scientific reports

Check for updates

OPEN

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

Shahid Khan

The calcium calmodulin (Ca^{2+}/CaM) dependent protein kinase II (CaMKII) decodes Ca^{2+} frequency oscillations. The CaMKIIα isoform is predominantly expressed in the brain and has a central role in learning. I matched residue and organismal evolution with collective motions deduced from the atomic structure of the human CaMKIIα holoenzyme to learn how its ring architecture abets function. Protein dynamic simulations showed its peripheral kinase domains (KDs) are conformationally coupled via lateral spread along the central hub. The underlying β-sheet motions in the hub or association domain (AD) were deconvolved into dynamic couplings based on mutual information. They mapped onto a coevolved residue network to partition the AD into two distinct sectors. A second, energetically stressed sector was added to ancient bacterial enzyme dimers for assembly of the ringed hub. The continued evolution of the holoenzyme after AD–KD fusion targeted the sector's ring contacts coupled to the KD. Among isoforms, the α isoform emerged last and, it alone, mutated rapidly after the poikilotherm–homeotherm jump to match the evolution of memory. The correlation between dynamics and evolution of the CaMKII AD argues single residue substitutions fine-tune hub conformational spread. The fine-tuning could increase CaMKIIα Ca^{2+} frequency response range for complex learning functions.

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 holoenzymes decode Ca^{2+} pulses[2]. The multi-subunit holoenzyme architecture, a two-ring stack of subunits with mirror symmetry[3], is unique among members of the large calmodulin-dependent kinase (CaMK) family. Recent reviews have detailed the relationship of CaMKII to the CaMK family[4] and its role in synaptic function[4,5].

In all CaMKII isoforms (α,β,γ,δ), individual subunits consist of a canonical kinase domain (KD) with a C-terminal pseudo-substrate regulatory segment adjacent to a CaM-binding domain (R) connected via flexible linkers to an association domain (AD) that forms the central hub. The Ca^{2+}/CaM triggered dissociation and capture of R by an adjacent KD enables trans-phosphorylation of R residue T286 in CaMKIIα (T287 in other isoforms). Activation by threonine trans-phosphorylation is common to all isoforms. CaMKIIα is the dominant isoform in the brain where αβ heterooligomers form by expression of the β isoform. Single polymorphisms in the human α and β isoforms result in intellectual disabilities[6–8]. Activation by T286 trans-phosphorylation occurs in CaMKII dimer constructs as well as in holoenzymes[9,10]. The study of isoform variability has focused on the alternatively spliced KD–AD linkers that modulate tissue localization, interactions with the actin cytoskeleton and the balance between activating and inhibitory phosphorylation[11–16]. Variations in the conserved AD responsible for ring assembly were not as well analyzed until recently. We now know that kinase activity is modulated by endogenous ligands that bind to the hub[17] and determined by differences between α and β isoform hubs as established by chimeric constructs[18]. Four of the six CaMKIIα variants found in patients with schizophrenia localized to the AD[8]. A CaMKIIβ hub deletion has revealed isoform-specific differences in holoenzyme formation[19]. Finally, human CaMKIIα R peptides interact with and destabilize their hubs[20], consistent with activation dependent subunit exchange[21,22]. Here, motivated by these studies, I have analyzed hub dynamics and evolution to evaluate the importance of the ringed architecture for CaMKII function.

[1]Molecular Biology Consortium, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [2]SBA School of Science and Engineering, LUMS, Lahore, Pakistan. [3]Laboratory of Cell Biology, NINDS, NIH, Bethesda, MD 20892, USA. email: smkhan@lbl.gov

An atomic structure of the human CaMKIIα holoenzyme is available, together with crystal structures of the CaMKII AD superfamily from many microbial and metazoan species. In addition, many amino acid sequences from this family have been deposited. I used a sub-assembly of the CaMKIIα holoenzyme to generate a conformational ensemble in silico. I analyzed this ensemble to map the long-range conformational spread between KDs via the AD hub. Residue contacts were analyzed to determine energetically stressed sectors. The increased energetic cost is compensated by enhanced function, typically binding propensity[23], as illustrated by a protein databank survey that reported stabilization of energetically stressed contacts clustered near protein–protein association interfaces by complex formation[24]. Multiple sequence alignment (MSA) of the CaMKII AD sequences identified coevolved residue contacts also shown to be important for function[25]. The most strongly coevolved contacts localized to the AD interfaces. I further compared phylogeny based on the MSA with that based on crystal structures and sensory behavior. I found that, first, variations of eukaryotic sequences most closely related to the human CaMKIIα AD mapped to the lateral ring contacts rather than the primordial dimer contact that was targeted in the bacterial sequences. Second, the behavioral phylogeny based on learning assays correlated with the sequence phylogeny behavior of the α, but not the other isoforms. I conclude that the AD hub propagates lateral conformational spread, based on the dynamics, and that the residue contacts that mediate this spread are important for function, based on the energetics and evolution. I propose that the distinct evolutionary trajectory of the α isoform reflects tuning of conformational spread in the ringed holoenzyme to extend the CaMKII response range to Ca^{2+} pulses for complex learning tasks in the brain.

Results

Hub β-sheet dynamics mediate long-range KD–KD coupling. The crystal structure of the two-stack human CaMKIIα holoenzyme has six subunits per stack (3SOA.PDB). The AD has distinct intra-stack and inter-stack contacts. The holoenzyme may be partitioned into three tetramers that each contain both contacts. In the tetramer ACGI shown in Fig. 1a, subunits A, C form the inter-stack dimer (Vert-Dim) contact. The (A, I) and (C, G) lateral dimers (Lat-Dim) form intra-stack contacts. The KDs also contact the ADs. Intra-subunit ("clip") contacts of the KD CaM-binding domain with its AD and inter-subunit ("spur") contacts of the KD DFG activation loop with the AD of the adjacent lateral subunit are seen in the autoinhibited holoenzyme[26]. Conformational ensembles were generated from the ACGI tetramer with tCONCOORD[27]. In brief, tCONCOORD generates another conformer from the tetramer crystal structure by random displacement of its atoms within limits, followed by iterative correction to eliminate bond violations until all bonding constraints are satisfied and a new structure is obtained (see "Methods" (Protein Dynamics) for parameters). The process is then repeated until an ensemble of the desired size is obtained. The tetramer root-mean-square fluctuation (rmsf) profile reported the mean fluctuations for the C^{α} atoms relative to the reference crystal structure obtained for this tCONCOORD ensemble. It was evident from the rmsf profile that small fluctuations of the hub AD domains were coupled to large motions of the associated KD's (Fig. 1b). The C^{α} atoms of the residue contacts at the Vert-Dim interfacial hinge were more rigid (0.29 ± 0.002 nm) than contacts for the Lat-Dim hinge (0.355 ± 0.016 nm) or the AD–KD "clip" (0.292 ± 0.005 nm) consistent with previous work[21,28].

I used principal component analysis (PCA) to determine the extent and nature of the long-range collective motions (Fig. 1c, Supplementary Video S1). These motions were encoded by the first three principal components (PCs) ("Methods" (Protein Dynamics)) that constituted a large fraction (0.7 = 0.46(PC1) + 0.14 (PC2) + 0.10 (PC3)) of all motions measured for the tetramer tCONCOORD ensemble. First, I estimated constraints due to the formation of the tetramer on the motion of the component subunits. A tCONCOORD ensemble for a monomer extracted from the holoenzyme was generated and similarly analyzed. The PC1PC2PC3 plots showed that the tetramer subunit motions had reduced amplitude, but increased anisotropy relative to the monomer. The monomer had a reduced PC1 fraction (0.34). Second, I used GROMACS geometric functions to map the PC motions from the tetramer ensemble onto the key contacts. These consisted of orthogonal tilt motions at the Vert-Dim contact and anisotropic β_2–β_5 sheet bending and twisting motions at the Lat-Dim contacts. The central hinge was the more rigid Vert-Dim contact rather than the weaker Lat-Dim contact that primarily transmitted the hub β sheet deformations to adjacent subunits. The principal PC1–PC3 components coupled KD–KD twist and rotations. The crankshaft (rotation + extension) KD motion is coupled to the AD–AD β_2–β_6 sheet bending and twisting modes. The measurements taken together reveal the KD–AD-KD coupling mechanism as detailed in Supplementary Fig. S1.

