Conformational Spread Drives the Evolution of the Calcium-calmodulin Protein Kinase II

Shahid Khan (khansh@nih.gov)
National Institute of Neurological Disorders and Stroke

Research Article

Keywords: calcium calmodulin (Ca2+CAM), protein kinase II (CaMKII) , learning, organismal evolution, CaMKIIa holoenzyme

Posted Date: May 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-506808/v1

05/10/'21
Submitted for publication

Conformational spread drives the evolution of the calcium-calmodulin protein kinase II

Shahid Khan

Molecular Biology Consortium, Lawrence Berkeley National Laboratory,
Berkeley, CA 94720
Laboratory of Cell Biology, NINDS, NIH, Bethesda, MD 20892.

Abstract

The calcium calmodulin (Ca^{2+}CAM) dependent protein kinase II (CaMKII) decodes Ca^{2+} frequency oscillations. It has a central role in learning. I matched residue and organismal evolution to collective motions deduced from the atomic structure of the human CaMKIIα holoenzyme. Protein dynamic simulations and bioinformatic analysis showed its stacked ring architecture conformationally couples kinase domains (KDs) via its central hub. The simulations revealed underlying β-sheet collective motions in the hub $\alpha\beta$ association domain (AD) map onto a coevolved residue network and partition it into two distinct sectors. The holoenzyme evolved in metazoans by stabilization of ancient enzyme dimers and fold elongation to create a second, metastable sector for ring assembly. Continued evolution targeted the ring contacts for lateral conformational spread. The α isoform, predominantly expressed in the brain, emerged last and evolved rapidly in sync with the poikilotherm-homeotherm jump in the evolution of memory. The correlation between CaMKII dynamics and phylogenetics argues single residue evolution fine-tunes hub conformational spread. The central role of CaMKII ringed architecture In the brain could be to increase Ca^{2+} frequency response range for complex learning functions.

Introduction

The frequency decoding of calcium pulses by calcium calmodulin-dependent kinase II (CaMKII) is central to CaMKII control of memory in the brain [1]. Remarkably, individual enzymes decode Ca^{2+} pulses. The multi-subunit holoenzyme architecture, a two-ring stack of subunits with mirror symmetry, is unique among members of the calmodulin-dependent (Ca^{2+}CaM) kinase family [2]. Individual subunits consist of a canonical kinase domain (KD) with a C-terminal pseudo-substrate regulatory segment (R) connected via flexible linkers to an association domain (AD) that forms the central hub (**Fig. 1A**).

What is the importance of the ringed architecture for the CaMKII function? Attention has focused on the variable alternatively spliced linkers that modulate interactions with the actin cytoskeleton and the balance between activating and inhibitory phosphorylation[3-8]. The sequence variability of the conserved AD has not been as well analyzed. I present here an analysis of this variability to follow-up recent crystal structures and biochemical assays that have highlighted AD flexibility and its role in activation triggered subunit exchange[9-11]. The analysis is coupled to the study of how the flexibility encodes long-range conformational spread between KDs and traces the responsible variations during the evolution of the CaMKIIα isoform, dominantly expressed in the brain[4]. Correlations between CaMKIIα evolution and behavioural complexity suggest the ringed architecture tunes CaMKII response to calcium pulses.

Results

Hub β-sheet dynamics mediate long-range KD-KD coupling.

A tetramer (subunits A, C, G, I) that encapsulates all domain contacts was extracted from the human holoenzyme structure (PDB:3SOA). The ADs of subunits A, C formed the inter-stack dimer (Vert-Dim) and made lateral dimer contacts (Lat-Dim) with ADs of subunits G, I as well as cis ("clip") and trans ("spur") KD contacts. Conformational ensembles were generated with tCONCOORD from this structure (**Methods**). The tetramer root-mean-square fluctuation (rmsf) profile showed the associated KD's amplified the Vert-Dim interfacial hinge motions (**Fig. 1B**). This hinge is more rigid (0.29 $\pm$ 0.002 nm) than the Lat-Dim hinge (0.355 $\pm$ 0.016 nm) consistent with previous work[9,12] or the AD-KD "clip" contact (0.292 $\pm$ 0.005 nm). I used principal component analysis (PCA) to classify long-range collective motions within the tetramer. These motions replicated the modes deduced from a survey of 743 protein data bank (PDB) structures as a fundamental property of collective motions mediated by β-folds[13]. The first three PCs (PC1

– PC3) reported orthogonal tilt motions at the Vert-Dim contact and anisotropic β_2-β_5 sheet bending and twisting motions at the Lat-Dim contacts. In PC1, crankshaft (rotation + extension) KD motion coupled to the AD-AD β_2-β_6 sheet bending and twisting modes **(Fig. 1Ci-iv)**. The Lat-Dim fluctuations were evident in the tetramer hub PC1. The PC1-PC3 motions for the ADs in the tetramer had reduced amplitude and anisotropy relative to the isolated monomer.

I used two tools to correlate local dynamics with molecular evolution (**Methods**). First, I assessed the energetic cost of residue contacts. The frustration (ΔE_{fr}) was the score of the stabilization energy of the native contact relative to all possible contacts. This measure showed that the contacts made by the regulatory segment R with both the KD C-lobe and AD were energetically stressed (**Fig. 1Di**). Second, I constructed mutual-information based networks to encode the dynamics of local fragments as 1D-strings (**Supplementary Fig. 1**). The DFG α_3-β_2 loop was the central peak common to the network centrality profiles of the monomer and the tetramer KDs. The R T286 fragment (R_{286}) is the central node in the monomer but is suppressed in the tetramer; a possible consequence of the inter-subunit clip KD-AD contact at the calcium calmodulin-binding site located at the other end of the R helix ($R_{300\text{-}306}$). The dynamic couplings indicate that the spur KD-AD transmits AD motions to rotation-translation of KD helices $\alpha_{6\text{-}7,9}$ (**Fig 1Dii**).

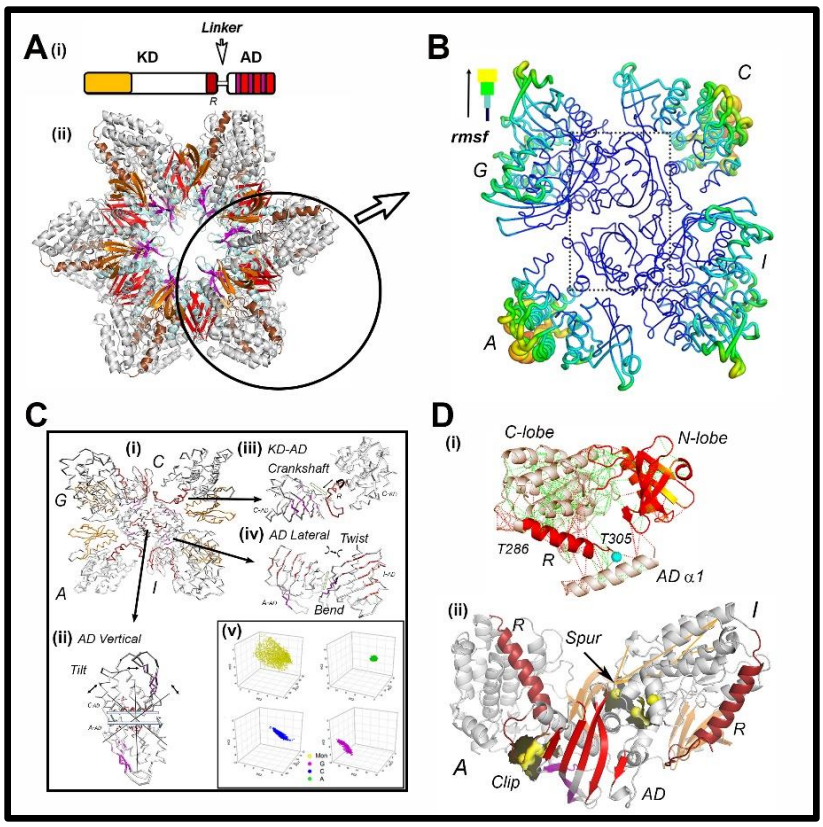


Fig. 1: A. Architecture. (i) Subunit. *Disordered linkers with varying length and composition connect the kinase domain (KD (N-lobe (orange), C-lobe (white)) with the association domain (AD). The pseudo-substrate, regulatory segment (R (brown) binds Ca^{2+}CAM. The AD β-sheet forms vertical (red) and lateral (magenta) holoenzyme contacts.* ***(ii) Assembly.*** *The ADs form the central hub in the multi-subunit holoenzyme (CaMKIIα PDB:3SOA). A tetramer (circle) was extracted for analysis of conformational fluctuations.* ***B. Flexure.*** *The flexibility (rmsf) profile derived from the tetramer conformational ensemble.* ***C. Collective Motions. (i)*** *A conformation in the CaMKIIα tetramer tCONCOORD ensemble (**Supplementary Video S1**).* ***(ii) Vertical dimer.*** *Tilt.* ***(iii) Lateral dimer.*** *Bend and twist.* ***(iv) KD-AD.*** *Crankshaft (extension + rotation). Rectangles represent β sheet long axes.* ***(v).*** *PCIPC2PC3 plots of the human CaMKIIα AD;* ***(i)*** *monomer and* ***(ii-iv)*** *the tetramer (**Supplementary Video S2**).* ***D. KD-AD Coupling.*** ***(i) R contacts.*** *Energy frustration - (stable (green), stressed (red)).* ***(ii) Dynamic network.*** *The KD-AD contact residues and surface (yellow) are shown. Dynamic couplings (orange lines). 3D-views in **Supplementary Video S3-4**.*

Hub AD fold evolution reflects its dynamics.

