Structural constraints and functional divergences in CASK evolution

Leslie LaConte and Konark Mukherjee

Abstract

CASK (Ca\(^{2+}\)/calmodulin-activated serine kinase) is a synaptic protein that interacts with the cytosolic tail of adhesion molecules such as neurexins, syncam and syndecans. It belongs to the MAGUK (membrane-associated guanylate kinase) family of scaffolding proteins which are known to decorate cell–cell junctions. CASK is an essential gene in mammals, critical for neurodevelopment. Mutations in the CASK gene in humans result in phenotypes that range from intellectual disability to lethality. Despite its importance, CASK has a single genetic isoform located in the short arm of the X chromosome near an evolutionary breakpoint. Surprisingly, CASK is a non-essential gene in invertebrates and displays functional divergence. In the present article, we describe the phylogenetic differences in existing CASK orthologues. The CASK gene has undergone a huge expansion in size (∼55-fold). Almost all of this expansion is a direct result of an increase in the size of the introns. The coding region of CASK orthologues, and hence the protein, exhibit a high degree of evolutionary conservation. Within the protein, domain arrangement is completely conserved and substitution rates are higher in the connecting loop regions [L27 (Lin2, Lin7)] than within the domain. Our analyses of single residue substitutions and genotype–phenotype relationships suggest that, other than intronic expansion, the dramatic functional changes of CASK are driven by subtle (non-radical) primary structure changes within the CASK protein and concomitant changes in its protein interactors.

Introduction

CASK (Ca\(^{2+}\)/calmodulin-activated serine kinase), a peripheral plasma membrane protein, belongs to the MAGUK (membrane-associated guanylate kinase) family of proteins [1]. Among all known MAGUKs, CASK is unique in containing a CaMK (Ca\(^{2+}\)/calmodulin-dependent kinase) domain at its N-terminus [2]. Additionally, CASK contains two L27 (Lin2, Lin7) domains, a PDZ domain, and an integrated SH3 (Src homology 3) and GuK (guanylate kinase) domain [3]. In all known animal species, a single gene codes for CASK. In humans, CASK is present on the short arm of the X chromosome [4]. Deletion and truncation mutations in the X chromosome region coding for the CASK protein have been documented in several studies investigating X-linked mental retardation [5–7]. Often, the patients with such mutations are female and display structural deformities in the brain, and head, neck and face region [8–10]. Missense mutations in the CASK gene have also been found in male patients with X-linked mental retardation [11], underlining its indispensable function in proper brain development and clinical importance.

It has long been recognized that the functions of a gene, and even entire signalling pathways involving the gene, demonstrate strong evolutionary conservation from simple unicellular organisms to humans. This has driven the way biological experiments are carried out and has led to the development of several model systems. CASK, however, strays from this established pattern of conserved protein function. CASK was discovered in three different animal groups in the same year (1996): nematodes (Caenorhabditis elegans), arthropods (Drosophila melanogaster) and mammals (Mus musculus). It was found in a screen for vulval defects in C. elegans, where it presumably functions in the proper localization of Let23 tyrosine kinase [12]. In Drosophila, CASK was initially speculated to be an alternative CaMK (Caki) important for proper locomotion [13]. In mouse, CASK was identified as an interacting partner of neurexin, a presynaptic adhesion molecule [3]; this association led to the proposal that CASK acts as a presynaptic scaffold in vertebrates. However, in all of the above-mentioned animal models, deletion of CASK did not alter either the formation or the structure of the neuronal synapse. Cask-null mice die within hours of birth and display a highly penetrant cleft palate [14]. Mutations in the CASK gene in humans are associated with mental retardation and craniofacial abnormalities [9]. However, in the invertebrate models characterized to date, CASK is not essential for survival; deletion of CASK leads to a vulvaless phenotype in C. elegans and lowered mobility in D. melanogaster. The lowered mobility of CASK-null fruitflies, however, does...
not seem to arise from a defect in motor neurons or premotor neuropil, and no obvious structural or developmental abnormalities are observed [15]. These distinct and unrelated phenotypes of CASK deletion in different animal models are difficult to reconcile and underscore a clear need to understand the evolution of the CASK gene. Interestingly, CASK has been shown to be an unusual kinase presumably fashioned by evolutionary necessity [16,17]. Although CASK phylogeny has been studied in relation to other MAGUKs [18,19], a study aimed at understanding the evolution of the entire CASK gene itself has not been carried out. In the present article, we describe the major constraints that have shaped the evolution of CASK and differences which might account for the varied phenotypic observations in different animal models.