I constructed mutual-information (nMI) based networks to encode the local fragment dynamics of local fragments as 1D-strings for comparison with the molecular evolution metrics derived from the MSAs (Fig. 1d). The network analysis revealed how AD flexibility is constrained in the tetramer relative to the free monomer. 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 Ca^{2+}/CaM binding site located at the other end of the R helix ($R_{300–306}$). The spur KD–AD contact transmits AD motions to rotation-translation of KD helices $\alpha_{6–7,9}$. In the tetramer, 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 the centrality plots consistent with the top network couplings (Supplementary Fig. S1).

The dynamic network reflects hub residue coevolution. I used two tools to correlate local dynamics with molecular evolution ("Methods" (Structure Analysis)). First, I assessed the energetic cost of residue contacts to diagnose metastable binding surfaces as noted in the Introduction. The frustration score (ΔE_{fr}) reported the stabilization energy of the native contact relative to all possible contacts. It was used to partitions contacts into sub-populations with low ("relaxed") or high ("stressed") cost energetics. Second, I scored residue coevolu-

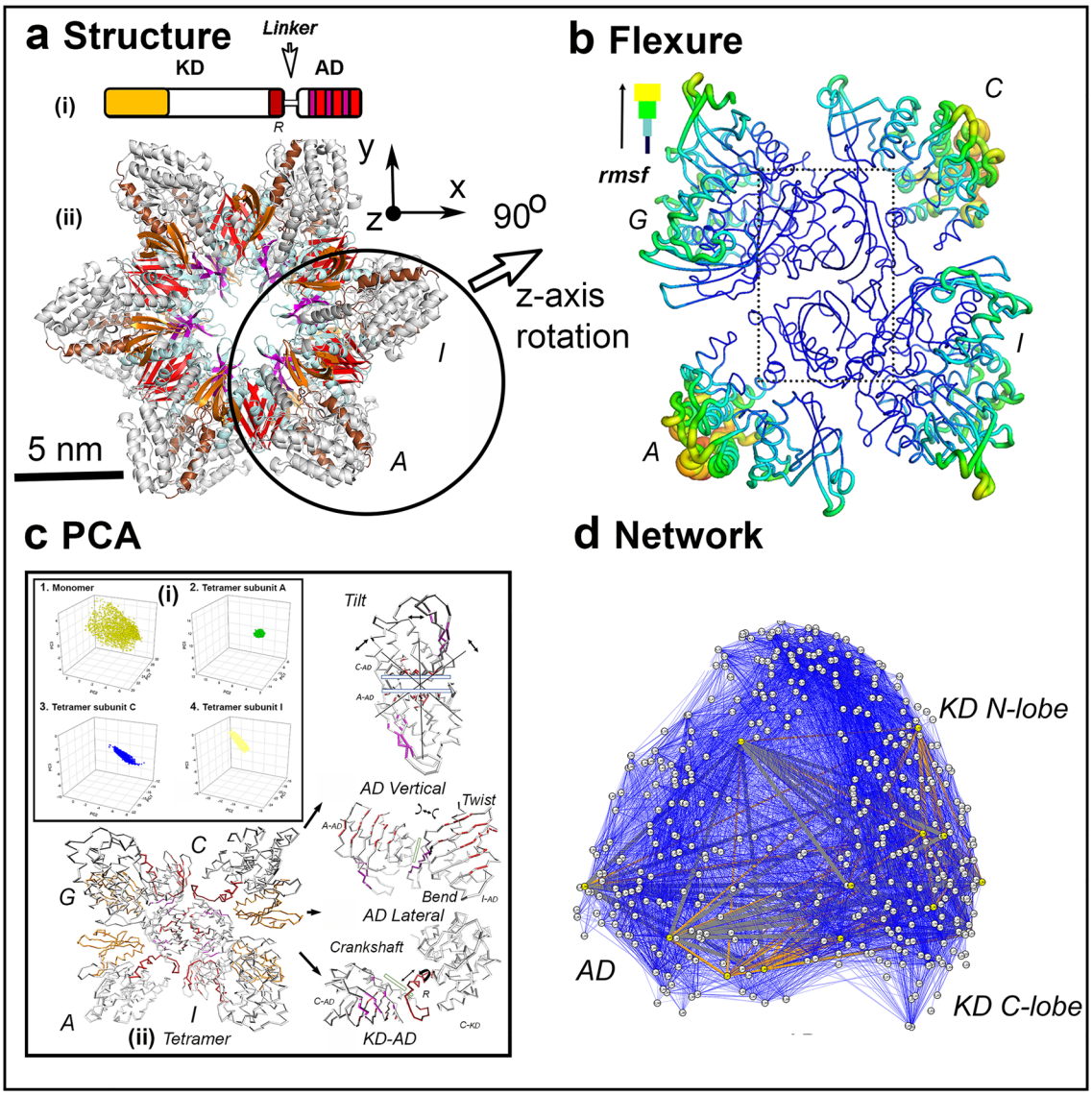


Figure 1. (**a**) Architecture. (i) Subunit. Disordered linkers with varying lengths and composition connect the kinase domain (KD (N-lobe (orange), C-lobe (white)) with the association domain (AD). The pseudo-substrate, regulatory segment (R (brown) binds Ca^{2+}/CaM. The AD β-sheet forms vertical (red) and lateral (magenta) ring contacts. (ii) Assembly. The ADs form the central hub in the multi-subunit holoenzyme (CaMKIIα 3SOA.PDB). A tetramer (circle) was extracted for analysis of conformational fluctuations. (**b**) Flexure. The flexibility (rmsf)) profile derived from the tetramer conformational ensemble. The tetramer orientation shown is rotated by 90° relative to the orientation in panel a (Supplementary Video S1). (**c**) Principal Component Analysis (PCA). (i). PCIPC2PC3 plots of the human CaMKIIα AD; (ii) A conformation in the CaMKIIα tetramer tCONCOORD ensemble (Supplementary Video S2) used for PCA, with PCs 1–3 mapped onto the key contacts (AD Vertical (Vert-Dim), Tilt. AD Lateral (Lat-Dim), Bend and twist. KD–AD. Crankshaft (extension + rotation)). Rectangles represent β sheet long axes. (**d**) Monomer network. Nodes = Residues (Circles). Edges = Dynamic couplings between 4-residue fragments (lines color-coded according to nMI score (high (orange) → low (blue)). Source listed in Table S1.

tion. The score (S_s) represented the evolutionary coupling strength between two residue positions, again indicative of a functionally important contact.

The spur KD–AD contact coupled AD fluctuations to α-helices in the KD C-lobe (Fig. 2a). The contacts made by the regulatory segment R with both the KD C-lobe and AD were energetically stressed based on the ΔE_{fr} scores (Fig. 2b). Interfacial couplings that propagated across and along the two stacks dominated the AD dynamic network. The Vert-Dim dynamic couplings connected the β_2–β_4 sheet center in one AD with the α_2–β_1 and β_3–β_4 loops in the other AD across the rigid (β_3–β_4, β_6 strands, α_2–α_3 loop) contact. The top (1%) dynamic couplings at the Lat-Dim contact of helix α_3 with the adjacent β_3–β_5 loops, changed AD β-sheet curvature (Fig. 2c). The ΔE_{fr} profiles revealed the energetic cost of the evolution of CaMKIIα Lat-Dim contact (Fig. 2d). Its structural elements formed stressed interactions relative to the relaxed contacts associated with the Vert-Dim interface.