The clues to CaMKII evolution are found in the AD hub. The CaMKIIα AD has a common fold, but low sequence homology, with the dimeric yeast nuclear pore complex component NTF2 and a bacterial dehydratase[14,15]. High-throughput genome sequencing and X-ray crystallography[16,17] identified the CaMKII-AD superfamily, sometimes cited as the NTF2 superfamily (PF08332). The residue conservation, coevolution and energetics bridged fold evolution and molecular mechanics (**Methods**).

The network analysis showed AD flexibility is constrained in the tetramer relative to the free monomer. The Vert-Dim had prominent β_2,β_3 and β_6 peaks while the Lat-Dim had prominent peaks for

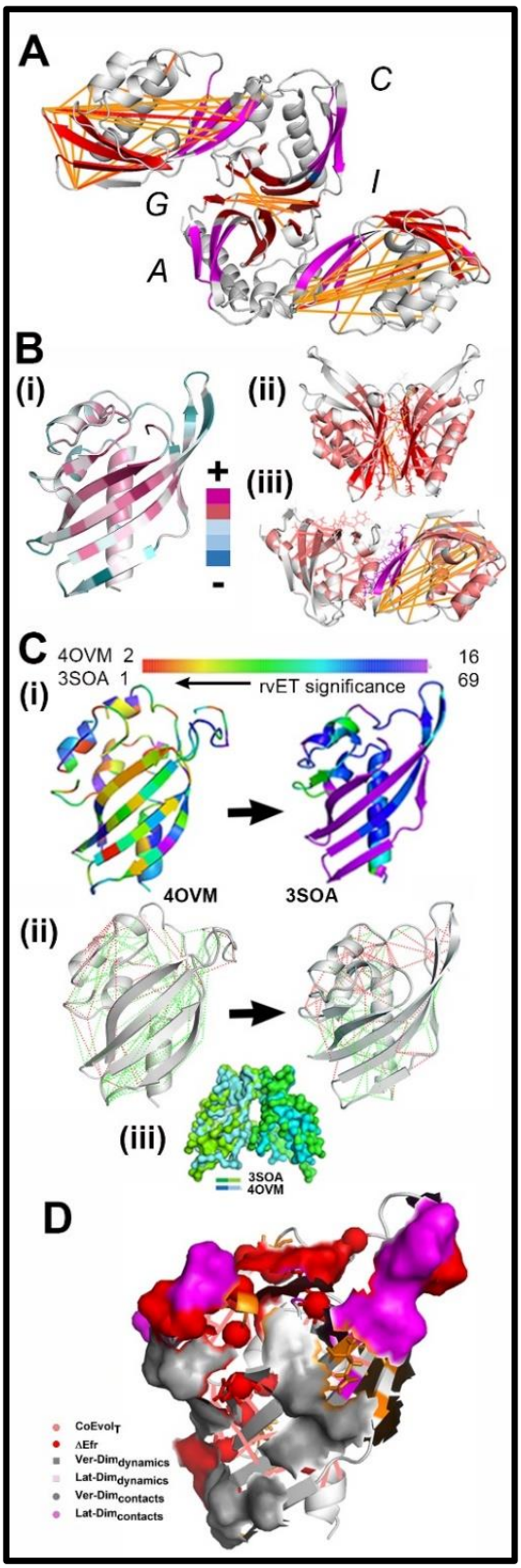
the α_1-α_2 loop, α_3 and β_4-β_5 junction in their centrality plots consistent with the top couplings The Vert-Dim dynamics were defined by couplings that connected the β_2-β_4 sheet centre in one AD with the α_2-β_1 and β_3-β_4 loops in the other either side of the rigid contact (β_3-β_4, β_6 strands, α_2-α_3 loop). The top (1%) dynamic couplings at the Lat-Dim contact of helix α_3 with the adjacent AD β_3–β_5 loops, affected β-sheet curvature in the latter (**Fig. 2A**). The PF08332 MSA surface conservation profile identified the Vert-Dim β_2-β_6 interface as the most well-conserved. This MSA seeded an expanded sequence set for analysis of residue coevolution. The top 15% of significant coevolved residue pairs, typically adjacent to the conserved residue positions, superimposed with the dynamic couplings in maps of the Vert-Dim and Lat-Dim subcomplexes. The coevolved pairs formed two clusters – the larger cluster stitched helices $\alpha_{2\text{-}3}$ with strands $\beta_{1\text{-}3}$, while the smaller one bonded the central $\beta_{2\text{-}5}$ sheet with long helix α_1 (**Fig. 2B**)

The CaMKII-AD fold is a curved β-sheet (β_2-β_6) anchored to a long helix α_1. Its evolution is seen by comparison of the crystal structures of a dimeric *Streptomyces* enzyme (PDB:4OVM) with the human CaMKIIα (PDB:3SOA) Vert-Dim (**Fig, 2C**). The 3SOA fold had a longer helix α_1 and β-sheet for the formation of Lat-Dim contacts. The EV trace, built from sequence homologs, was superimposed on each structure. The ET MSAs were constructed from 20 4OVM.pdb (E value < 10^{-2}) and nearly 500 3SOA.pdb (E value < 10^{-60}) sequence homologs. For 3SOA.pdb, the KD DFG loop had the most significant score (1.7$\pm$1.0), followed by the $R_{286\text{-}305}$ segment score (22.2$\pm$1.7), both key determinants of kinase activation. The rvET profiles revealed broad evolution of the 4OVM fold while the evolution of the 3SOA fold was localized to the Lat-Dim contact. The ΔE_{fr} profiles showed the energetic cost of the evolution of CaMKIIα Lat-Dim contact. Its structural elements formed metastable interactions relative to the sector associated with the Vert-Dim contact that retained favourable, interactions seen in the bacterial (4OVM.pdb) homolog. The superimposed two dimer structures illustrated the conservation of the Vert-Dim contact.

The overlap of the evolution metrics with the dynamic network is summarized in **Fig. 2D**. The composite $\alpha_{2\text{-}3}\beta_{1\text{-}3}$ contact network and the central β-sheet hinge were the principal drivers of CaMKII-AD evolution. Coevolved contacts between helix α_1 and the β sheet maintained the cross-section of the hydrophobic core used, in bacteria, for sequestration, acid-base cyclization, and isomerization of aromatic polyenes. The long-range mechanical relay explains how hydrogen bonds control shear of the hydrophobic core sidechains to control subunit stoichiometry, most simply by regulating β-sheet curvature[18]. It also propagates inter-subunit conformational spread via the emergence of the Lat-Dim contact.

The brain CaMKII α isoform emerged most recently from ancient bacterial enzymes.

The evolution metrics highlighted structural modulators of the AD-fold, but the phylogenetic analysis was required for their chronology. The PF08332 sequences were clustered, and a global phylogenetic tree constructed from the cluster representatives. The tree demarcated into prokaryotic, eukaryotic, and archaeal branches (**Fig. 3A**). The most populous node (I (n=45)) contained enzymes from extremophilic nitrate and sulfur-reducing bacteria for aromatic compound biosynthesis. The long lengths characteristic of the bacterial branches indicated a high rate of evolution. Three smaller nodes and a more distant fungal node diverged with node I in a major split from the tree stem that branched with several outliers before it reached archaeal and eukaryotic representatives. The second most populous node (II (n=35)), represented by the rat CaMKIIβ AD, was composed of metazoan sequences. It had short branch lengths consistent with constrained evolution upon integration with the KD into multiple phosphorylation pathways. Protozoan sequences formed the third most populous node (III (n =18)). The examples of horizontal gene transfer (HGT) included a marine bacterium in node III and protozoan fungal and green algae around node I. These examples notwithstanding, the monophyletic tree linked ancient extremophiles and other bacterial species to mammalian and archaeal relatives. The dimer, ring and CaMKII kinase assemblies segregated to nodes I, III and II, respectively.


A
C
I
G
A
B
(i)
(ii)
(iii)
+
-
C
4OVM 2
3SOA 1
16
69
rvET significance
(i)
4OVM
3SOA
(ii)
(iii)
3SOA
4OVM
D
CoEvol$_T$
ΔEfr
Ver-Dim$_{dynamics}$
Lat-Dim$_{dynamics}$
Ver-Dim$_{contacts}$
Lat-Dim$_{contacts}$

133

Fig. 2: The dynamics and evolution of the CaMKII AD. A. Tetramer-ACGI AD dynamics. *The top dynamic couplings computed between 4-residue fragments (yellow (weak) -> orange -> red (strong)).* ***B. Evolution. (i)*** *Residue conservation.* ***(ii, iii)*** *The superposition of the dynamic (thin orange) and coevolved (thick salmon) couplings adjacent to the* ***(ii)*** *vertical and* ***(iii)*** *lateral AD contacts.* ***C.*** *Structural evolution.* ***(i)*** *The evolutionary trace for structures of an ancient bacterial enzyme (PDB;40VM) versus human CaMKIIα (PDB:3SOA).* ***(ii)*** *Residue contact energetics of the 4OVM and 3SOA structures.* ***(ii)*** *Conservation of the dimer contact (RMSD = 0.69 nm).* ***D.*** *The 3SOA-AD structure with the superposed interdomain contacts, energetically frustrated contacts, dynamic and coevolved couplings. 3D-views in* ***Supplementary Video S5-10****.*

I selected 22 crystal structures of the CaMKII-AD superfamily, in addition to human CaMKIIα, to detail fold evolution. The 2D-heatmap (**Supplementary Fig. S2**A) and pseudo-phylogenetic tree (**Fig. 3A. Box**) constructed from the DALI Z-scores demarcated prokaryotic and eukaryotic structures. The α_1 N-terminus, loops at either end of helix α_3, the β_4– β_5 loop, and β_5 C-terminus were the variable elements. The dimer was the dominant assembly (n =11), followed by holoenzymes (n=7) monomers (n = 2), a heterodimer. trimer and ring. The multiple oligomeric states in *Streptomyces*, an ancient bacterial lineage[19] in tree node I, could have sharpened response to limiting nutrients in primordial environments. The protozoan ring assemblies had similar architecture to the 3SOA AD hub, while the marine bacterium *Pirellula* sp.SH-Sr6A assembles a fourteen-subunit oligomer that may also form a homologous ring structure[18].

