study were 9 species in Archaea and 22 species in Bacteria. The ePK domains of eukaryotes were extracted by Pfam. The distant homologs of the ePKs in the prokaryotes were identified by PSI-BLAST. In our search, we used prokaryotic amino acid sequences and S. cerevisiae ePK domain sequences as subjects (a dataset) and S. cerevisiae ePK domain sequences of major groups as queries. And then, the prokaryotic amino acid sequences of the results were used as query. The threshold value for inclusion in the position-specific matrix used for iterations was set to 1 and the final expectation value and score were set to $10^{-4}$ and 75, respectively. Then the structural similarities between ePKs and distant homologs of ePKs were investigated by using 3D-PSSM2.6 and FUGUE. Phylogenetic trees from the distance matrix were drawn by Phylip. The kinase domain sequences were aligned by ClustalW ver.1.7 and their phylogenetic trees were drawn by "njplot" of ClustalX ver.1.81. Some sequence sets with weak similarity were aligned by parallel PRRN.
The similarities between the proteins coded in the neighborhood of the ePK genes on the genomes were investigated as follows. First, similarities between the proteins of 15 genes on both sides of the ePK genes on the genome were detected by BLAST, that is, a BLAST search was executed for all pairs of ePK genes with 31 proteins. The threshold for a statistically meaningful E value was set to 0.01. Even if the threshold were set to 0.001, the results would not be very different. When there are four homologous genes: a1, a2, b1, and b2, where a1 and a2 are on the same chromosome and b1 and b2 are on the other chromosome, two types of gene duplication are considered. One is “successive duplication”: a1→b1, a1→a2, b1→b2 and the other is “simultaneous gene duplication after gene duplication”: (a1→a2)→(b1, b2). (b1 and b2 are duplicated by a1 and a2, respectively.) Here “→” indicates gene duplication, and S(a1, a2) indicates the translated protein alignment score between a1 and a2. To distinguish between “successive duplication” and “simultaneous gene duplication after gene duplication”, the alignment scores, S, of these duplicated genes were compared. In the case of “successive duplication”, the protein alignment score was expected to be S(a1,a2) > S(a2, b2), while in the case of “simultaneous gene duplication after gene duplication”, it was expected to be S(a1,a2) < S(a2,b2), if the duplication occurred in full-length genes. The former cases were removed from further analysis, because we investigated the similarity of the proteins to look for homologous genes because of duplication of the genome region including ePK.

3 Results and Discussions

Among the ABC1, RIO1, O-sialoglycoprotein, and AQ578 families, which have low sequence similarity with ePKs, the ABC1 family was found to be the most plausible origin of ePKs in eukaryotes by structural information. However, since there is no steadfast evidence either for or against the horizontal gene transfer in the bacterial Pkn family, it remains unclear which family ABC1- or Pkn- is the origin of ePKs in eukaryotes. Since most sequences of both families have a high similarity to the sequence of the AGC/CaMK group, the origin of ePKs is most likely to have first evolved into the AGC and/or CaMK groups and then ePKs expanded as a result of gene duplication regardless of which family is the origin of ePKs in eukaryotes. The existence of cAMP-dependent kinases of protozoa and their similarity to those of eukaryotes, as well as a weak similarity of cyclin-dependent-like kinases of protozoa to those of eukaryotes support this idea. The phylogenetic trees constructed based on the number of similar-sequence proteins in the neighborhood of the ePK genes on the genome are similar to the phylogenetic trees constructed based on multiple alignments of the ePK domains, which also supports the divergence of ePKs as a result of gene duplication. These results indicate that even if the Pkn-family genes are genes horizontally transferred from eukaryotes, there is a high possibility that this horizontal transfer occurred before the divergence of protozoa such as G. lamblia and early eukaryotes. Further, the similarity of ePKs and that of the proteins and genomes in the neighborhood of ePKs on the genomes indicate that the MEK and MEKK families diverged from the MAPK family, and that the MEKKK(STE20) and a scaffold protein (STE5) diverged from the MEK family and the neighborhood genome/gene of MAPK(KSS1) genes, respectively, and that SHO1, which is not an ePK but upstream component of MEKK (STE11), diverged from the neighborhood genome/gene of MEKK family. The correlation coefficients of the phylogenetic trees of MAPKs and MEKs, MEKs and MEKKs, and MAPKs and MEKKs were quite high in the pathways without cross-talk, which indicates that MAP kinase pathways have co-evolved by interacting with each other. The high correlation coefficient will be useful to predict in predicting of unknown pathway components.

Finally, the classification and comparison of ePK families in ePKs in eukaryotes showed the families involved in specific functions such as immune system, nervous system, and apoptosis were expanded in higher organisms. We concluded that the variety of signaling pathways in higher eukaryotes were mainly due to the variety of upstream components of the signaling pathways including tyrosine kinases and the complexity of MAP kinase pathways with cross-talk.

References