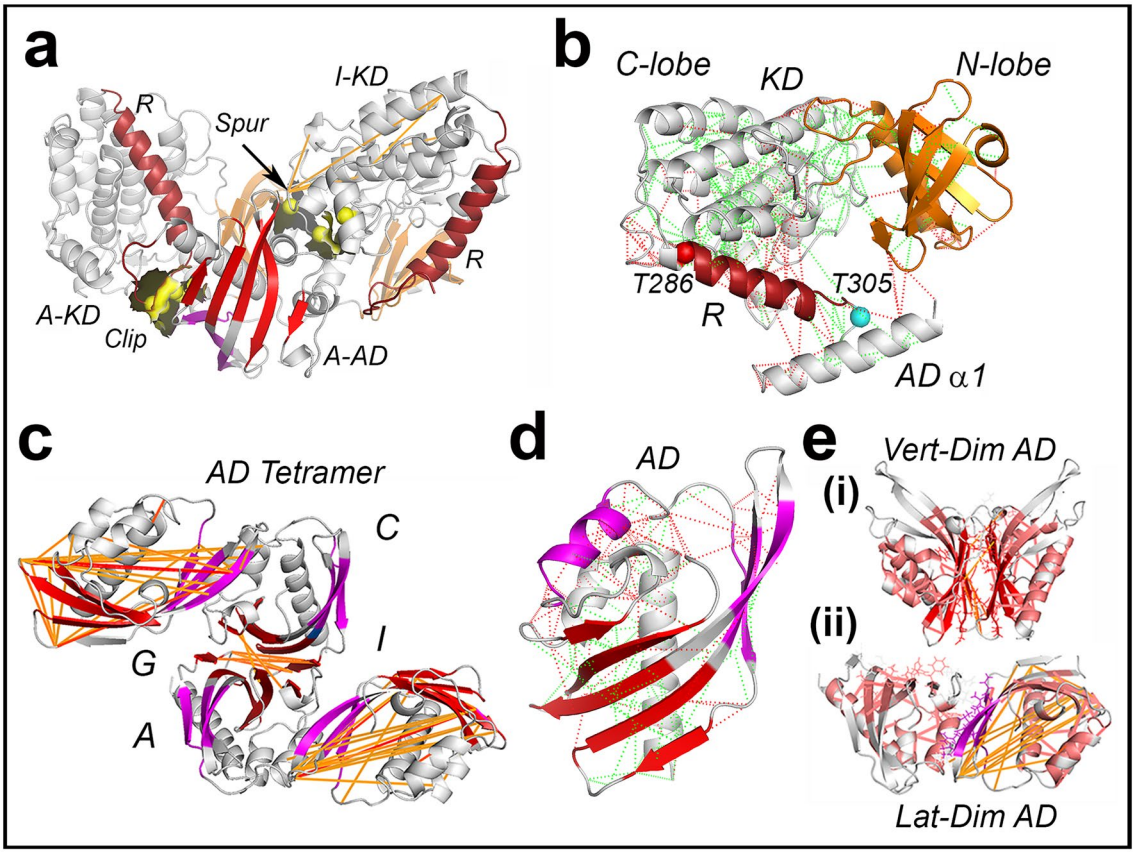


Figure 2. The dynamics and evolution of the CaMKII AD. Hub interfaces and R helix are color-coded as in Fig. 1a. (**a**,**b**). KD–AD Contact. (**a**). Dynamics. The top dynamic couplings computed between 4-residue fragments (yellow (weak) → orange → red (strong)). The KD–AD contact residues and surface (yellow). (**b**) Energetics. Energy frustration—(relaxed (green), stressed (red)). Spheres denote T286 (red), T305 (cyan). (**c**,**d**) AD Fold. (**c**) Interfacial dynamics. ACGI AD tetramer. The dynamic couplings span the interfaces (Vert-Dim (red), Lat-Dim (purple). Supplementary Video S3). (**d**) Energetics. Energy frustration scores are color coded as in B. (**e**) Residue Coevolution. The superposition of the dynamic (thin orange lines) and coevolved (thick salmon lines) couplings adjacent to the (i) Vert-Dim and (ii) Lat-Dim contacts (stick sidechains). 3D-views in Supplementary Videos S4–S7. Source listed in Table S1.

The strongly coevolved residue contacts scaffolded the dynamic couplings as seen in maps of the Vert-Dim and Lat-Dim dimer complexes. The contacts formed two sectors—the larger sector stitched helices α_{2-3} with strands β_{1-3} at the Lat-Dim contact, while the smaller one bonded the central β_{2-5} sheet at the Vert-Dim contact with long helix α_1 (Fig. 2e).

High-throughput genome sequencing and X-ray crystallography[29,30] identified the CaMKII-AD superfamily, sometimes cited as the NTF2 superfamily (PF08332) due to the common fold between the CaMKIIα AD and the dimeric yeast nuclear pore complex component NTF2[31,32]. The PF08332 MSA seeded an expanded sequence set for analysis of residue coevolution. The overlap of the evolution metrics with the complete AD dynamic network is shown in Fig. 3. The composite $\alpha_{2-3}\beta_{1-3}$ contact network and the central β-sheet hinge were the principal drivers of CaMKII-AD coevolution. Coevolved contacts between helix β_1 and the β sheet maintained the cross-section of the hydrophobic core. The long-range mechanical relays coupled the hydrogen-bonded hydrophobic core sidechains with β-sheet curvature[33]. The fluctuations in β-sheet curvature drove lateral intra-stack conformational spread via the Lat-Dim contacts.

The evolution of the CaMKIIα AD targets the Lat-Dim contacts. I studied phylogeny to link structure to speciation, with the tree of life constructed from ribosomal RNA sequences as reference[34]. I superimposed all available crystal structures of the CaMKII-AD superfamily (n = 22), in addition to human CaMKIIα, to understand AD fold evolution. The phylogenetic tree (Fig. 4a) constructed from the DALI superimposition Z-scores demarcated prokaryotic and eukaryotic structures. The dimer was the dominant assembly (n = 11), followed by holoenzymes (n = 7) monomers (n = 2), a heterodimer. trimer and ring. Multiple superfamily members with different oligomeric states were found in *Streptomyces*, an ancient bacterial lineage[35]. 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[33].

The CaMKII-AD fold is a curved β-sheet (β_2–β_6) anchored to a long helix α_1. The PF08332 MSA surface conservation profile identified the Vert-Dim β_2–β_6 interface as the most conserved. The 2D-heatmap revealed that

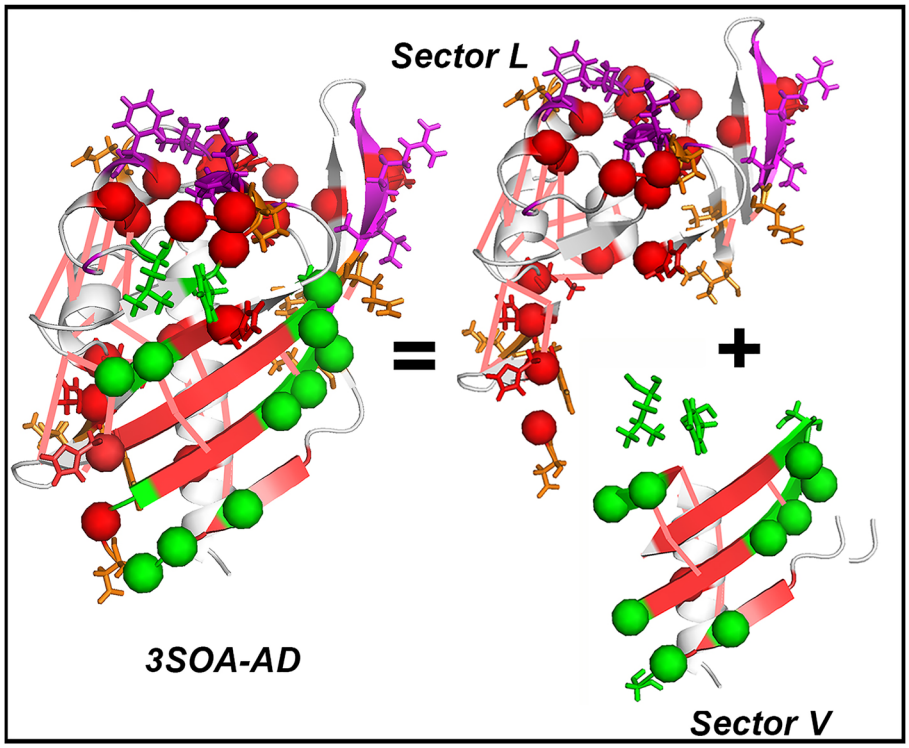


Figure 3. The two sectors of the CaMKIIα AD. The coevolved residue network of the primordial vertical dimer (Sector V) has energetically relaxed residues (green spheres) at the Vert-Dim contact interface (red β strands). The coevolved residue network (thick salmon lines) of the Lat-Dim contact (Sector L) has energetically stressed residues (red spheres) at the contact interface (purple α-helix, β strands). Stick representations denote residues at the interface (purple, green, red) or part of adjacent dynamic couplings (gold). 3D-view in Supplementary Video S8. Source listed in Table S1.