The CaMKII kinases have been found, thus far, in metazoans. I constructed phylogenetic trees from one thousand sequences most homologous to the *Caenorhabditis elegans* CaMKII, an ancient CaMKII with well-characterized structure and biochemistry to trace holoenzyme evolution (**Fig. 3B**). The nematodes (n = 18) formed the base of the stem that bifurcated to arthropod or chordate representatives. Insects (n = 8) and arachnids (n = 17) formed dedicated arthropod group nodes. Chordate as well as arthropod sequences segregated to a large mixed node (n = 106). The two chordate nodes (n = 554, 242) contained 271δ, 111γ and 7β isoform sequences with the β sequences all within the larger node. Out of the four isoforms (α, β, γ, δ), no sequences of the "α" isoform were found.

I, therefore, gathered one thousand closest homologs of the rat neuronal α and β isoforms to trace their evolution. The resulting sequences (234α, 352β, 830γ, 575δ) were clustered (n = 34) for tree construction. This tree branched according to isoform rather than phyla (**Fig. 3C**) in contrast to the *C. elegans* rooted tree, suggested it reflected differences in tissue-dependent isoform expression within organisms. These differences could arise from adaptive selection for tissue-specific signal phospho-relays orchestrated by the KD and alternatively-spliced linker variants[4]. I compared domain tree topologies to test this idea (**Supplementary Fig. S2B**). The similarity score (~ 0.6), although weaker, was comparable to

the scores (> 0.8) obtained for proteins with strongly interacting domains such as ribosomal components and the F_1 ATP synthase subunits[20]. Thus, AD-KD coupling underlies the evolution of the CaMKII isoforms.

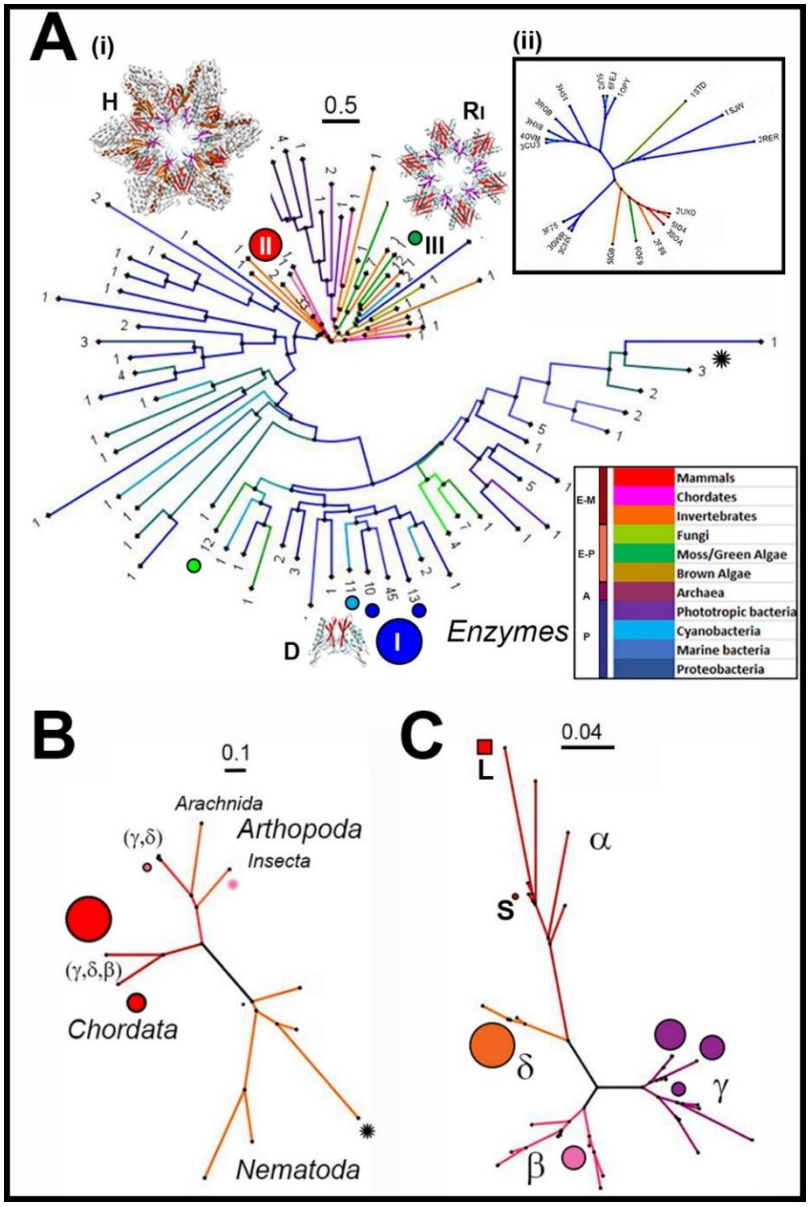


Fig. 3: CaMKII phylogenetics. A.(i) CaMKII AD superfamily tree. *The colour-coded tree (E-M (metazoan), E-P (protozoa), A (archaea), P (prokaryotes)) was constructed from sequences representing 81 PF08332. The major nodes are composed of bacteria (I (P. n =45)), vertebrate (II (E-M. n = 33)) and diatom/fungal (III (E-P. n =)) clusters, followed by four smaller (3 bacterial and a fungal node in a large group that includes node-1. The AD assemblies associated with nodes I-III (D (dimer, β contact (red)), Ri (ring, lateral β contact (magenta)), H (holoenzyme. R (brown, KD C-lobe (orange))) are shown. The asterisk indicates a possible common ancestor.* ***(ii) Structural phylogenetics.*** *Tree of 23 crystal structures of the CaMKII-AD homologs based on DALI scores (colour-coded as in (i).* ***B. Metazoan CaMKII-AD evolution.*** *Tree based on clustering of 1000 homologs of the C. elegans CaMKII (black asterisk). Midge (pink asterisk).*

C. Isoform evolution. *Tree based on clustering of 2000 homologs of the human CaMKII α and β isoforms. Circles mark major clusters (colour = phylum (A, B), isoform (C); diameter = membership), except for the most distant, larger CaMKII α node (L. square).*

The poikilotherm – homeotherm transition is a major step in CaMKIIα evolution

A more detailed analysis of the divergence and species composition of the α and β branches of the tree (Fig. 3C) provided insights into α isoform evolution. The δ isoform was the least diverse, followed by the β isoform. The major β isoform node was contaminated (~ 10%) with γ isoform sequences reflecting a close evolutionary relationship between these isoforms ($\overline{\Delta X}_{\beta-\delta} = 0.116 \pm 0.008$) (**Methods**). The node species composition was evenly balanced between poikilothermic and homeothermic (mammalian) vertebrates. The smaller α node (α_S) though most closely related to the other isoforms, nevertheless diverged significantly from them ($\overline{\Delta X}_{\alpha S-\beta\gamma\delta}$= 0.208±0.014 versus $\overline{\Delta X}_{\beta\gamma\delta}$ = 0.148+0.009). It consisted dominantly of poikilotherms. The larger node (α_L) diverged even more ($\Delta X_{\propto L-S} = 0.126$). It consisted largely of mammalian particularly anthropoid sequences consistent with the rapid evolution of memory in homeotherms (**Fig. 4A**) due to the development of the hypothalamus for temperature homeostasis and the associated amygdaloid complex[21].