**Gene size**

We first surveyed CASK genes from different clades of the animal kingdom, examining size and exon–intron organization. The cDNA that codes for human CASK is only 2.7 kbp, however the size of the human CASK gene (~408 kbp) is much larger than the average human gene (28 kbp [20]). CASK gene orthologues differ in size approximately 55-fold (Figure 1), with invertebrate CASK genes being much smaller than those of vertebrates. The smallest known CASK gene orthologue exists in a multicellular organism belonging to the placozoan class, *Trichoplax adhaerens* (Figure 1). The size of this CASK orthologue is only ~8 kbp, approximately the same size as the mature CASK transcript in humans [21]. The largest known CASK gene orthologue is seen in primitive new world monkeys called common marmosets (*Callithrix jacchus*) and exceeds the human CASK gene size by 50 kbp.

Despite the tremendous increase in the size of the CASK gene over the course of animal evolution, the changes in the number of introns have been minimal, and most of the gene size increase can be attributed to an increase in the size of the introns (Figure 1). The key question is whether this increase in intron size relates to the critical functions unique to mammalian CASK. Introns are involved in cellular activities that include (i) supplying non-coding RNAs such as microRNAs, (ii) regulating transcription, (iii) influencing alternative splicing and trans-splicing, (iv) nuclear export, (v) regulating the stability and decay of mRNA, (vi) exon shuffling, and (vii) increasing the crossing-over of coding sequences during meiosis [22]. One consequence of the expansion of intron length in the CASK gene could be increased complexity in the regulation of CASK transcription. CASK is expressed in many tissue types [21]. The variety and types of cells increase in more complex multicellular organisms, and this increase in complexity is accompanied by the need for precise tissue-specific regulation of expression. Indeed, long introns seen in human and mouse genes have been proposed to be related to developmental timing of gene expression [23], suggesting that CASK might play a more critical role during mammalian development than in organisms such as *Drosophila* and *C. elegans*. CASK’s extraordinarily large introns, largely concentrated in the early portion of the gene as is commonly observed [24], may contain intron-mediated enhancement signals. When the human and mouse CASK gene sequences were probed

---

**Figure 1** | Evolutionary expansion in CASK gene size occurs primarily in introns

Length of orthologues of the human CASK gene (black) and total length of introns contained within the corresponding sequence (grey). The broken line represents the longest CASK orthologue gene represented in invertebrates.
with IMEter, version 2.0 [25], a bioinformatics tool designed to detect such signals, the first two introns were predicted to moderately enhance expression. Larger introns can also contribute to the production of more alternatively spliced variants. The human CASK gene encodes 17 transcripts, a significantly larger number than its orthologues in C. elegans and Drosophila, each of which produce six CASK transcripts; the surprising length of the mammalian CASK gene may thus provide more functional flexibility for mammalian CASK. Another intriguing consequence of the vast expansion of intron length in the mammalian CASK gene is the possible exonization of conserved intron sequences [26]. In fact, Dembowski et al. [27] recently showed that CASK contains an inducible exon in the GuK domain that is expressed only upon neuronal stimulation. In addition, the mammalian CASK gene locus contains two orphan rhodopsin-like GPCRs (G-protein-coupled receptors) embedded in its introns [28]. It is reasonable to conclude that the increases in intron length seen in the mammalian CASK gene probably contribute to the novel and essential function(s) unique to the CASK protein in mammals through a variety of mechanisms. It has been proposed that the CASK gene underwent a duplication event in the vertebrate lineage, followed by a loss of the exons encoding the CaMK domain, to give rise to the MPPI (membrane palmitoylation protein 1) gene [18,22]. However, the MPPI gene is unable to adequately compensate for the lack of vertebrate CASK gene function, as evinced by the lethality of the CASK knockout in mice and the developmental impact of CASK mutations in humans. This suggests a critical role for the CASK CaMK domain. There is no known orthologue of MPPI in invertebrates. The CASK gene in invertebrates, however, is known to produce a protein with an alternative N-terminus (CASK-α in Drosophila and Lin-2b in nematodes) [15] (Figure 2A). In Drosophila, this alternative protein (CASK-α; Figure 2A) starts with a signature palmitoylation sequence and may represent an MPPI-like protein. The functions of these transcripts have yet to be addressed, but this might be an evolutionary strategy that allows for the generation of a diverse set of invertebrate MAGUK proteins with fewer genes [15]. Drosophila that lack CASK-β (the mammalian CASK orthologue) maintain viability, fertility and circadian rhythm, indicating that canonical CASK is non-essential in invertebrates, unlike vertebrates.