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. A global phylogenetic tree constructed from the PF08332 sequences related structural to organismal evolution. The tree representation was sufficient for this aim even though large evolutionary diversity in single-gene families is more accurately represented as a network[36]. The tree was demarcated into prokaryotic, eukaryotic, and archaeal clusters consistent with the phylogeny based on the crystal structures (Supplementary Fig. S2). The dimer has been recognized as an ancestral assembly module[33], but its evolution based on sequence was not previously tracked from primordial bacteria to humans due to the low homology.

I used Evolutionary Trace (EV trace) to identify selection forces over short evolutionary timescales at two distant stages in the evolution of the superfamily. The EV trace was then mapped onto 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 was the key determinant (rvET = 1.7 ± 1.0), followed by the $R_{286–305}$ segment (rvET = 22.2 ± 1.7), both important for kinase activation. The rvET AD profiles revealed that in bacteria (4OVM.PDB) the fold underwent global evolution. In contrast, AD fold evolution in metazoan CaMKII holoenzymes (3SOA.PDB) was localized to the Lat-Dim contact rather than the Vert-Dim contact. Its rate was modest compared to KD evolution.

Finally, the structures of a *Streptomyces* enzyme (4OVM.PDB) and the human CaMKIIα AD dimer extracted from human CaMKIIα (3SOA.PDB) were superimposed (Fig. 4c). The CaMKIIα AD fold had elongated helix α_1 and β-sheet segments relative to the bacterial enzyme that were utilized to form Lat-Dim contacts. The structural conservation of the Vert-Dim contact was indicated by the root mean square deviation (RMSD) between the common C^{α} backbone atoms of the superimposed structures (0.69 nm) and supported its sequence conservation deduced from the PF08332 MSA. I conclude that the Vert-Dim contact evolved most rapidly in bacteria, becoming fixed during metazoan evolution. The Lat-Dim contact was the focus of continuing evolution of the CaMKII AD fold in primates, in concert with KD evolution.

The poikilotherm–homeotherm transition is a major step in CaMKIIα evolution. The CaMKII holoenzyme structures were too few to trace the evolution of the CaMKIIα isoform, so I used CaMKII sequences instead. First, to understand the emergence of isoforms, I constructed phylogenetic trees from one thousand sequences most homologous to the *Caenorhabditis elegans* CaMKII, an ancient CaMKII with well-characterized structure and biochemistry (Fig. 5a). *C. elegans* lacks isoforms but has alternatively spliced variants. 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 seg-

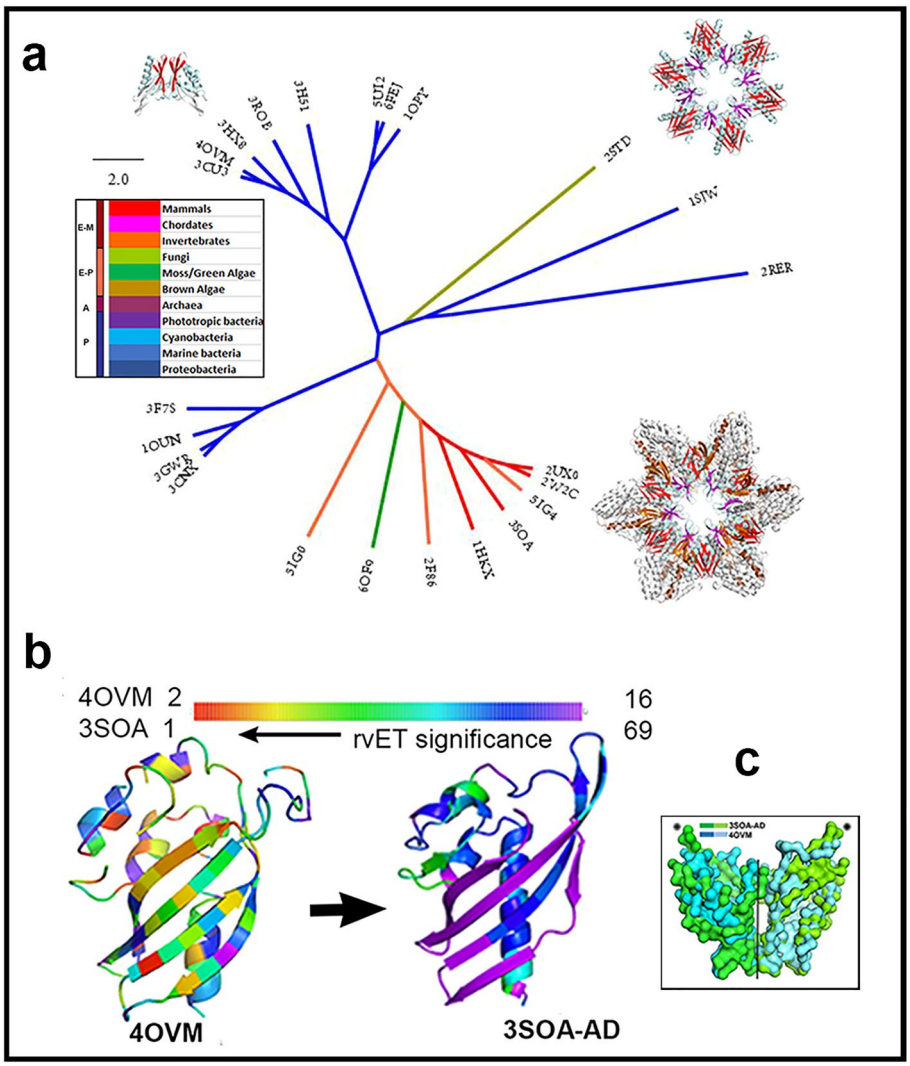


Figure 4. CaMKII phylogenetics (from 3D Structures). (**a**). Tree of 23 crystal structures of the CaMKII-AD homologs (color-coded by phyla) based on DALI scores. (**b**). The evolutionary trace (EV trace) for structures of an ancient bacterial enzyme (40VM.PDB) versus human CaMKIIα (3SOA.PDB). 3D Views in Supplementary Videos S9–S10. (**C**). Conservation of the dimer contact (RMSD = 0.69 nm). Source listed in Table S1.

regated 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. 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 α isoform 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. 5b) in contrast to the *C. elegans* rooted tree. This branching pattern extended to the individual domain tree topologies. The similarity score between the KD and AD trees was ~ 0.6 (Fig. 5c). This score was weaker but comparable to the scores (> 0.8) obtained for proteins with strongly interacting domains such as ribosomal components and the F_1 ATP synthase subunits[37]. Thus, the domain coevolution, reflected in the similarity score is consistent with KD–AD interactions. In addition, the isoform dependent branching pattern shows that the variation of the domains within isoforms is less than between isoforms. This is also the case with the linkers where isoform conservation was more readily detected as the linker sequences are more variable[4]. I speculate that the conservation within isoforms and the domain coevolution may reflect adaptive selection for tissue-specific signal phospho-relays orchestrated by the KD[12].

Analysis of the divergence and species composition of the major nodes of this tree provided insights into isoform evolution. The δ isoform was the least diverse with a single node. Its species composition is partitioned between poikilothermic (P) and homeothermic (H) vertebrates in ratio, $H{:}P_{\delta} = 2.24$ (Fig. 5d). The species compositions of the other isoform nodes are shown in Fig. 5e. 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"(Phylogenetics)). The H:P ratios of the β (1.1) and the larger γ nodes (1.02, 1.66), but not the smallest γ node (γ_S), were like δ. $H{:}P_{\gamma S} = 3.35$. The α cluster was markedly different from the other isoform clusters. The smaller α node (α_S) though most closely related to the other isoforms, nevertheless diverged significantly

from them ($\overline{\Delta X}_{\alpha S-\beta\gamma\delta} = 0.208 \pm 0.014$ versus $\overline{\Delta X}_{\beta\gamma\delta} = 0.148 \pm 0.009$, $\overline{\Delta X}_{\gamma S-\beta\gamma\delta} = 0.084 \pm 0.062$). It consisted dominantly of poikilotherms ($H{:}P_{\alpha S} = 0.1$). The larger node (α_L) diverged even more ($\Delta X_{\propto L-S} = 0.126$). The α_L node consisted largely of mammalian particularly anthropoid sequences ($H{:}P_{\alpha L} = 5.66$) consistent with the rapid evolution of memory in primates (Fig. 5d)[38].

Discussion

This study has associated two dynamic processes that occur over dramatically different timescales—the macromolecular motions ($< 10^{-4}$ s) of the assembly with the evolution of the CaMKII holoenzyme over billions of years ($> 10^{14}$ s). The emergence of the CaMKII holoenzyme from ancient enzymes and the correlation with behavior involved distinct transitions coupled to fundamental changes in life forms (Fig. 6).