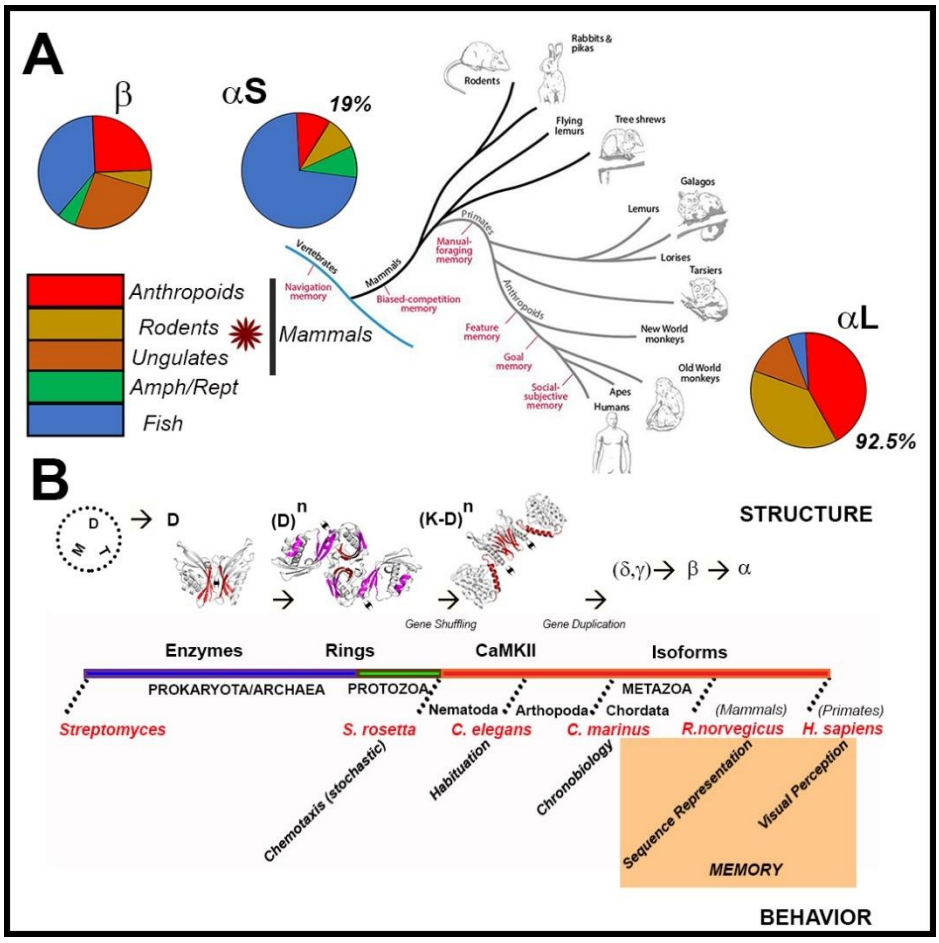


Fig. 4: A. CaMKIIα and Memory. *The major CaMKIIα phylogenetic nodes straddle the evolution of memory (tree from[22] with permission). Homeotherm (mammals (brown asterisk)) fractions for the nodes (I, II) are shown.* ***B. The correlation between CaMKII architecture and behaviour. Architecture.*** *The selection of the dimer (D) from other structures (M = monomer, T = tetramer) in bacteria seeded the emergence of ringed hub assemblies. The fusion with the kinase (K) domains coincided with the emergence of multicellularity (secondary structures colour coded as in Fig. 1). Diversity was created by linker alternative splicing and enhanced by the generation of isoforms.* ***Behaviour****. Work on model organisms suggests CaMKII evolution peaks with the development of cognitive memory.*

Discussion

This study has associated two dynamic processes that occur over dramatically different timescales - the evolution of the CaMKII holoenzyme over billions of years (>10^{14}s) with the macromolecular motions (<10^{-4}s) of the assembly. The emergence of the CaMKII holoenzyme from ancient enzymes and the correlation with behaviour involved distinct transitions coupled to fundamental changes in life forms (**Fig. 4B**). First, ring assemblies formed from dimeric enzymes. The comparison with the *Streptomyces* enzyme supports spectroscopic [23,24] and structural [25] evidence that the holoenzyme formed by serial extension of vertically oriented dimer homologs. Second, the increased biosphere gene pool associated with the

emergence of multicellular organisms led to HGT based shuffling events[26] that probably resulted in fusion with an ancestral KD. Bacterial, archaeal, and eukaryotic serine-threonine kinases have a common evolutionary origin [27,28], while Ca^{2+}CAM dependent protein kinases are present in nitrogen-fixing bacteria [29], The chemotactic choanoflagellate *Salpingoeca rosetta*, that switches between unicellular and multicellular lifestyles, has Homer and other primordial synaptic scaffolding proteins in addition to CaMKII[30]. Third, alternative splicing, important for example in midge *Clunio marinus* chronobiology[7] and short and long-term habituation in nematode *C. elegans* mechanosensory neurons[31], preceded CaMKII isoforms. Fourth, after the appearance of isoforms most likely by gene duplication events[4]. advanced memory mechanisms from timed sequence representation in rodents (*Rattus norvegicus*) to visual memory and perception in humans (*Homo sapiens*) emerged in mammals[32]. These mechanisms required a variable threshold and expanded range of electrical, hence post-synaptic Ca^{2+}, stimulation frequencies regulated by neuromodulators and associative learning[33,34]. The evolution of the α isoform jumps correlated with the transition from poikilotherms to homeotherms to reflect the increase in behavioural complexity.

What might be the role of the long-range dynamics reported here? A dimer is sufficient for CaMKII kinase activation[24]. In neurons, the multimeric CaMKII architecture allows multivalent binding to the actin cytoskeleton[35,36] or multiple partners at the post-synaptic membrane[37]. However, it is not clear what role hub flexibility has in these functions. Furthermore, actin-binding is the weakest for the α isoform[38] and unlikely to exert selection pressure. The most attractive explanation proposed thus far for the importance of hub dynamics is activation-triggered subunit exchange[10]. Such a mechanism would benefit from, though not require lateral conformational spread. It does not, however, explain the phylogenetic differences between the α and β isoforms since both undergo activation-triggered subunit exchange[9].

I, therefore, propose an alternative hypothesis, namely that conformational spread tunes conformational transitions as in the bacterial flagellar motor[39] to select and optimize response to a broad range of Ca^{2+} pulse frequencies. It explains the utilization of an energetically metastable, conformationally plastic sector for ring assembly. Evolution of the ancient dimer contact slowed after ring assemblies appeared allowing the hub to function as a semi-rigid connector module while the lateral contact formed a fine-grained dynamic code that continues to evolve as illustrated by Fig. 2C. Linker splicing is too coarse a mechanism and α has the fewest alternatively spliced linker variants among isoforms. The conformational dynamics of the R segment, a major KD network node, are coupled to hub dynamics and will be modulated upon calcium calmodulin-binding[40], subunit capture[41] and substrate occupancy[42] irrespective of whether the AD-KD contacts in the zero-linker holoenzyme are retained in finite-length

linker constructs. It is conceivable that both subunit variation, due to activation-triggered exchange, and KD undocking upon activation facilitate the combinatorial increase in the frequency range regulated by the conformational spread in the hub.

Methods

Phylogenetics: The multiple sequence alignments (MSAs) of the CaMKII AD (PF08332) and the protein kinase domain (PF00069) were downloaded from the Pfam database (*www.Pfam.org*[43]). The thousand closely related homologs for each of the *C. elegans* CaMKII, rat CaMKIIα and CaMKIIβ sequences were identified with Uniprot (*www.UniProt.org*[44]). The sequences were clustered with CD-Hit[45] at the 0.8 cutoff threshold. Hierarchical clustering with the 0.8 cutoff, followed by a 0.6 cutoff was used for the PF08332 and *C. elegans* sequence sets. The MSAs of the cluster representatives with constructed with MUSCLE[46]. Crystal structures were downloaded from Protein Data Bank (*www.rcsb.org*[47]).

Unrooted trees were constructed with FastTree using the JTT model of amino acid evolution. Correlation matrices of paired tip distances (Xi, Yi) were constructed for each tree. Isoform diversity was estimated by the mean tip distance ($\overline{\Delta X} = (\sum_{t=1}^{n} \Delta X)/n)$). The similarity between the CaMKII AD (X) and KD (Y) tree topologies was measured as the r score [20].

$$r = \sum_{i=1}^{n}(Yi - \bar{Y})(Xi - \bar{X})/\{\sqrt{\sum_{i=1}^{n}(Yi - \bar{Y})^2}\sqrt{\sum_{i=1}^{n}(Xi - \bar{X})^2}\} \qquad \text{Equation 1}$$

Sequence Analysis: The Pfam PF08332 MSA (1842 sequences) was input into ConSurf [48] for estimation of residue conservation. Additional CaMKKI-AD superfamily sequences of isolates from diverse habitats and clinical repositories were added to PF08332 in GREMLIN (*www.gremlin.org*[49]). The expanded dataset (16,485 sequences) was submitted for MSA construction with HHblits ($E < 10^{-6}$, 4 iterations, 75% coverage). The top 30 couplings (> 0.995 significance) were mapped on the 3D structure of the human holoenzyme.

Structure Analysis: The topology of the 3D crystal structures was analyzed with CCP4[50] and DALI[51]. The DALI C^{α}-C^{α} distance correlation matrix alignment optimizes the correspondence between aligned residues pairs from multiple structures. The results were represented as a heatmap. The DALI scores Z_{AB}, a metric for the correspondence between structures A and B corrected for the geometric mean length and the standard deviation was used for the construction of pseudo-phylogenetic dendrograms.