Protein phylogeny and structural conservation
In general, the evolution of CASK aligns well with our understanding of animal evolution. The only striking difference occurs with the tunicate CASK gene representative. Tunicates, which are the animals closest to vertebrates, formed a distinct separate clade with insects when the CASK phylogeny was generated (Figure 2B). Cnidarian CASK exhibits a higher similarity with human CASK compared with C. elegans CASK, supporting the notion of accelerated evolution in nematode worms [29]. Similarly, the distance between vertebrate and tunicate CASK might

Figure 2 | Phylogenetic comparison of CASK gene products
(A) Domain arrangement of classical CASK and alternative CASK proteins produced by Drosophila and Caenorhabditis CASK genes. (B) CASK protein sequences were obtained from NCBI database, alignment and phylogenetic tree were constructed using http://www.phylogeny.fr/ (the number denotes substitutions per site).
The sequence and structure of individual CASK domains are highly conserved (A) Sequences of the designated CASK orthologues were aligned using the Multalin View in Chimera [35]. Overall percentage identity between the human sequence and each orthologue was calculated, as well as the percentage sequence identity for each domain (green, higher than overall sequence identity; red, lower). Domains were based on human CASK sequence (NP_003679.2) as follows: CAMK (residues 13–322), L27 (residues 347–470), PDZ (residues 499–572), SH3 (residues 615–681) and GuK (residues 773–912). (B) Overlay of homology models generated using Modeller 9.9 [36] for the Drosophila (blue) and Placozoan CASK (purple) CAMK domain, PDZ domain and GuK domain based on available X-ray crystal structures from corresponding mammalian CASK structures (CAMK, 3C0I, PDZ, 1KWA, GuK, 1KGd).