The structural phylogeny shows the CaMKII AD evolved via elongation and enhanced curvature of its β-sheet. The comparison of the *Streptomyces* and human CaMKIIα Vert-Dim dimers (Fig. 4c) supports spectroscopic[10,40] and structural[41] evidence that rings form by serial extension of vertically oriented dimer homologs. The elongated β-sheet topology assembles the hub and propagates conformational fluctuations across Lat-Dim as well as Vert-Dim contacts. The principal motions obtained from the PCA replicated the modes deduced from a survey of 743 PDB structures as fundamental collective motions driven by β-folds that propagate perpendicular to the plane of the fold[42]. What might be the role of these motions?

The KD–KD coupling mediated via the hub documented here is evidence for subunit cooperativity. CaMKII dimer assemblies have a lower Ca^{2+}/CaM half-response dose (EC_{50}) for kinase activation[10], albeit similar Hill coefficients dominated by T286 trans-phosphorylation. The cooperativity can rationalize the different dimer and holoenzyme kinase activities. Holoenzyme structures obtained by cryo-electron tomography[43,44], as opposed to the 3SOA.PDB crystal structure analyzed here, reveal that the inactive holoenzyme predominantly exists in an extended conformation without AD–KD contacts. The compliant, flexible linkers could act as mechanical low-pass filters to attenuate the KD–AD coupling, being wound in response to hub hinge motions. The hub interfacial (Vert-Dim, Lat-Dim) dynamics that orchestrate the overall KD–KD coupling should remain unaltered. The activated holoenzyme is also extended, but the R segments may interact with the hub[20]. The multimeric CaMKII architecture further allows multivalent binding to the actin cytoskeleton[45,46] or multiple partners at the post-synaptic membrane in neurons[47], and flexibility could be important for multivalent binding. Although models have utilized dimer and simpler geometries to reduce computational cost this study underscores the value of accurate models[48] with interfacial energies constrained by structural data for modelling kinase activation.

This study documents that the α isoform has a unique evolutionary trajectory relative to the other isoforms. It emerged last with expression largely in the brain but evolved more rapidly than the rest. The link between CaMKII and sensory behavior required AD fusion with an ancestral KD after ring assemblies formed. Bacterial, archaeal, and eukaryotic serine-threonine kinases have a common ancestor[49,50], while Ca^{2+}/CaM dependent protein kinases are present in nitrogen-fixing bacteria[51] so unicellular organisms could have provided the ancestral KD. Horizontal gene transfer-based shuffling events and the increased biosphere gene pool associated with the emergence of multicellular organisms presumably led to the fusion[52]. An ancestral CaMKII is found in the choanoflagellate *Salpingoeca rosetta* that alternates between unicellular and multicellular lifestyles, has a sensory response (chemotaxis), Homer and other primordial synaptic scaffolding proteins[53]. Alternative splicing, important in midge *Clunio marinus* chronobiology[15] and short and long-term habituation in nematode *C. elegans* mechanosensory neurons[54] preceded CaMKII isoforms, that were most likely generated by gene duplication events[12]. Advanced memory mechanisms from timed sequence representation in rodents (*Rattus norvegicus*) to visual memory and perception in humans (*Homo sapiens*) emerged in mammals[55]. These mechanisms required variable thresholds and expanded range of electrical, hence post-synaptic Ca^{2+}, stimulation frequencies regulated by neuromodulators and associative learning[56,57]. Linker splicing could be too coarse a mechanism to execute such tasks without pathological consequences[8]. Furthermore, α has few spliced variants consistent with its limited tissue distribution[12]. The development of the hypothalamus for temperature homeostasis and the associated amygdaloid complex has been argued to have facilitated the evolution of advanced memory[38]. Strikingly, the rapid evolution of the α isoform followed the poikilotherm to homeotherm jump in contrast to the evolution of the other isoforms. Other physiological processes such as cardiovascular function are also affected in major ways. However δ, the isoform with the greatest similarity to α, is the major isoform expressed in the heart and has a more diverse splice variant and expression profile[12]. Furthermore, actin-binding capability important for heart function is the weakest for, and unlikely to exert selection pressure on, the α isoform[58]. Finally, the evolutionary trace mapped onto the CaMKIIα holoenzyme indicates that the evolution of the α isoform targets the Lat-Dim contact instrumental for conformational spread.

I, therefore, propose that conformational spread in the multimeric ring assembly tunes conformational transitions as in the bacterial flagellar motor[59] to select and optimize the cooperative CaMKIIα kinase response to a broad range of Ca^{2+} pulse frequencies. The realization that a small αβ domain is partitioned into two distinct sectors based on the evolution of its residue contacts is a remarkable result that both explains the mechanics of the KD–KD coupling and rationalizes the maintenance of an energetically metastable, conformationally plastic sector. Evolution of the ancient Vert-Dim contact slowed after ring assemblies appeared allowing this contact to function as a conserved, semi-rigid connector module while the Lat-Dim contact formed a fine-grained dynamic code for KD–KD coupling with coevolved contacts that continue to change (Figs. 3, 4b). The code can be configured in two ways; by variation in the strength of the lateral interfacial contacts[60] and/or by change of subunit stoichiometry[59] mediated, in part, by activation-triggered subunit exchange[20,22]. The hub's lateral sector contacts could be regulated by such exchange consistent with disassembly by R peptides[20]. The conformational fluctuations of the R segment, a major KD network node coupled to hub dynamics, will be modulated upon Ca^{2+}/CaM binding[61], subunit capture[9] and substrate occupancy[62]. The formation of αβ heterooligomers, subunit

Figure 5. CaMKII phylogenetics (from 1D Sequences). (**a**). Metazoan CaMKII-AD evolution. Tree based on 1000 clustered homologs of the *C. elegans* CaMKII (black asterisk). Midge (pink asterisk). (**b**). Isoform evolution. Tree based on 2000 clustered homologs of the rat CaMKII α and β isoforms. Squares denote the most distant, larger CaMKIIα node (L) and the smallest y node with H:P ratios > $\overline{H:P} \pm 2$ s (1.5 ± 1.33), where $\overline{H:P}$ is the mean of the other nodes. Circles mark major clusters (diameter = membership; color = (i) phylum; (ii) isoform), Note difference in scale bar from that in Fig. S2C. (**c**) AD–KD coupling. The AD versus KD similarity plot compared the KD and AD phylogenetic trees color coded by isoform as in panel b (Box) The dual color symbols identify isoform pairs. The correlation coefficient, r = 0.59 with the best-fit (solid line) ± 95% confidence intervals (dashed lines). (**d**). Evolution of memory. Phylogenetic tree based on behavioral assays (from[39]). Arrows show the major bifurcations associated with the emergence of homeotherms (blue) and primates (red). Box: The distribution of the poikilotherms (blue, cyan) and homeotherms (rouge, red, orange, salmon) kingdoms in the sequences of the δ isoform. (**e**). Phylogenetic Species Diversity: Species distributions for major nodes of the α, β and γ isoforms based on kingdom (i) and homeostasis (ii). Source listed in Table S1.

stoichiometry variation, and KD–AD interactions could all enhance the combinatorial increase in the frequency range regulated by the hub.

Methods

Sequence analysis. The Pfam PF08332 MSA (1842 sequences) was input into ConSurf[63] 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[64]). The expanded dataset (16,485 sequences (seq)) was submitted for MSA construction with HHblits ($E < 10^{-6}$, 4 iterations, 75% coverage of the 3SOA AD (135 residues length (len)). Coevolution strength was given by the raw score (S_r) a function of the pseudo-likelihood learning procedure, entropic correction, normalization (seq/len = 16,485/135) and separation between residue positions (> 3). The scaled score, S_s), is the normalized score ($S_r/\overline{S_r}$). The top 30 couplings ($S_s > 1.4 > 0.995$ contact probability, (P_c)), with validated contact, out of a total of 187 ($S_s > 0.5 > 0.265$ P_c) were mapped on the 3D structure of the human holoenzyme. The 186 couplings represented 1% of the possible couplings between residue positions.