The Evolutionary Trace traces the evolution of functional residues[52]. The homologous sequences for the human CaMKIIγ AD (PDB:2UX0) and CaMKIIα (PDB:3SOA) were downloaded, clustered, and used for dendrogram construction. Mutations localized at splits in the dendrogram identified possible functional sites. A contiguous patch of such residues identified a functional surface. The real value ET ($rvET$) score integrates the entropy and dendrogram location of each residue position in the MSA weighted for evolutionary distance.

$$rvET_i = 1 + \sum_{n=1}^{N=1} w_{node}(n) \sum_{g=1}^{n} w_{group}(g) xs_i \qquad \text{Equation 2}$$

where w_{node} and w_{group} are the phylogenetic tree nodes and tips, respectively. The s_i is the information entropy that measures the frequency of occurrence, (f_{ia}), of amino acid a in residue position i within the MSA.

$$s_i = -\sum_{a=1}^{20} f_{ia}\, ln(f_{ia}) \qquad \text{Equation 3}$$

The frustration index, ΔEfr, computes the energies of the native residue contacts relative to the distribution of decoy energies, obtained by randomizing the identities of the residues in the native (ij) contacts with n randomly selected amino acid combinations (h) [53].

$$\Delta Efr_{ij} = \left(\Delta E_{ij}^{N} - \overline{\Delta E_{i'j'}^{D}}\right)/\sqrt{(1/n)\sum_{k=1}^{h}\left(\Delta E_{i'j'}^{D} - \overline{\Delta e_{i'j'}^{D}}\right)^2} \qquad \text{Equation 4}$$

The native contact is "minimally frustrated" if its energy ΔE_{ij}^{N} is at the lower end of $\Delta E_{i,j}^{D}$, decoy energy distribution (mean $\overline{\Delta E_{ij}^{D}}$). The contact is "highly-frustrated" if the converse is true. Contacts with an index higher than 0.78 and lower than -1 are taken as minimally frustrated and highly frustrated, respectively. A case study of the integration of $rvET$ and ΔEfr to understand protein design is provided for calmodulin[54].

Protein Dynamics: The monomer subunit A, the tetramer complex (subunits ACGI) and the ACGI tetramer AD human were extracted in silico from the human CaMKII holoenzyme structure (PDB:3SOA). The tetramer contained all lateral and vertical dimer contacts represented in the intact holoenzyme. Conformational ensembles of these structures were generated in Gromacs (*www.gromacs.org*[55]) with tCONCOORD[56] with a 2.2 solvation score, as described previously [57]. The conformer 3D structures were encoded as a 1D-string of four-residue fragments with a structural alphabet based on representative fragment states (letters) determined from frequently occurring conformations in 798 high-resolution X-ray structures [58]. The resulting array of 1D strings was used to derive a network of dynamic couplings based on normalized mutual information (nMI) with GSATools[59]). The correlation of conformational changes in a pair of protein segments *(i, j)* was calculated as normalized mutual information (nMI) between the associated columns in the structural string alignment.

$$nMI(C_i; C_j) = (I(C_i; C_j) - \varepsilon(C_i; C_j))/H_{C_iC_j} \quad \text{Equation 5}$$

where $C_i$and C_j are the relevant columns in the 1D string alignment, $I(C_i; C_j)$ is the mutual information between them, $H_{C_iC_j}$is the joint entropy, and $\varepsilon(C_i; C_j)$is the expected finite-size error. The top couplings (nMI > 0.15) were mapped on the 3D structures with a Pymol plugin. The contribution of a node to the network scaled with its connectivity, estimated by the eigenvector centrality, *E*, calculated directly from the correlation matrix:

$$E.(M)_{corr} = E.\lambda \quad \text{Equation 6}$$

where $(M)_{corr}$ is the correlation matrix and λ the corresponding eigenvalue.

The essential collective motions were obtained by PCA [60]. PCs were generated by diagonalization of the covariance matrix of C^{α} positions derived from the tCONCOORD ensembles. The variance of the states was taken as a measure of "motion" with the first few PCs representing "slow" larger amplitude motions than those recorded by the later PCs on a relative, though not absolute, timescale.

Table 1 lists the software and algorithms used.

Quantification and Statistical Analysis

Phylogenetics and Evolution: The local support used by FastTree, instead of traditional bootstrap values, is the estimation based on 1000 trials of the best probability of each split as assessed by the minimal evolution criterion [61]. The GREMLIN analysis of the HHblits AD MSA identified 186 coevolved couplings above the significance threshold (132.9 = (sequence number (16,485)/sequence length (131))[49], of which the top 30 (> 0.995 significance) were mapped onto the crystal structure. The Pearson's coefficient was used to assess the similarity between KD and AD tree topologies.

TABLE 1: Software and Algorithms

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CCP4	[50]	*www.ccp4.ac.uk/*
CD-Hit	[45]	*weizhongli-lab.org/cd-hit*
Cytoscape	-	*cytoscape.org*
DALI	[51]	*ekhidna2.biocenter.helsinki.fi/dali*
EV-Trace	[52]	*evolution.lichtargelab.org*
FastTree	[61]	*www.microbesonline.org/fasttree/*
FigTree	-	*evomics.org/resources/software/molecular-evolution-software/figtree*
Frustratometer	[53]	*frustratometer.qb.fcen.uba.ar*
GREMLIN	[49]	*www.gremlin.org*
Gromacs	[55]	*www.gromacs.org*
GSATools	[59]	*pandinilab.org/gsatools.html*
MUSCLE	[62]	*www.drive5.com/muscle/*
Pfam	[43]	*www.Pfam.org*
Protein Data Bank	[47]	*www.rcsb.org*
Pymol	-	*www.pymol.org*
tCONCOORD	[56]	*www3.mpibpc.mpg.de/groups/de_groot/dseelig/tconcoord.html*
Uniprot	[44]	*www.UniProt.org*

Dynamics: 66,536 (16^4) equilibrium conformations were generated for the monomer and tetramer structures extracted from PDB:3SOA. The overlap between ensemble subsets was >99% when subset size was < ¼ of the total ensemble, as reported previously for CaMKII kinase domain structures[63]. The top network couplings mapped onto the AD crystal structure represented pairs above the 2σ significance threshold in the distribution obtained after correction for the finite size error.

Data and Code Availability

The tree dendrograms, the GREMLIN job (ID 1592362472), subsets of the tCONCOORD ensembles, and the GSATools network files have been deposited in Mendeley (*https://www.mendeley.com/reference-manager/library/collections/d81a4fb0-c1d5-4ee1-8a81-d31e0a34575d/).* The PCA trajectories and structural models have been uploaded as Supplementary information.

References

1 Lisman, J. The CaM kinase II hypothesis for the storage of synaptic memory. *Trends Neurosci.* **17**, 406-412, doi:0166-2236(94)90014-0 [pii] (1994).

2 Bayer, K. U. & Schulman, H. CaM Kinase: Still Inspiring at 40. *Neuron* **103**, 380-394, doi:10.1016/j.neuron.2019.05.033 (2019).

3 Brocke, L., Srinivasan, M. & Schulman, H. Developmental and regional expression of multifunctional Ca2+/calmodulin-dependent protein kinase isoforms in rat brain. *J. Neurosci.* **15**, 6797-6808 (1995).

4 Tombes, R. M., Faison, M. O. & Turbeville, J. M. Organization and evolution of multifunctional Ca(2+)/CaM-dependent protein kinase genes. *Gene* **322**, 17-31, doi:10.1016/j.gene.2003.08.023 (2003).

5 O'Leary, H., Lasda, E. & Bayer, K. U. CaMKIIbeta association with the actin cytoskeleton is regulated by alternative splicing. *Mol. Biol. Cell* **17**, 4656-4665 (2006).

6 Bhattacharyya, M. *et al.* Flexible linkers in CaMKII control the balance between activating and inhibitory autophosphorylation. *eLife* **9**, doi:10.7554/eLife.53670 (2020).

7 Kaiser, T. S. *et al.* The genomic basis of circadian and circalunar timing adaptations in a midge. *Nature* **540**, 69-73, doi:10.1038/nature20151 (2016).

8 GuptaRoy, B. *et al.* Alternative splicing of Drosophila calcium/calmodulin-dependent protein kinase II regulates substrate specificity and activation. *Brain Res. Mol. Brain Res.* **80**, 26-34, doi:10.1016/s0169-328x(00)00115-7 (2000).

9 Bhattacharyya, M. *et al.* Molecular mechanism of activation-triggered subunit exchange in Ca(2+)/calmodulin-dependent protein kinase II. *eLife* **5**, doi:10.7554/eLife.13405 (2016).

10 Stratton, M. *et al.* Activation-triggered subunit exchange between CaMKII holoenzymes facilitates the spread of kinase activity. *eLife* **3**, e01610, doi:10.7554/eLife.01610 (2014).

11 Sloutsky, R. *et al.* Heterogeneity in human hippocampal CaMKII transcripts reveals allosteric hub-dependent regulation. *Sci Signal* **13**, doi:10.1126/scisignal.aaz0240 (2020).

12 Torres-Ocampo, A. P. *et al.* Characterization of CaMKIIalpha holoenzyme stability. *Protein Sci.* **29**, 1524-1534, doi:10.1002/pro.3869 (2020).

13 Fenwick, R. B., Orellana, L., Esteban-Martin, S., Orozco, M. & Salvatella, X. Correlated motions are a fundamental property of beta-sheets. *Nat Commun* **5**, 4070, doi:10.1038/ncomms5070 (2014).

14 Hoelz, A., Nairn, A. C. & Kuriyan, J. Crystal structure of a tetradecameric assembly of the association domain of Ca2+/calmodulin-dependent kinase II. *Mol. Cell* **11**, 1241-1251, doi:10.1016/s1097-2765(03)00171-0 (2003).

15 Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275-282 (2002).

16 Huang, T. *et al.* Crystal structure of SgcJ, an NTF2-like superfamily protein involved in biosynthesis of the nine-membered enediyne antitumor antibiotic C-1027. *J. Antibiot. (Tokyo).* **69**, 731-740, doi:10.1038/ja.2016.88 (2016).