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Species</th>
<th>CAMK</th>
<th>L27</th>
<th>PDZ</th>
<th>SH3</th>
<th>GuK</th>
<th>Overall Sequence Identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Placozoan</td>
<td>Trichoplax adhaerens</td>
<td>45%</td>
<td>28%</td>
<td>47%</td>
<td>39%</td>
<td>43%</td>
<td>38%</td>
</tr>
<tr>
<td>Cnidarian</td>
<td>Nematostella vectensis</td>
<td>60%</td>
<td>41%</td>
<td>70%</td>
<td>55%</td>
<td>71%</td>
<td>54%</td>
</tr>
<tr>
<td>Nematode</td>
<td>Caenorhabditis elegans</td>
<td>45%</td>
<td>30%</td>
<td>78%</td>
<td>69%</td>
<td>68%</td>
<td>50%</td>
</tr>
<tr>
<td>Arthropod</td>
<td>Drosophila melanogaster</td>
<td>64%</td>
<td>38%</td>
<td>81%</td>
<td>71%</td>
<td>71%</td>
<td>60%</td>
</tr>
<tr>
<td>Tunicate</td>
<td>Oikopleura dioica</td>
<td>56%</td>
<td>24%</td>
<td>59%</td>
<td>49%</td>
<td>42%</td>
<td>43%</td>
</tr>
<tr>
<td>Platyhelminth</td>
<td>Schistosoma mansoni</td>
<td>69%</td>
<td>47%</td>
<td>72%</td>
<td>72%</td>
<td>66%</td>
<td>61%</td>
</tr>
<tr>
<td>Mollusk</td>
<td>Lymania stagnalis</td>
<td>72%</td>
<td>52%</td>
<td>76%</td>
<td>67%</td>
<td>71%</td>
<td>63%</td>
</tr>
<tr>
<td>Chordata</td>
<td>Xenopus tropicalis</td>
<td>97%</td>
<td>82%</td>
<td>100%</td>
<td>99%</td>
<td>96%</td>
<td>94%</td>
</tr>
<tr>
<td>Chordata</td>
<td>Mus musculus</td>
<td>97%</td>
<td>82%</td>
<td>100%</td>
<td>100%</td>
<td>99%</td>
<td>95%</td>
</tr>
<tr>
<td>Chordata</td>
<td>Homo sapiens</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Also be indicative of accelerated evolution in the urochordate clade. The overall residue identity between the simplest CASK and human CASK is 38% (Figure 3A). Within the vertebrate clade, the residue identity typically exceeds 93%, with all mammalian CASK sequences being virtually identical (Figure 3A). CASK is a multidomain protein with a highly conserved domain arrangement. Each protein domain can fold independently. We therefore looked at the domain-specific conservation of residues. As seen in Figure 3(A), residues within the domains of CASK are better conserved than linker sequences, with the PDZ domain exhibiting the highest conservation, and the L27 domain, which links the CaMK domain and the PDZ domain, displaying the least conservation. The high conservation of the PDZ domain across the phyla indicates that interactions involving this domain must have been established and optimized very early in evolutionary history.

High-resolution three-dimensional structures for several of the highly conserved domains of CASK have been solved [16,30,31]. Homology modelling of three of these domains (CAMK, PDZ and GuK) was performed to study the impact of evolutionary changes in the primary sequence on the conservation of structure. Using the human CASK domains as structural templates, homology models were calculated on the basis of alignments of the CASK primary sequence of D. melanogaster (NP_001163681.1) and T. adhaerens (XP_002111823.1) (Figure 3B). The T. adhaerens orthologue is the most distantly related CASK sequence to the human structures and has correspondingly low overall sequence identity (38%). Even the largest average backbone RMSD (root mean square deviation) [0.3686 Å (1 Å = 0.1 nm)] observed between a pool of ten homology models of the Drosophila CASK GuK domain and the human CASK GuK domain (Figure 3B) is relatively small, strongly suggesting...
that the global three-dimensional structure of these domains is likely to be conserved in CASK orthologues.

**Residue change and functional evolution**

Single residue substitutions can have dramatic effects on protein–protein interactions, even when overall tertiary structure is maintained. The CaMK domain of CASK interacts with three neuronal proteins (Mint-1, caskin and liprin), probably in a competitive manner, but CaMKs themselves do not interact with these proteins. These CASK-specific interactions can be traced to a single aspartate-to-valine substitution (at Val^{117}) in the CaMK domain of CASK [32,33]. Such subtle changes can thus give rise to novel functions by promoting novel interactions. The overall evolutionary substitutions in CASK are severely constrained by the fundamental functions, and hence three-dimensional structures, of individual domains. Careful observation, however, reveals substitutions in critically situated residues that have a major impact on CASK’s overall function. CASK is the only known kinase which is inhibited by divalent ions such as Mg^{2+} [16]. This unique property of CASK is largely dependent on substitution on just four residues lining the nucleotide-binding pocket [17]. Three of these four critical substitutions are absent from the Trichoplax CASK sequence, which retains residues identical with the canonical CaMKs, suggesting that the Trichoplax CASK might be a conventional Mg^{2+}-utilizing kinase [17]. Functionally, the atypical Mg^{2+}-sensitive kinase function of CASK could convey three potential advantages: (i) limiting substrate specificity to stoichiometrically complexed proteins, (ii) the ability to sense divalent cation levels in the milieu, and/or (iii) the ability to sense the level of total ATP level in the milieu.