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[65]). The thousand closely related homologs for each of the *C. elegans* CaMKII, rat CaMKIIα and CaMKIIβ sequences were identified with UniProt (www.UniProt.org[66]). The sequences were clustered with CD-Hit[67] at the 0.8 cutoff threshold. Hierarchical clustering with the 0.8 cutoff, followed by a 0.6 cutoff was used for the PF08332, *C. elegans* and composite CaMKIIαβ sequence sets. The sequences were assigned to isoform either by direct readout of the UniProt headers or by comparison against the rat isoforms and selection by lowest E-value The MSAs of the cluster representatives were constructed with MUSCLE[68]. Crystal structures were downloaded from Protein Data Bank (www.rcsb.org[69]).

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 ($\overline{\Delta X} = \left(\sum_{i=1}^{n} \Delta X/n\right)$). The similarity between the CaMKII AD (X) and KD (Y) tree topologies was measured as the r score[37].

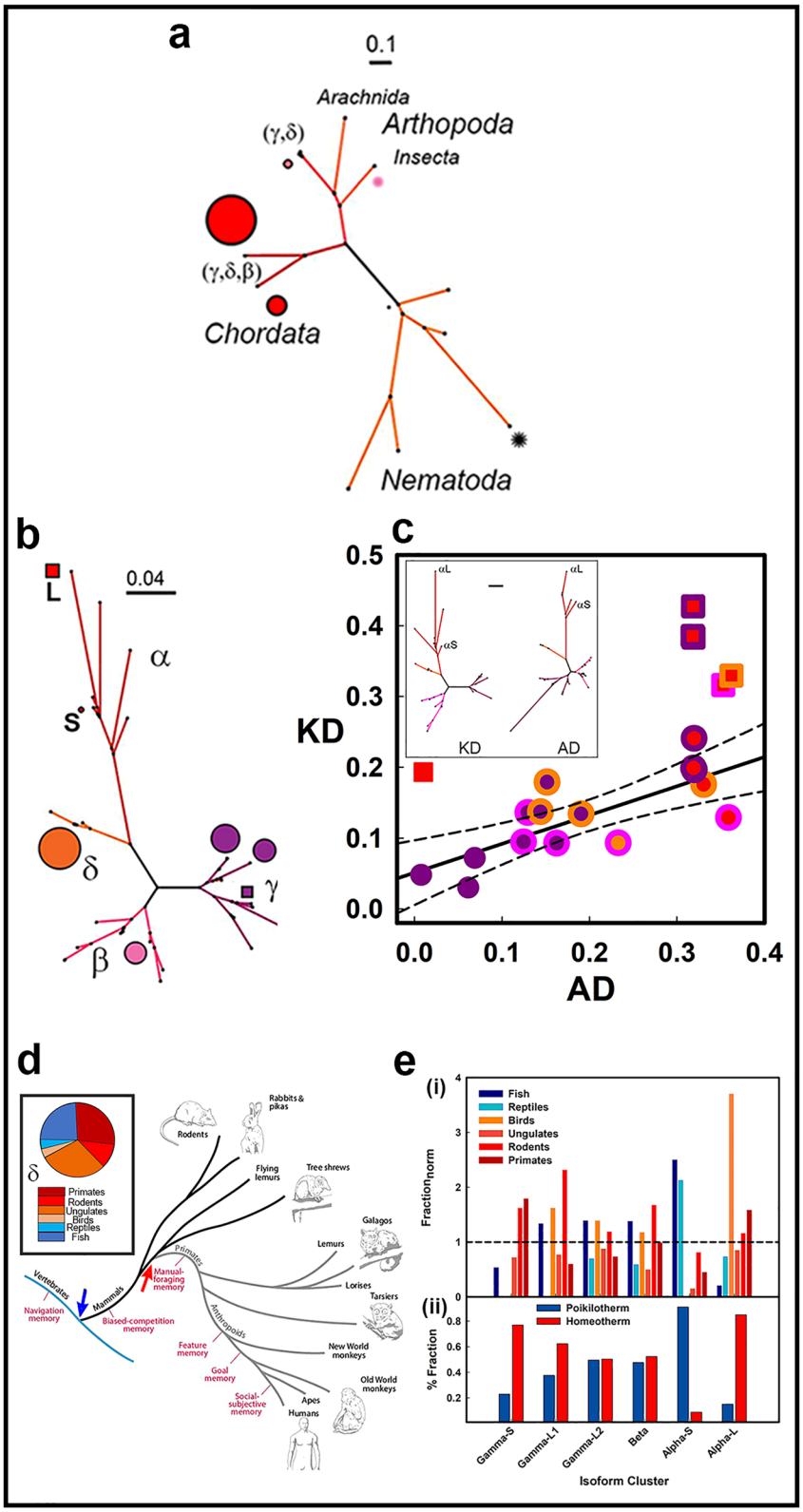
$$r = \sum_{i=1}^{n} (Yi - \overline{Y})(Xi - \overline{X}) / \left\{ \sqrt{\sum_{i=1}^{n} (Yi - \overline{Y})^2} \sqrt{\sum_{i=1}^{n} (Xi - \overline{X})^2} \right\} \quad (1)$$

Structure analysis. The topology of the 3D crystal structures was analyzed with CCP4[70] and DALI[71]. 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[72]. The homologous sequences for the bacterial AD (4OVM.PDB) and human CaMKIIα (3SOA.PDB) 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.

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

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.


a
0.1
Arachnida
Arthopoda
(γ,δ)
Insecta
(γ,δ,β)
Chordata
Nematoda
b
0.04
L
α
S
δ
γ
β
c
KD
AD
αL
αS
d
δ
Primates
Rodents
Ungulates
Birds
Reptiles
Fish
Rodents
Rabbits & pikas
Flying lemurs
Tree shrews
Galagos
Lemurs
Lorises
Tarsiers
New World monkeys
Old World monkeys
Apes
Humans
Primates
Anthropoids
Mammals
Vertebrates
Manual-foraging memory
Navigation memory
Biased-competition memory
Feature memory
Goal memory
Social-subjective memory
e
(i)
Fraction norm
Fish
Reptiles
Birds
Ungulates
Rodents
Primates
(ii)
% Fraction
Poikilotherm
Homeotherm
Gamma-S
Gamma-L1
Gamma-L2
Beta
Alpha-S
Alpha-L
Isoform Cluster

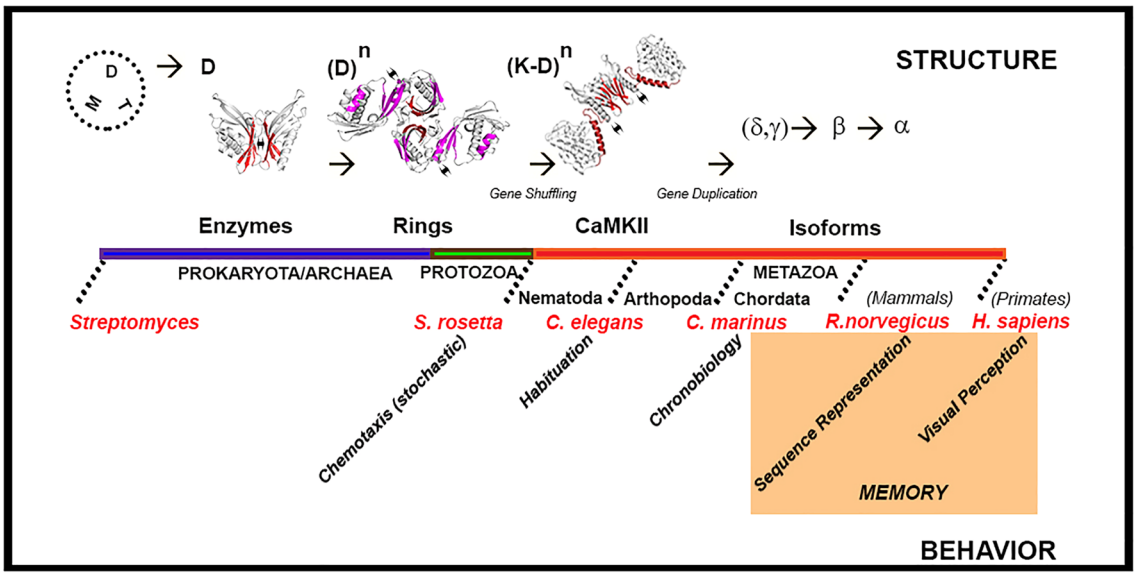


Figure 6. The correlation between the structural evolution of CaMKII and sensory behavior. Structure. The selection of the dimer (D) from other structures (M = monomer, T = tetramer) in bacteria seeded the emergence of ringed hub assemblies. The fusion with the kinase (K) domains coincided with the emergence of multicellularity (secondary structures color coded as in Fig. 1). Diversity was created by linker alternative splicing and enhanced by the generation of isoforms. Behavior. Work on model organisms suggests CaMKII evolution peaks with the development of cognitive memory. Advanced memory mechanisms (orange block). Source listed in Table S1.

$$s_i = -\sum_{a=1}^{20} f_{ia} ln(f_{ia}) \tag{3}$$

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)[73].