17 Eberhardt, R. Y. *et al.* Filling out the structural map of the NTF2-like superfamily. *BMC Bioinformatics* **14**, 327, doi:10.1186/1471-2105-14-327 (2013).

18 McSpadden, E. D. *et al.* Variation in assembly stoichiometry in non-metazoan homologs of the hub domain of Ca(2+) /calmodulin-dependent protein kinase II. *Protein Sci.* **28**, 1071-1082, doi:10.1002/pro.3614 (2019).

19 Chevrette, M. G. *et al.* Taxonomic and Metabolic Incongruence in the Ancient Genus Streptomyces. *Front Microbiol* **10**, 2170, doi:10.3389/fmicb.2019.02170 (2019).

20 Pazos, F. & Valencia, A. Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng.* **14**, 609-614, doi:10.1093/protein/14.9.609 (2001).

21 Loonen, A. J. M. & Ivanova, S. A. Circuits regulating pleasure and happiness: evolution and role in mental disorders. *Acta Neuropsychiatr* **30**, 29-42, doi:10.1017/neu.2017.8 (2018).

22 Baldwin, M. K., Cooke, D. F. & Krubitzer, L. Intracortical Microstimulation Maps of Motor, Somatosensory, and Posterior Parietal Cortex in Tree Shrews (Tupaia belangeri) Reveal Complex Movement Representations. *Cereb. Cortex* **27**, 1439-1456, doi:10.1093/cercor/bhv329 (2017).

23 Thaler, C., Koushik, S. V., Puhl, H. L., 3rd, Blank, P. S. & Vogel, S. S. Structural rearrangement of CaMKIIalpha catalytic domains encodes activation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6369-6374, doi:10.1073/pnas.0901913106 (2009).

24 Sarkar, P. *et al.* Deciphering CaMKII Multimerization Using Fluorescence Correlation Spectroscopy and Homo-FRET Analysis. *Biophys. J.* **112**, 1270-1281, doi:10.1016/j.bpj.2017.02.005 (2017).

25 Rellos, P. *et al.* Structure of the CaMKIIdelta/calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLoS biology* **8**, e1000426 (2010).

26 Mikhailovsky, G. & Gordon, R. Shuffling type of biological evolution based on horizontal gene transfer and the biosphere gene pool hypothesis. *Biosystems* **193-194**, 104131, doi:10.1016/j.biosystems.2020.104131 (2020).

27 Stancik, I. A. *et al.* Serine/Threonine Protein Kinases from Bacteria, Archaea and Eukarya Share a Common Evolutionary Origin Deeply Rooted in the Tree of Life. *J. Mol. Biol.* **430**, 27-32, doi:10.1016/j.jmb.2017.11.004 (2018).

28 Pereira, S. F., Goss, L. & Dworkin, J. Eukaryote-like serine/threonine kinases and phosphatases in bacteria. *Microbiol. Mol. Biol. Rev.* **75**, 192-212, doi:10.1128/MMBR.00042-10 (2011).

29 Levy, J. *et al.* A putative Ca2+ and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* **303**, 1361-1364, doi:10.1126/science.1093038 (2004).

30 Burkhardt, P. *et al.* Evolutionary insights into premetazoan functions of the neuronal protein homer. *Mol. Biol. Evol.* **31**, 2342-2355, doi:10.1093/molbev/msu178 (2014).

31 Bozorgmehr, T., Ardiel, E. L., McEwan, A. H. & Rankin, C. H. Mechanisms of plasticity in a Caenorhabditis elegans mechanosensory circuit. *Front Physiol* **4**, 88, doi:10.3389/fphys.2013.00088 (2013).

32 Murray, E. A., Wise, S. P. & Graham, K. S. Representational specializations of the hippocampus in phylogenetic perspective. *Neurosci. Lett.* **680**, 4-12, doi:10.1016/j.neulet.2017.04.065 (2018).

33 Romano, D. R., Pharris, M. C., Patel, N. M. & Kinzer-Ursem, T. L. Competitive tuning: Competition's role in setting the frequency-dependence of Ca2+-dependent proteins. *PLoS Comput Biol* **13**, e1005820, doi:10.1371/journal.pcbi.1005820 (2017).

34 Derrick, B. E. & Martinez, J. L., Jr. Frequency-dependent associative long-term potentiation at the hippocampal mossy fiber-CA3 synapse. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10290-10294, doi:10.1073/pnas.91.22.10290 (1994).

35 Wang, Q. *et al.* Assemblies of calcium/calmodulin-dependent kinase II with actin and their dynamic regulation by calmodulin in dendritic spines. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18937-18942, doi:10.1073/pnas.1911452116 (2019).

36 Khan, S., Downing, K. H. & Molloy, J. E. Architectural Dynamics of CaMKII-Actin Networks. *Biophys. J.* **116**, 104-119, doi:10.1016/j.bpj.2018.11.006 (2019).

37 Khan, S. *et al.* Sequestration of CaMKII in dendritic spines in silico. *J. Comput. Neurosci.* **31**, 581-594 (2011).

38 Khan, S., Conte, I., Carter, T., Bayer, K. U. & Molloy, J. E. Multiple CaMKII Binding Modes to the Actin Cytoskeleton Revealed by Single-Molecule Imaging. *Biophys. J.* **111**, 395-408, doi:10.1016/j.bpj.2016.06.007 (2016).

39 Ma, Q., Nicolau, D. V., Jr., Maini, P. K., Berry, R. M. & Bai, F. Conformational spread in the flagellar motor switch: a model study. *PLoS Comput Biol* **8**, e1002523, doi:10.1371/journal.pcbi.1002523 (2012).

40 Rosenberg, O. S., Deindl, S., Sung, R. J., Nairn, A. C. & Kuriyan, J. Structure of the autoinhibited kinase domain of CaMKII and SAXS analysis of the holoenzyme. *Cell* **123**, 849-860, doi:10.1016/j.cell.2005.10.029 (2005).

41 Chao, L. H. *et al.* Intersubunit capture of regulatory segments is a component of cooperative CaMKII activation. *Nature structural & molecular biology* **17**, 264-272, doi:10.1038/nsmb.1751 (2010).

42 Nguyen, T. A. *et al.* Covert Changes in CaMKII Holoenzyme Structure Identified for Activation and Subsequent Interactions. *Biophys. J.* **108**, 2158-2170, doi:10.1016/j.bpj.2015.03.028 (2015).

43 El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res* **47**, D427-D432, doi:10.1093/nar/gky995 (2019).

44 UniProt, C. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* **47**, D506-D515, doi:10.1093/nar/gky1049 (2019).

45 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152, doi:10.1093/bioinformatics/bts565 (2012).

46 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

47 Burley, S. K. *et al.* RCSB Protein Data Bank: Sustaining a living digital data resource that enables breakthroughs in scientific research and biomedical education. *Protein Sci.* **27**, 316-330, doi:10.1002/pro.3331 (2018).

48 Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-350, doi:10.1093/nar/gkw408 (2016).

49 Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030, doi:10.7554/eLife.02030 (2014).

50 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).

51 Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**, 128-140, doi:10.1002/pro.3749 (2020).

52 Mihalek, I., Res, I. & Lichtarge, O. A family of evolution-entropy hybrid methods for ranking protein residues by importance. *J. Mol. Biol.* **336**, 1265-1282, doi:10.1016/j.jmb.2003.12.078 (2004).

53 Parra, R. G. *et al.* Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. *Nucleic Acids Res* **44**, W356-360, doi:10.1093/nar/gkw304 (2016).

54 Tripathi, S., Waxham, M. N., Cheung, M. S. & Liu, Y. Lessons in Protein Design from Combined Evolution and Conformational Dynamics. *Sci Rep* **5**, 14259, doi:10.1038/srep14259 (2015).

55 Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845-854, doi:10.1093/bioinformatics/btt055 (2013).

56 Seeliger, D., Haas, J. & de Groot, B. L. Geometry-based sampling of conformational transitions in proteins. *Structure* **15**, 1482-1492, doi:10.1016/j.str.2007.09.017 (2007).

57 Pandini, A., Morcos, F. & Khan, S. The Gearbox of the Bacterial Flagellar Motor Switch. *Structure* **24**, 1209-1220, doi:10.1016/j.str.2016.05.012 (2016).

58 Pandini, A., Fornili, A., Fraternali, F. & Kleinjung, J. Detection of allosteric signal transmission by information-theoretic analysis of protein dynamics. *FASEB J.* **26**, 868-881, doi:10.1096/fj.11-190868 (2012).

59 Pandini, A., Fornili, A., Fraternali, F. & Kleinjung, J. GSATools: analysis of allosteric communication and functional local motions using a structural alphabet. *Bioinformatics* **29**, 2053-2055, doi:10.1093/bioinformatics/btt326 (2013).

60 Amadei, A., Linssen, A. B., de Groot, B. L., van Aalten, D. M. & Berendsen, H. J. An efficient method for sampling the essential subspace of proteins. *J. Biomol. Struct. Dyn.* **13**, 615-625, doi:10.1080/07391102.1996.10508874 (1996).

61 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS one* **5**, e9490, doi:10.1371/journal.pone.0009490 (2010).

62 Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113, doi:10.1186/1471-2105-5-113 (2004).