Mammalian CASK is capable of two interactions (with caskin [34] and liprin [32]) not seen in other CASK orthologues. We were unable to co-immunoprecipitate either caskin or liprin with Drosophila CASK (L. LaConte and K. Mukherjee, unpublished work). The peptide necessary to bind CASK is absent from Drosophila caskin; similarly, the CASK interaction with α-liprin requires a loop in liprin which is absent from invertebrate α-liprin [32]. Such changes in molecular partners can have a profound effect on the function of a molecule which is otherwise conserved. Many of the CASK residues involved in the CASK–liprin interaction (Glu^{271}, Asp^{272}, Glu^{280} and Tyr^{286}) [32] are not present in invertebrate CASK. Specifically, Tyr^{286} is replaced by a threonine residue in Drosophila CASK and a positively charged lysine residue in C. elegans CASK. The functional importance of this vertebrate-specific tyrosine is emphasized by the fact that the mutation Y268H is the cause of mental retardation in humans [11]. In human CASK, position 28 is arginine, but in Trichoplax CASK, it is leucine. A back mutation to leucine in an Italian family has been identified as the cause of a neurodevelopmental anomaly, FG syndrome [10], suggesting a gain of function associated with this leucine-to-arginine mutation in the human protein. A similar phenomenon is observed at CASK position 396; a proline-to-serine mutation in human CASK is associated with mental retardation [11], and yet this position in Drosophila is normally occupied by a serine residue. These observations all suggest that substitution at single critically placed residues in CASK might, throughout evolution, lead to gains in function without affecting the overall architecture of the protein.

**Conclusion**

Animal models are frequently used to understand the biological basis of disease conditions relevant to humans. Functional conservation of many genes and their products has been documented, validating the use of these animal models. Despite such conservation, there are functional divergences which are shaped by evolutionary pressure. Such changes have signatures at both the genetic and protein levels. A detailed knowledge of the phylogenetic divergence at the genetic and protein level allows us to better utilize and interpret data from various animal models. CASK, a MAGUK protein, is critical for neurodevelopment in mammals and essential for survival. In invertebrates, however, CASK function is divergent. In C. elegans, Lin-2-null (CASK orthologue) animals exhibit a vulval phenotype without any apparent neuronal defect. In Drosophila, CASK is present predominantly in the nervous system, as in vertebrates, but other than a locomotion phenotype, CASK-β-null fruitflies are identical with control fruitflies, with no major neurodevelopmental phenotypes. This dissimilarity of phenotype across clades argues for a newly acquired function of CASK in vertebrates. Other than an increase in intronic size, CASK is a well-conserved gene. Most substitutions within the coding region of CASK fail to cause major changes in its three-dimensional structure, suggesting that CASK has a function that is conserved across multiple clades. The CASK gene encodes an unusual protein kinase fused to a scaffolding MAGUK protein. This common structural theme suggests that the likely function of CASK combines a slow kinase activity dependent on the cellular divalent cation-to-ATP ratio with its scaffolding properties. CASK may therefore function as either a divalent cation sensor or an ATP sensor. Surprisingly, mutations in human CASK that parallel substitutions from lower organisms give rise to neurodevelopmental defects. Vertebrate CASK has also acquired specific variations within its coding region that explain changes in its interactions with other proteins and its enzymatic activity. In addition to changes in CASK itself, there are noticeable changes in its interacting partners such as caskin and liprin, which cause commensurate changes in CASK function. We hypothesize that these critical changes in CASK and its protein partners have optimized CASK function for different evolutionary niches.

**Acknowledgement**

We thank Vrushali Chavan for her input.
Funding

This work was funded by a start-up package to K.M. from Virginia Tech Carilion Research Institute (VTcRI).

References


19 Muller, G.V., Muller, I.M. and Werner, E.G. (2000) Sponge proteins are more similar to those of Homo sapiens than to Caenorhabditis elegans. Biol. J. Linn. Soc. 71, 821-828


Received 24 April 2013
doi:10.1042/BS020130061