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

The native contact is "minimally frustrated" if its energy ΔE_{ij}^{N} is at the lower end of the $\Delta E_{i'j'}^{D}$ 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 were taken as minimally frustrated ("relaxed") and highly frustrated ("stressed"), respectively. A case study of the integration of *rvET* and ΔEfr to understand protein design is available for calmodulin[74].

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 (3SOA.PDB). The tetramer contained all lateral and vertical dimer contacts represented in the intact holoenzyme. Conformational ensembles of these structures were generated in Gromacs 4.5.7 (www.gromacs.org[75]) with tCONCOORD[27] as described previously[76]. Full atom detail is preserved as a new structure is generated by random displacement of the constituent atoms within limits (2 nm^3) followed by correction, up to 500 iterations, to eliminate bond violations. The solvent is implicit—solvent atoms are not simulated for increased computational speed, but a solvation parameter estimates the distance-dependent probability of a water molecule next to a particular atom for the prediction of unstable hydrogen bonds. This parameter was set to 2.2.

The essential collective motions were obtained by PCA[77]. The PCs were generated by diagonalization of the covariance matrix of C^{α} positions derived from the tCONCOORD ensembles. The variance of the PCs as given by the eigenvalues 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 timescale.

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[78]. The resulting array of 1D strings was used to derive a network of dynamic couplings based on normalized mutual information (*nMI*) with GSATools[79]). 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.

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

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 66 top couplings ($nMI > 0.15$, fragment separation > 4 residues) were mapped on the 3D structures with a Pymol plugin[80]. Of these, 7 spanned the Vert-Dim interface, while 21 spanned the Lat-Dim interface.

The contribution of a node to the network scaled with its connectivity, estimated by the eigenvector centrality, E, calculated directly from the correlation matrix:

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

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

Table S1 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[81]. 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))[64], of which the top 30 (> 0.995 significance) were mapped onto the crystal structure. Pearson's coefficient was used to assess the similarity between KD and AD tree topologies.

Dynamics. 66,536 (16^4) equilibrium conformations were generated for the monomer and tetramer structures extracted from 3SOA.PDB. The overlap between ensemble subsets was > 99% when subset size was < ¼ of the total ensemble, as reported previously for CaMKII KD structures[80]. The top network couplings mapped onto the AD crystal structure represented pairs above the 2s significance threshold in the distribution obtained after correction for the finite size error.

Data 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.

Received: 26 December 2021; Accepted: 3 May 2022
Published online: 19 May 2022

References

1. Lisman, J. The CaM kinase II hypothesis for the storage of synaptic memory. *Trends Neurosci.* **17**, 406–412. https://doi.org/10.1016/0166-2236(94)90014-0 (1994).
2. De Koninck, P. & Schulman, H. Sensitivity of CaM kinase II to the frequency of Ca^{2+} oscillations. *Science* **279**, 227–230. https://doi.org/10.1126/science.279.5348.227 (1998).
3. Morris, E. P. & Torok, K. Oligomeric structure of alpha-calmodulin-dependent protein kinase II. *J. Mol. Biol.* **308**, 1–8. https://doi.org/10.1006/jmbi.2001.4584 (2001).
4. Bayer, K. U. & Schulman, H. CaM Kinase: Still inspiring at 40. *Neuron* **103**, 380–394. https://doi.org/10.1016/j.neuron.2019.05.033 (2019).
5. Hell, J. W. CaMKII: Claiming center stage in postsynaptic function and organization. *Neuron* **81**, 249–265. https://doi.org/10.1016/j.neuron.2013.12.024 (2014).
6. Kury, S. *et al.* De novo mutations in protein kinase genes CAMK2A and CAMK2B cause intellectual disability. *Am. J. Hum. Genet.* **101**, 768–788. https://doi.org/10.1016/j.ajhg.2017.10.003 (2017).
7. Akita, T. *et al.* De novo variants in CAMK2A and CAMK2B cause neurodevelopmental disorders. *Ann. Clin. Transl. Neurol.* **5**, 280–296. https://doi.org/10.1002/acn3.528 (2018).
8. Brown, C. N. *et al.* Characterization of six CaMKIIalpha variants found in patients with schizophrenia. *iScience* **24**, 103184. https://doi.org/10.1016/j.isci.2021.103184 (2021).
9. Chao, L. H. *et al.* Intersubunit capture of regulatory segments is a component of cooperative CaMKII activation. *Nat. Struct. Mol. Biol.* **17**, 264–272. https://doi.org/10.1038/nsmb.1751 (2010).
10. Sarkar, P. *et al.* Deciphering CaMKII multimerization using fluorescence correlation spectroscopy and homo-FRET analysis. *Biophys. J.* **112**, 1270–1281. https://doi.org/10.1016/j.bpj.2017.02.005 (2017).
11. Brocke, L., Srinivasan, M. & Schulman, H. Developmental and regional expression of multifunctional Ca^{2+}/calmodulin-dependent protein kinase isoforms in rat brain. *J. Neurosci.* **15**, 6797–6808 (1995).
12. 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. https://doi.org/10.1016/j.gene.2003.08.023 (2003).
13. 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).
14. Bhattacharyya, M. *et al.* Flexible linkers in CaMKII control the balance between activating and inhibitory autophosphorylation. *Elife* https://doi.org/10.7554/eLife.53670 (2020).
15. Kaiser, T. S. *et al.* The genomic basis of circadian and circalunar timing adaptations in a midge. *Nature* **540**, 69–73. https://doi.org/10.1038/nature20151 (2016).
16. 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. https://doi.org/10.1016/s0169-328x(00)00115-7 (2000).
17. Leurs, U. *et al.* GHB analogs confer neuroprotection through specific interaction with the CaMKIIalpha hub domain. *Proc. Natl. Acad. Sci. U. S. A.* https://doi.org/10.1073/pnas.2108079118 (2021).
18. Sloutsky, R. *et al.* Heterogeneity in human hippocampal CaMKII transcripts reveals allosteric hub-dependent regulation. *Sci. Signal* https://doi.org/10.1126/scisignal.aaz0240 (2020).