63 Pandini, A., Schulman, H. & Khan, S. Conformational coupling by trans-phosphorylation in calcium calmodulin dependent kinase II. *PLoS Comput Biol* **15**, e1006796, doi:10.1371/journal.pcbi.1006796 (2019).

Acknowledgements

I thank Drs Howard Schulman and Thomas Reese for their comments on the manuscript. This study was supported, in part, by the NIH, NINDS intramural research program. It utilized the NIH HPC Biowulf computational cluster (*http://hpc.nih.gov*).

Author Contribution

S.K. conceived the idea, did the work and wrote the manuscript

Competing Interest

The author declares no competing interest.

Figures

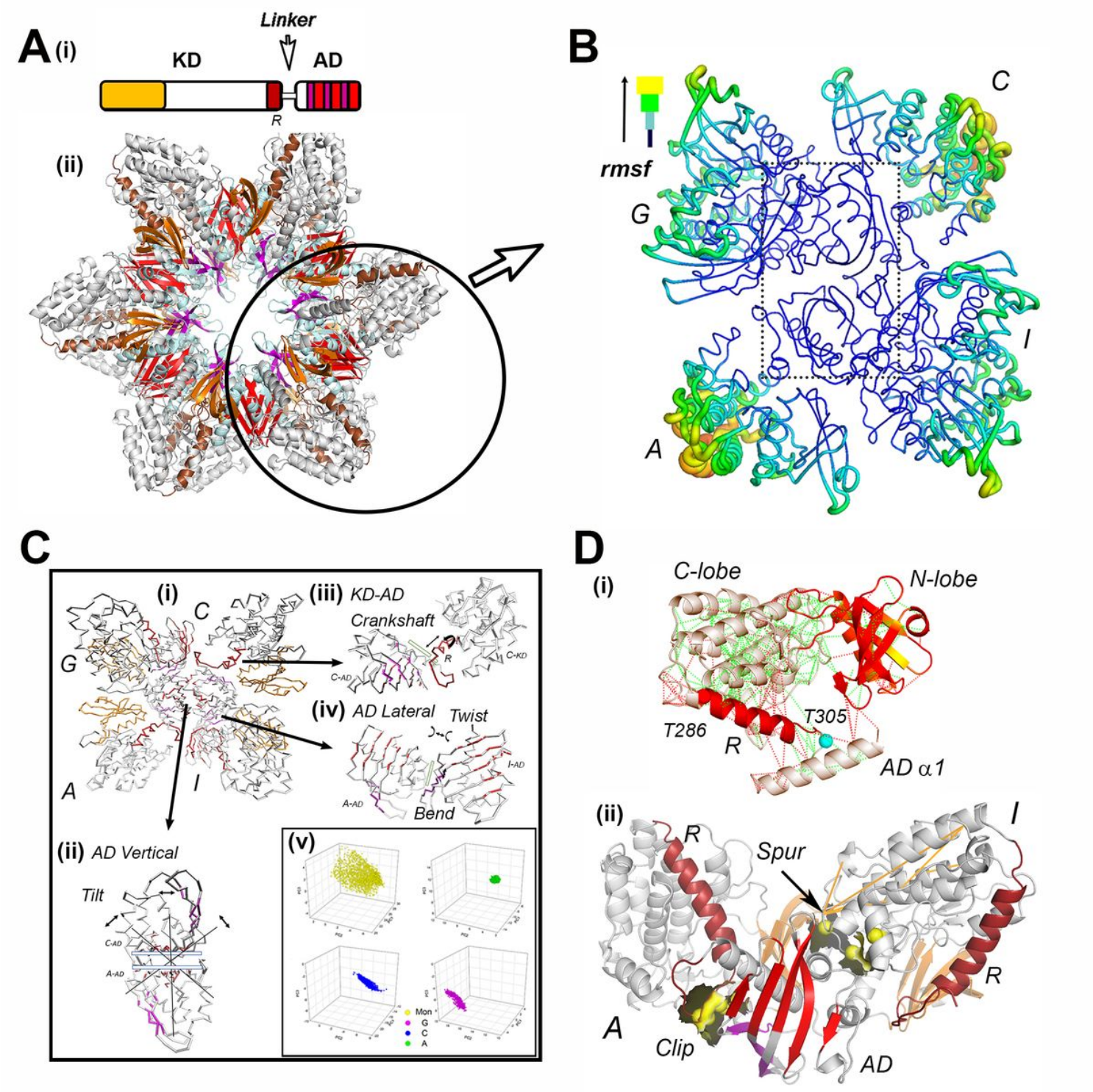


Figure 1

A. Architecture. (i) Subunit. Disordered linkers with varying length and composition connect the kinase domain (KD (N-lobe (orange), C-lobe (white)) with the association domain (AD). The pseudo-substrate, regulatory segment (R (brown) binds Ca2+CAM. The AD β-sheet forms vertical (red) and lateral (magenta) holoenzyme contacts. (ii) Assembly. The ADs form the central hub in the multi-subunit holoenzyme

(CaMKII α PDB:3SOA). A tetramer (circle) was extracted for analysis of conformational fluctuations. B. Flexure. The flexibility (rmsf) profile derived from the tetramer conformational ensemble. C. Collective Motions. (i) A conformation in the CaMKII α tetramer tCONCOORD ensemble (Supplementary Video S1). (ii) Vertical dimer. Tilt. (iii) Lateral dimer. Bend and twist. (iv) KD-AD. Crankshaft (extension + rotation). Rectangles represent β sheet long axes. (v). PCIPC2PC3 plots of the human CaMKIIα AD; (i) monomer and (ii-iv) the tetramer (Supplementary Video S2). D. KD-AD Coupling. (i) R contacts. Energy frustration - (stable (green), stressed (red)). (ii) Dynamic network. The KD-AD contact residues and surface (yellow) are shown. Dynamic couplings (orange lines). 3D-views in Supplementary Video S3-4.

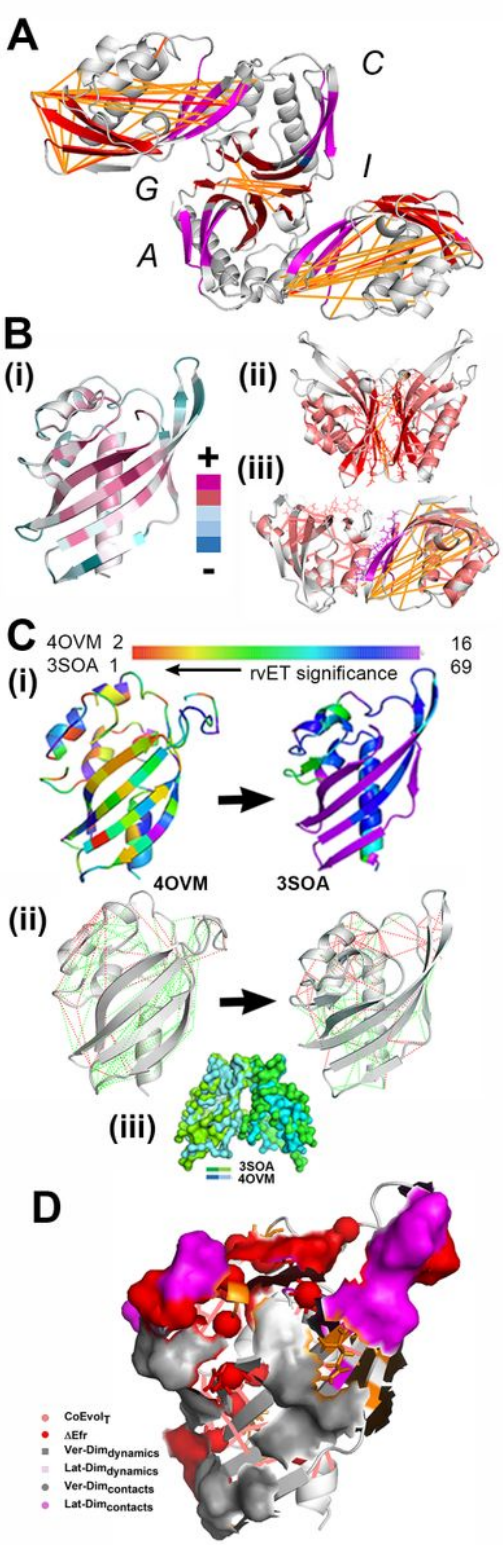


Figure 2

The dynamics and evolution of the CaMKII AD. A. Tetramer-ACGI AD dynamics. The top dynamic couplings computed between 4-residue fragments (yellow (weak) -> orange -> red (strong)). B. Evolution. (i) Residue conservation. (ii, iii) The superposition of the dynamic (thin orange) and coevolved (thick salmon) couplings adjacent to the (ii) vertical and (iii) lateral AD contacts. C. Structural evolution. (i) The evolutionary trace for structures of an ancient bacterial enzyme (PDB;40VM) versus human CaMKIIα

(PDB:3SOA). (ii) Residue contact energetics of the 4OVM and 3SOA structures. (ii) Conservation of the dimer contact (RMSD = 0.69 nm). D. The 3SOA-AD structure with the superposed interdomain contacts, energetically frustrated contacts, dynamic and coevolved couplings. 3D-views in Supplementary Video S5-10.