19. Cook, S. G. *et al.* Analysis of the CaMKIIalpha and beta splice-variant distribution among brain regions reveals isoform-specific differences in holoenzyme formation. *Sci. Rep.* **8**, 5448. https://doi.org/10.1038/s41598-018-23779-4 (2018).
20. Karandur, D. *et al.* Breakage of the oligomeric CaMKII hub by the regulatory segment of the kinase. *Elife* https://doi.org/10.7554/eLife.57784 (2020).
21. Bhattacharyya, M. *et al.* Molecular mechanism of activation-triggered subunit exchange in Ca(2+)/calmodulin-dependent protein kinase II. *Elife* https://doi.org/10.7554/eLife.13405 (2016).
22. Stratton, M. *et al.* Activation-triggered subunit exchange between CaMKII holoenzymes facilitates the spread of kinase activity. *Elife* **3**, e01610. https://doi.org/10.7554/eLife.01610 (2014).
23. Ferreiro, D. U., Komives, E. A. & Wolynes, P. G. Frustration, function and folding. *Curr. Opin. Struct. Biol.* **48**, 68–73. https://doi.org/10.1016/j.sbi.2017.09.006 (2018).
24. Ferreiro, D. U., Hegler, J. A., Komives, E. A. & Wolynes, P. G. Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 19819–19824 (2007).
25. Morcos, F. *et al.* Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E1293-1301. https://doi.org/10.1073/pnas.1111471108 (2011).
26. Chao, L. H. *et al.* A mechanism for tunable autoinhibition in the structure of a human Ca2+/calmodulin- dependent kinase II holoenzyme. *Cell* **146**, 732–745. https://doi.org/10.1016/j.cell.2011.07.038 (2011).
27. Seeliger, D., Haas, J. & de Groot, B. L. Geometry-based sampling of conformational transitions in proteins. *Structure* **15**, 1482–1492. https://doi.org/10.1016/j.str.2007.09.017 (2007).
28. Torres-Ocampo, A. P. *et al.* Characterization of CaMKIIalpha holoenzyme stability. *Protein Sci.* **29**, 1524–1534. https://doi.org/10.1002/pro.3869 (2020).
29. 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. https://doi.org/10.1038/ja.2016.88 (2016).
30. Eberhardt, R. Y. *et al.* Filling out the structural map of the NTF2-like superfamily. *BMC Bioinform.* **14**, 327. https://doi.org/10.1186/1471-2105-14-327 (2013).
31. Hoelz, A., Nairn, A. C. & Kuriyan, J. Crystal structure of a tetradecameric assembly of the association domain of Ca^{2+}/calmodulin-dependent kinase II. *Mol. Cell* **11**, 1241–1251. https://doi.org/10.1016/s1097-2765(03)00171-0 (2003).
32. Huse, M. & Kuriyan, J. The conformational plasticity of protein kinases. *Cell* **109**, 275–282 (2002).
33. 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. https://doi.org/10.1002/pro.3614 (2019).
34. Ciccarelli, F. D. *et al.* Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283–1287. https://doi.org/10.1126/science.1123061 (2006).
35. Chevrette, M. G. *et al.* Taxonomic and metabolic incongruence in the ancient genus streptomyces. *Front Microbiol.* **10**, 2170. https://doi.org/10.3389/fmicb.2019.02170 (2019).
36. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022–3027. https://doi.org/10.1093/molbev/msab120 (2021).
37. Pazos, F. & Valencia, A. Similarity of phylogenetic trees as indicator of protein-protein interaction. *Protein Eng.* **14**, 609–614. https://doi.org/10.1093/protein/14.9.609 (2001).
38. Loonen, A. J. M. & Ivanova, S. A. Circuits regulating pleasure and happiness: Evolution and role in mental disorders. *Acta Neuropsychiatr.* **30**, 29–42. https://doi.org/10.1017/neu.2017.8 (2018).
39. 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. https://doi.org/10.1093/cercor/bhv329 (2017).
40. 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. https://doi.org/10.1073/pnas.0901913106 (2009).
41. Rellos, P. *et al.* Structure of the CaMKIIdelta/calmodulin complex reveals the molecular mechanism of CaMKII kinase activation. *PLoS Biol.* **8**, e1000426 (2010).
42. 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. https://doi.org/10.1038/ncomms5070 (2014).
43. Myers, J. B. *et al.* The CaMKII holoenzyme structure in activation-competent conformations. *Nat. Commun.* **8**, 15742 (2017).
44. Buonarati, O. R. *et al.* CaMKII versus DAPK1 Binding to GluN2B in ischemic neuronal cell death after resuscitation from cardiac arrest. *Cell Rep.* **30**, 1-8 e4. https://doi.org/10.1016/j.celrep.2019.11.076 (2020).
45. 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. https://doi.org/10.1073/pnas.1911452116 (2019).
46. Khan, S., Downing, K. H. & Molloy, J. E. Architectural dynamics of CaMKII-actin networks. *Biophys. J.* **116**, 104–119. https://doi.org/10.1016/j.bpj.2018.11.006 (2019).
47. Khan, S. *et al.* Sequestration of CaMKII in dendritic spines in silico. *J. Comput. Neurosci.* **31**, 581–594 (2011).
48. Pharris, M. C. *et al.* A multi-state model of the CaMKII dodecamer suggests a role for calmodulin in maintenance of autophosphorylation. *PLoS Comput. Biol.* **15**, e1006941. https://doi.org/10.1371/journal.pcbi.1006941 (2019).
49. 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. https://doi.org/10.1016/j.jmb.2017.11.004 (2018).
50. Pereira, S. F., Goss, L. & Dworkin, J. Eukaryote-like serine/threonine kinases and phosphatases in bacteria. *Microbiol. Mol. Biol. Rev.* **75**, 192–212. https://doi.org/10.1128/MMBR.00042-10 (2011).
51. Levy, J. *et al.* A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* **303**, 1361–1364. https://doi.org/10.1126/science.1093038 (2004).
52. Mikhailovsky, G. & Gordon, R. Shuffling type of biological evolution based on horizontal gene transfer and the biosphere gene pool hypothesis. *Biosystems* **193–194**, 104131. https://doi.org/10.1016/j.biosystems.2020.104131 (2020).
53. Burkhardt, P. *et al.* Evolutionary insights into premetazoan functions of the neuronal protein homer. *Mol. Biol. Evol.* **31**, 2342–2355. https://doi.org/10.1093/molbev/msu178 (2014).
54. Bozorgmehr, T., Ardiel, E. L., McEwan, A. H. & Rankin, C. H. Mechanisms of plasticity in a *Caenorhabditis elegans* mechanosensory circuit. *Front. Physiol.* **4**, 88. https://doi.org/10.3389/fphys.2013.00088 (2013).
55. Murray, E. A., Wise, S. P. & Graham, K. S. Representational specializations of the hippocampus in phylogenetic perspective. *Neurosci. Lett.* **680**, 4–12. https://doi.org/10.1016/j.neulet.2017.04.065 (2018).
56. 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. https://doi.org/10.1371/journal.pcbi.1005820 (2017).
57. 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. https://doi.org/10.1073/pnas.91.22.10290 (1994).
58. 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. https://doi.org/10.1016/j.bpj.2016.06.007 (2016).
59. 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. https://doi.org/10.1371/journal.pcbi.1002523 (2012).

60. Duke, T. A., Le Novere, N. & Bray, D. Conformational spread in a ring of proteins: A stochastic approach to allostery. *J. Mol. Biol.* **308**, 541–553. https://doi.org/10.1006/jmbi.2001.4610 (2001).
61. 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. https://doi.org/10.1016/j.cell.2005.10.029 (2005).
62. Nguyen, T. A. *et al.* Covert changes in CaMKII holoenzyme structure identified for activation and subsequent interactions. *Biophys. J.* **108**, 2158–2170. https://doi.org/10.1016/j.bpj.2015.03.028 (2015).
63. Ashkenazy, H. *et al.* ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344-350. https://doi.org/10.1093/nar/gkw408 (2016).
64. Ovchinnikov, S., Kamisetty, H. & Baker, D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *Elife* **3**, e02030. https://doi.org/10.7554/eLife.02030 (2014).
65. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432. https://doi.org/10.1093/nar/gky995 (2019).
66. UniProt, C. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515. https://doi.org/10.1093/nar/gky1049 (2019).
67. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150–3152. https://doi.org/10.1093/bioinformatics/bts565 (2012).
68. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797. https://doi.org/10.1093/nar/gkh340 (2004).
69. 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. https://doi.org/10.1002/pro.3331 (2018).
70. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242. https://doi.org/10.1107/S0907444910045749 (2011).
71. Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140. https://doi.org/10.1002/pro.3749 (2020).
72. 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. https://doi.org/10.1016/j.jmb.2003.12.078 (2004).
73. 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. https://doi.org/10.1093/nar/gkw304 (2016).
74. Tripathi, S., Waxham, M. N., Cheung, M. S. & Liu, Y. Lessons in protein design from combined evolution and conformational dynamics. *Sci. Rep.* **5**, 14259. https://doi.org/10.1038/srep14259 (2015).
75. Pronk, S. *et al.* GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854. https://doi.org/10.1093/bioinformatics/btt055 (2013).
76. Pandini, A., Morcos, F. & Khan, S. The gearbox of the bacterial flagellar motor switch. *Structure* **24**, 1209–1220. https://doi.org/10.1016/j.str.2016.05.012 (2016).
77. 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. https://doi.org/10.1080/07391102.1996.10508874 (1996).
78. 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. https://doi.org/10.1096/fj.11-190868 (2012).
79. 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. https://doi.org/10.1093/bioinformatics/btt326 (2013).
80. Pandini, A., Schulman, H. & Khan, S. Conformational coupling by trans-phosphorylation in calcium calmodulin dependent kinase II. *PLoS Comput. Biol.* **15**, e1006796. https://doi.org/10.1371/journal.pcbi.1006796 (2019).
81. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490. https://doi.org/10.1371/journal.pone.0009490 (2010).

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 contributions

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

Competing interests

The author declares no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-12090-y.

Correspondence and requests for materials should be addressed to S.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.