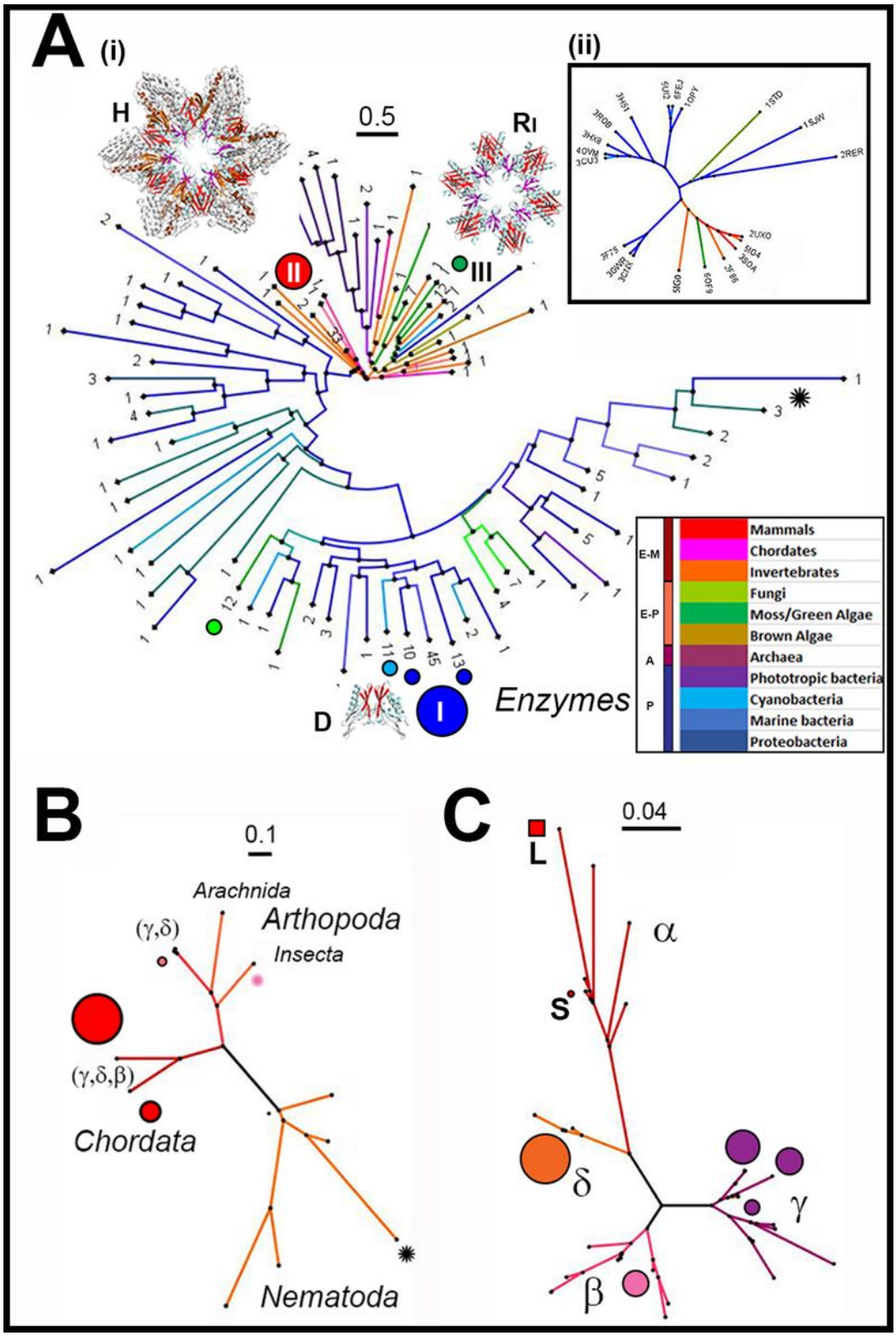


Figure 3

CaMKII phylogenetics. A.(i) CaMKII AD superfamily tree. The colour-coded tree (E-M (metazoan), E-P (protozoa), A (archaea), P (prokaryotes)) was constructed from sequences representing 81 PF08332. The major nodes are composed of bacteria (I (P. n =45)), vertebrate (II (E-M. n = 33)) and diatom/fungal (III (E-P. n =)) clusters, followed by four smaller (3 bacterial and a fungal node in a large group that includes node-1. The AD assemblies associated with nodes I-III (D (dimer, b contact (red)), Ri (ring, lateral β contact (magenta)), H (holoenzyme. R (brown, KD C-lobe (orange))) are shown. The asterisk indicates a possible common ancestor. (ii) Structural phylogenetics. Tree of 23 crystal structures of the CaMKII-AD homologs based on DALI scores (colour-coded as in (i). B. Metazoan CaMKII-AD evolution. Tree based on clustering of 1000 homologs of the C. elegans CaMKII (black asterisk). Midge (pink asterisk). C. Isoform evolution. Tree based on clustering of 2000 homologs of the human CaMKII α and β isoforms. Circles mark major clusters (colour = phylum (A, B), isoform (C); diameter = membership), except for the most distant, larger CaMKIIa node (L. square).

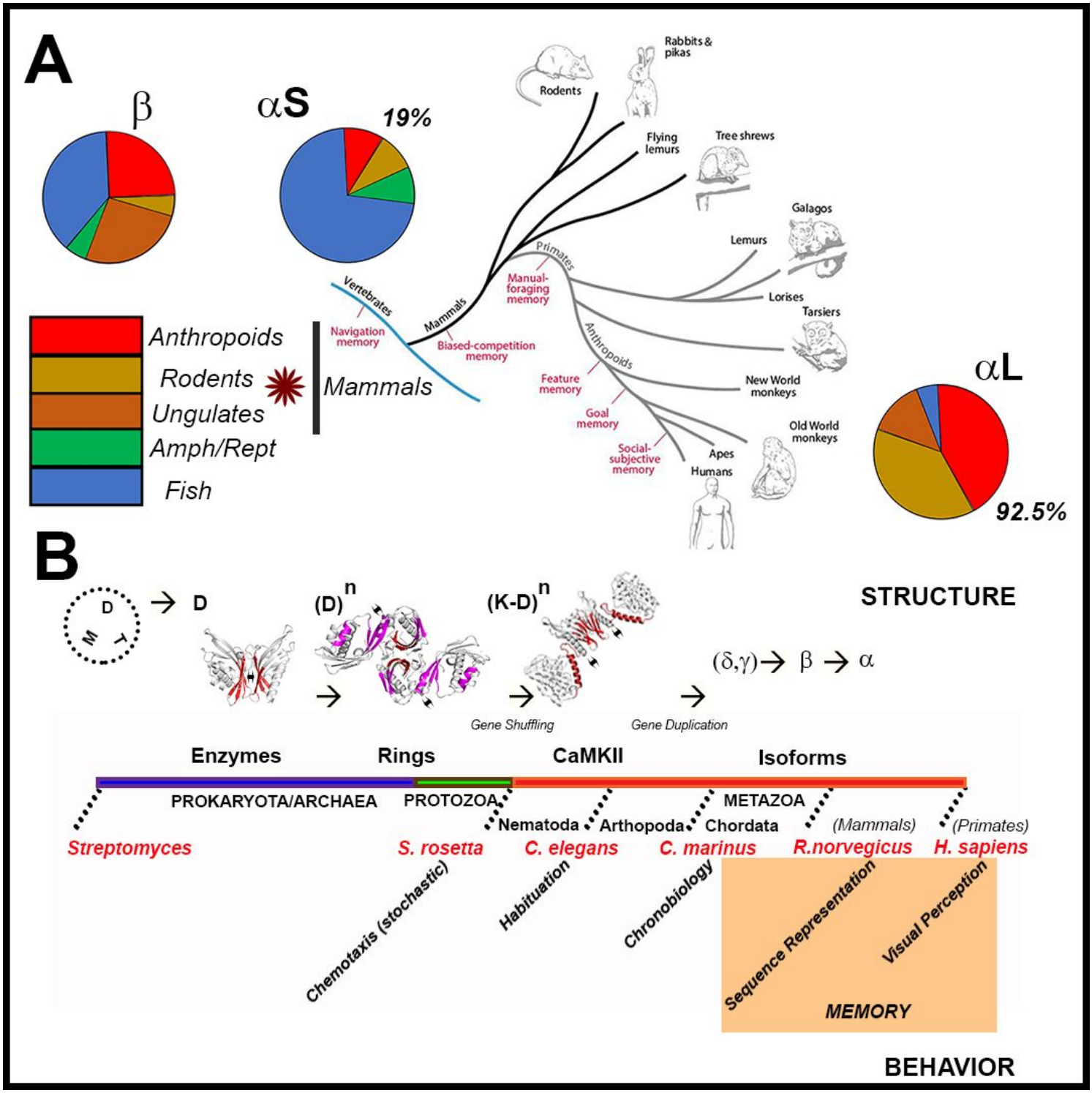


Figure 4

A. CaMKIIα and Memory. The major CaMKIIα phylogenetic nodes straddle the evolution of memory (tree from22 with permission). Homeotherm (mammals (brown asterisk)) fractions for the nodes (I, II) are shown. B. The correlation between CaMKII architecture and behaviour. Architecture. The selection of the dimer (D) from other structures (M = monomer, T = tetramer) in bacteria seeded the emergence of ringed hub assemblies. The fusion with the kinase (K) domains coincided with the emergence of multicellularity (secondary structures colour coded as in Fig. 1). Diversity was created by linker alternative splicing and

enhanced by the generation of isoforms. Behaviour. Work on model organisms suggests CaMKII evolution peaks with the development of cognitive memory.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf