

Novel transglutaminase-like peptidase and C2 domains elucidate the structure, biogenesis and evolution of the ciliary compartment

Dapeng Zhang and L. Aravind*

National Center for Biotechnology Information; National Library of Medicine; National Institutes of Health; Bethesda, MD USA

Keywords: ciliogenesis, transglutaminase-like, membrane, tubulin-tyrosine ligase, C2, transition zone, Y-shaped linkers, evolution, origin of eukaryotes, ciliopathy

In addition to their role in motility, eukaryotic cilia serve as a distinct compartment for signal transduction and regulatory sequestration of biomolecules. Recent genetic and biochemical studies have revealed an extraordinary diversity of protein complexes involved in the biogenesis of cilia during each cell cycle. Mutations in components of these complexes are at the heart of human ciliopathies such as Nephronophthisis (NPHP), Meckel-Gruber syndrome (MKS), Bardet-Biedl syndrome (BBS) and Joubert syndrome (JBTS). Despite intense studies, proteins in some of these complexes, such as the NPHP1-4-8 and the MKS, remain poorly understood. Using a combination of computational analyses we studied these complexes to identify novel domains in them which might throw new light on their functions and evolutionary origins. First, we identified both catalytically active and inactive versions of transglutaminase-like (TGL) peptidase domains in key ciliary/centrosomal proteins CC2D2A/MKS6, CC2D2B, CEP76 and CCDC135. These ciliary TGL domains appear to have originated from prokaryotic TGL domains that act as peptidases, either in a prokaryotic protein degradation system with the MoxR AAA+ ATPase, the precursor of eukaryotic dyneins and midasins, or in a peptide-ligase system with an ATP-grasp enzyme comparable to tubulin-modifying TTL proteins. We suggest that active ciliary TGL proteins are part of a cilia-specific peptidase system that might remove tubulin modifications or cleave cilia-localized proteins, while the inactive versions are likely to bind peptides and mediate key interactions during ciliogenesis. Second, we observe a vast radiation of C2 domains, which are key membrane-localization modules, in multiple ciliary proteins, including those from the NPHP1-4-8 and the MKS complexes, such as CC2D2A/MKS6, RPGRIP1, RPGRIP1L, NPHP1, NPHP4, C2CD3, AHI1/Joubertin and CEP76, most of which can be traced back to the last eukaryotic ancestor. Identification of these TGL and C2 domains aid in the proper reconstruction of the Y-shaped linkers, which are key structures in the transitional zone of cilia, by allowing precise prediction of the multiple membrane-contacting and protein-protein interaction sites in these structures. These findings help decipher key events in the evolutionary separation of the ciliary and nuclear compartments in course of the emergence of the eukaryotic cell.

Introduction

Eukaryotic cilia (or flagella) are fundamentally different in structural and mechanistic terms from the superficially similar motility organelles of the prokaryotic superkingdoms. Unlike prokaryotic flagella, which are extracellular organelles, these organelles are contiguous with the cytoplasm cell body and are supported by a distinct microtubular skeleton, the axoneme.¹ In keeping with this, studies over the past 25 years have revealed that cilia are not just organelles of motility, but a distinct subcellular compartment, which has important additional roles as a locus for sensory signal transduction and for regulatory sequestering of proteins away from the cytoplasm of the cell body.^{2,3} Recent studies have indicated that several hundreds of proteins reside transiently or permanently in cilia.^{4,5} Beyond the microtubular axonemal core (typically adopting the “9 + 2” microtubule configuration) and

motor proteins, the resident proteins form several distinct complexes, which are central to ciliary function and assembly.⁶ These include (1) the dynein-regulatory (nexin) complex, which connects the microtubule doublets of the cilium, and also links them to different dyneins;⁷ (2) the septin 2/7 complex, which forms a transport barrier at the base of the cilium;^{8,9} (3) the tubulin polyglutamylase complex,¹⁰ which covalently modifies the microtubules and regulates their interactions with the dyneins; (4) the Bardet-Biedl protein complex or BBsome;¹¹ (5) the intraflagellar-transport complexes or IFT complexes;^{12,13} (6) the MKS complex⁶ and (7) the NPHP 1-4-8 complex.¹⁴ The last four complexes play distinct roles in the assembly and membrane association of the ciliary cytoskeleton and trafficking of membrane proteins into the ciliary compartment.¹⁵⁻¹⁷

While the structure and biogenesis of cilia show some tissue-specific and phyletic differences, certain common features can

*Correspondence to: L. Aravind; Email: aravind@ncbi.nlm.nih.gov

Submitted: 08/21/12; Accepted: 09/03/12

<http://dx.doi.org/10.4161/cc.22068>

be discerned across eukaryotes. In general, ciliogenesis follows a unique series of steps, which distinguish it from the dynamics of other subcellular compartments. In each cell cycle, the first step in ciliogenesis is the maturation of the basal body from the centriole or the cognate microtubule-organizing center.¹⁸ In the simplest cases, the basal body directly migrates to the proximity of the cell membrane to initiate ciliogenesis.¹⁹ In other cases, the basal body is first capped by a double-membrane sheath, the ciliary vesicle, which brings the basal body close to the cell membrane by fusing with it.¹⁸ The fusion might result in a local invagination, the ciliary pocket, with which the incipient cilium (equivalent to the transition zone of the mature cilium) associates and serves as the center for further trafficking of proteins in and out of the ciliary compartment. The specific targeting of axonemal and ciliary membrane proteins to this region then allows further growth of the cilium to a predetermined length. Genetic and cytological studies suggest that the early steps of ciliogenesis, including the association of the ciliary body with the vesicle, or its docking close to the cell membrane, are dependent on components of the MKS and NPHP 1-4-8 complexes.^{15,16,20-22} The subsequent steps involving preliminary growth and extension of the cilium beyond the transition zone are dependent on the BBSome and IFT complexes.^{23,24}

The presence of cilia (flagella) at the base of all major eukaryotic lineages, and also their presence in most “excavate” lineages,²⁵ which are likely to be the earliest-branching eukaryotic clades, indicates that they are a shared derived character (synapomorphy) of the entire eukaryotic clade.^{26,27} Thus, the early evolution of cilia as a novel multifunctional organelle distinct from the motility systems observed in the prokaryotic superkingdoms is central to the question of eukaryotic origins. Answering this question is complicated by the emergence of a practically “fully formed” cilium in the last eukaryotic common ancestor (LECA), with apparently no intermediates or precursors. Some aspects of the early evolution of cilia are also related to the more general question of the emergence of subcellular compartments during the origin of eukaryotes. Maturation of the ciliary compartment and transport of proteins into it are dependent on a number of small GTPases of the extended Ras-like clade, such as Arl6, which functions with the BBSome, Arl13B, Arf4, Ran and Rab8.²⁸ As suggested by previous studies, the explosive radiation of these small GTPases in eukaryotes from precursors acquired from prokaryotes appears to have been part of not just the emergence of the ciliary compartment, but also other subcellular membrane-bound structures, such as the nucleus and the Golgi-vesicular complex.^{29,30} Indeed, origin of the nuclear and the ciliary compartments might be closely linked,³¹ as suggested by their shared trafficking of GTPase, Ran and the nucleoporins, which also restrict protein transport into the cilia.^{32,33} Computational studies on the sequence relationships of component proteins have provided key leads regarding the origin of the microtubular cytoskeleton and the dynein and kinesin motors.³⁴⁻³⁶ However, it should be kept in mind that the question of the origin of cilia is related to, but not the same as, explaining the provenance of tubulin or the motor proteins. Central to explaining the origin of cilia are scenarios that can account for the microtubular skeleton

and the motor proteins coming together as an assemblage with the membrane-associated protein complexes. In particular, elucidating the origin of the key players in early ciliogenesis (e.g., MKS and NPHP 1-4-8 complexes) and their links to the core ciliary components are likely to be of considerable value in explaining the early evolution of cilia. Our analysis of the B9 proteins, key components of the MKS complex, with identification of a distinct version of the C2 domain, have clarified certain aspects of the early evolution of cilia.³⁷ This and other protein sequence and structure analysis studies have clarified the evolutionary origins of certain components of the cilia.^{28,38} They have also provided new leads to better understand both the cell biology of ciliogenesis and pathologies that are central to a broad class of human diseases known as ciliopathies. Given the efficacy of these computational methods in dissecting the structures and functions of these proteins, we resorted to an in depth sequence analysis of key players in ciliogenesis to detect novel domains and predicted functional linkages to other ciliary components. As a consequence, we were able to reconstruct certain key events in the evolution of these complexes, predict new functions and components and provide a possible explanation for how the cytoskeleton-membrane interactions arose during the origin of eukaryotic cilia.

Results and Discussion

Sequence-structure analysis of components of the MKS and NPHP complexes. Given the central role of the MKS and NPHP 1-4-8 complexes in early ciliogenesis, we attempted to establish the affinities and provenance of key components of these complexes. Interestingly, analysis of sequences of proteins belonging to these complexes using the SEG program, with parameters adjusted to detect globular domains,³⁹ revealed that several of them contained globular regions that did not map to previously characterized domains. Hence, the relationships and structures of these distinct globular regions are vital to develop a better understanding of both the evolution and functions of these ciliary proteins. Table 1 displays a summary of primary components implicated in ciliogenesis that we analyzed in this study along with the known globular domains and those newly detected by us.

Detection of a conserved transglutaminase-like domain in the ciliary/centriolar proteins CC2D2A/MKS6, CC2D2B, CEP76 and CCDC135. The CC2D2A/MKS6⁴⁰ is a large protein in the MKS complex in which, previously, only a C2 domain (gi: 197209974, residues 1,040–1,200 in human CC2D2A; see below for details) could be detected, along with a N-terminal coiled-coil region (Fig. 1A). Further, we noticed that its C-terminal region (residues 1,320–1,430) is found in a paralogous protein CC2D2B (gi: 229577352, residues 100–220; human CC2D2B), which, however, lacks the C2 domain. A search of the non-redundant database with the PSI-BLAST program using this region as seed recovered homologous regions in CEP76,⁴¹ which is a centriole-associated protein, and CCDC135/FAP50/Lost Boys,⁴² which is a conserved ciliary protein tightly associated with the outer microtubule doublets of the

Table 1. Summary of the primary components of ciliary complexes with the domain architectures and function predictions

Protein complex	Gene name	Alternative name by syndrome				GI ID (human)	Domain architecture and mutation/deletion mapping	Functions
		NPHP	MKS	BBS	JBTS			
NPHP 1-4-8	NPHP1	NPHP1			JBTS4	189491774	<u>SH3+NPHP1-C2+α-helical domain</u>	Membrane localization
	NPHP4	NPHP4				23510323	NPHP4N-C2 +NPHP4C-C2+IG*3	Membrane localization
	RPGRIP1L	NPHP8	MKS5		JBTS7	118442834	<u>CC+RPGRIP1N-C2+PKC-C2 +RPGRIP1C-C2</u>	Membrane localization and protein interaction
	RPGRIP1					112734867	CC+ RPGRIP1N-C2+PKC-C2 +RPGRIP1C-C2	Membrane localization+protein interaction
MKS	MKS1		MKS1	BBS13		89242137	B9-C2	Membrane localization
	B9D1		MKS9			7661536	B9-C2	Membrane localization
	B9D2		MKS10			226371646	B9-C2	Membrane localization
	TCTN1				JBTS13	91208022	DUF1619	
	TCTN2		MKS8			74731861	DUF1619	
	TCTN3					91208025	DUF1619	
	CC2D2A		MKS6		JBTS9	197209974	<u>CC+CC2D2AN-C2 + CC2D2AC-C2 + TGL</u>	Membrane localization and protein interaction
	Meckelin/TMEM67	NPHP11	MKS3		JBTS6	317373389	TM	
	TMEM216		MKS2		JBTS2	115387120	TM	
TMEM237				JBTS14	113205077	TM		
Novel members of MKS and NPHP complexes	AHI1/JOUBERIN				JBTS3	31542701	AHI1-C2+WD40+SH3	Membrane localization
	C2CD3					148596944	C2CD3N-C2+PKC-C2*5	Membrane localization
	CEP76					21314728	CEP76-C2+TGL	Membrane localization+peptidase activity
	CCDC135/FAP50					223941912	TGL+CC	Peptidase activity
NPHP 5-6	CEP290	NPHP6	MKS4	BBS14	JBTS5	116241294	CC (Coiled coil)	
	IQCB1	NPHP5				3123054	TPRs+IQ_repeat	Protein interaction

The novel protein domains deleted or disrupted by ciliopathy mutations in humans are underlined.

axoneme ($e < 10^{-5}$ in iterations 1–2). Further iterations of this search recovered a region in several prokaryotic proteins (e.g., *Bacillus subtilis* protein YebA and the *Mycobacterium smegmatis* protein, gi: 118473669; 10^{-7} – 10^{-15} in iterations 3–4). Similar results were obtained in searches with the JACKHMMER program. The region of significant similarity shared by these prokaryotic proteins and the eukaryotic ciliary/centriolar proteins corresponded to the Transglutaminase-like (TGL) domain.

This suggested that the common globular domain shared by CC2D2A, CC2D2B, CEP76 and CCDC135 is a version of the TGL domain (Fig. 1A). The TGL domain adopts the papain-like peptidase fold, whose active site is typically comprised of a catalytic triad formed by a cysteine from an N-terminal helix, and a histidine and an acidic residue from successive strands of the core β -barrel.⁴³ Catalytically active versions are peptidases, peptide-N-glycanases, transglutaminases or deamidases, while

Table 1. Summary of the primary components of ciliary complexes with the domain architectures and function predictions (continued)

Protein complex	Gene name	Alternative name by syndrome				GI ID (human)	Domain architecture and mutation/deletion mapping	Functions
		NPHP	MKS	BBS	JBTS			
BBS	BBS1			BBS1		38257662	WD40+IG	Protein interaction
	BBS2			BBS2		20454827	WD40+IG	Protein interaction
	BBS7			BBS7		29029557	WD40+IG	Protein interaction
	BBS9			BBS9		97180305	WD40+CC+IG	Protein interaction
	BBS4			BBS4		160359000	TPR	Protein interaction
	BBS5			BBS5		74750959	PH	Membrane/lipid interaction
	BBS8			BBS8		308153511	TPR	Protein interaction
	BBS6			BBS6		25914754	Chaperonin	Folding/assembly
	BBS10			BBS10		100816407	Chaperonin	Folding/assembly
	BBS12			BBS12		40217788	Chaperonin	Folding/assembly
	ARL6			BBS3		14149815	Ras_like_GTPase	Localization
	TRIM32			BBS11		153792582	E3 ubiquitin-protein ligase	Ubiquitin modification
Inversin compartment	Inversin	NPHP2				68565551	ANKs	Protein interaction
	NPHP3	NPHP3	MKS7			68565783	CC+ NTPase_STAND+TPRs	ATP-dependent complex assembly
	NEK8	NPHP9				34098463	Protein kinase domain + RCC1s	Protein modification

The novel protein domains deleted or disrupted by ciliopathy mutations in humans are underlined.

catalytic inactive versions (e.g., Rad4 and Cyk3) function as peptide-binding domains.⁴⁴

An examination of the multiple sequence alignment of the core of the TGL domains from prokaryotic TGL proteins and the ciliary/centriolar proteins indicated that the former strongly conserve the catalytic triad, implying that there are active enzymes (Fig. 1A). However, the eukaryotic proteins present a more complex picture due to their drastic divergence relative to the TGL domain of their prokaryotic homologs. In order to understand their evolutionary history, we used the prokaryotic homologs as a root in a phylogenetic analysis of the TGL domains of the eukaryotic ciliary/centriolar proteins. This revealed three distinct clades of the TGL domains in eukaryotes, respectively prototyped by CCDC135, CEP76 and CC2D2A/MKS6 (Fig. 1B). The monophyly of these three distinct eukaryotic branches is supported by several sequence features that they share to the exclusion of all other TGL domains (Fig. 1A). They also share a unique C-terminal $\alpha+\beta$ extension to the core TGL domain that is likely to stack with the former to comprise the complete folded unit. Indeed, such extensions have been previously observed in other subfamilies of TGLs, such as the Rad4-PNGase subgroup.^{43,44} Of the three eukaryotic clades, CEP76 and CC2D2A/MKS6 form a strongly supported higher-order clade (Fig. 1B), which is characterized by a compact TGL domain. In contrast, CCDC135 has a large insert after the first strand of the core TGL domain (Fig. 1A), which is a common location for large inserts in several members of the TGL superfamily.⁴⁴ Examination of the phylogenetic patterns indicates that each of the three eukaryotic clades has representatives from all major ciliated eukaryotic lineages,

including the “excavate” taxa. Among the excavate taxa, *Giardia* and *Trichomonas* tend to form the basal-most branches, followed by kinetoplastids and heteroloboseans. This pattern is consistent with the previously inferred phylogeny of eukaryotic taxa^{25,45} and suggests that the three distinct clades of TGL domains had radiated prior to the LECA. Examination of the active site residues (Fig. 1A) reveals that all representatives of the CC2D2A/MKS6 clade are inactive. In the CEP76 clade, most representatives are inactive, though many of the animal sequences appear to possess intact active site residues. In the CCDC135 clade, one subset contains the usual active site cysteine, suggesting that they function as thiol peptidases. The other subset (e.g., in *Drosophila*; Fig. 1A) displays a serine in place of the catalytic cysteine, similar to what is observed in the TGL domain of the eukaryotic Menin proteins.^{46,47} Hence, these serine-containing TGL domains could be potentially active as serine peptidases, comparable to the previously noted case of the plasmodial SERA peptidase,⁴⁸ which displays a cathepsin-like version of the papain-like fold with an active site serine. Thus, the eukaryotic ciliary/centriolar TGLs are predicted to include both catalytically active (mainly from the CCDC135 clade) and inactive versions (mostly from the CEP76 and CC2D2A/MKS6 clades).

Functional and evolutionary implications of ciliary TGL domains: Inferences from operonic associations of the bacterial homologs. The prokaryotic homologs recovered in searches with the eukaryotic ciliary/centriolar TGL domains provide functional clues regarding their eukaryotic counterparts. We used contextual information gleaned from domain architectures and conserved gene neighborhoods (predicted operons) to

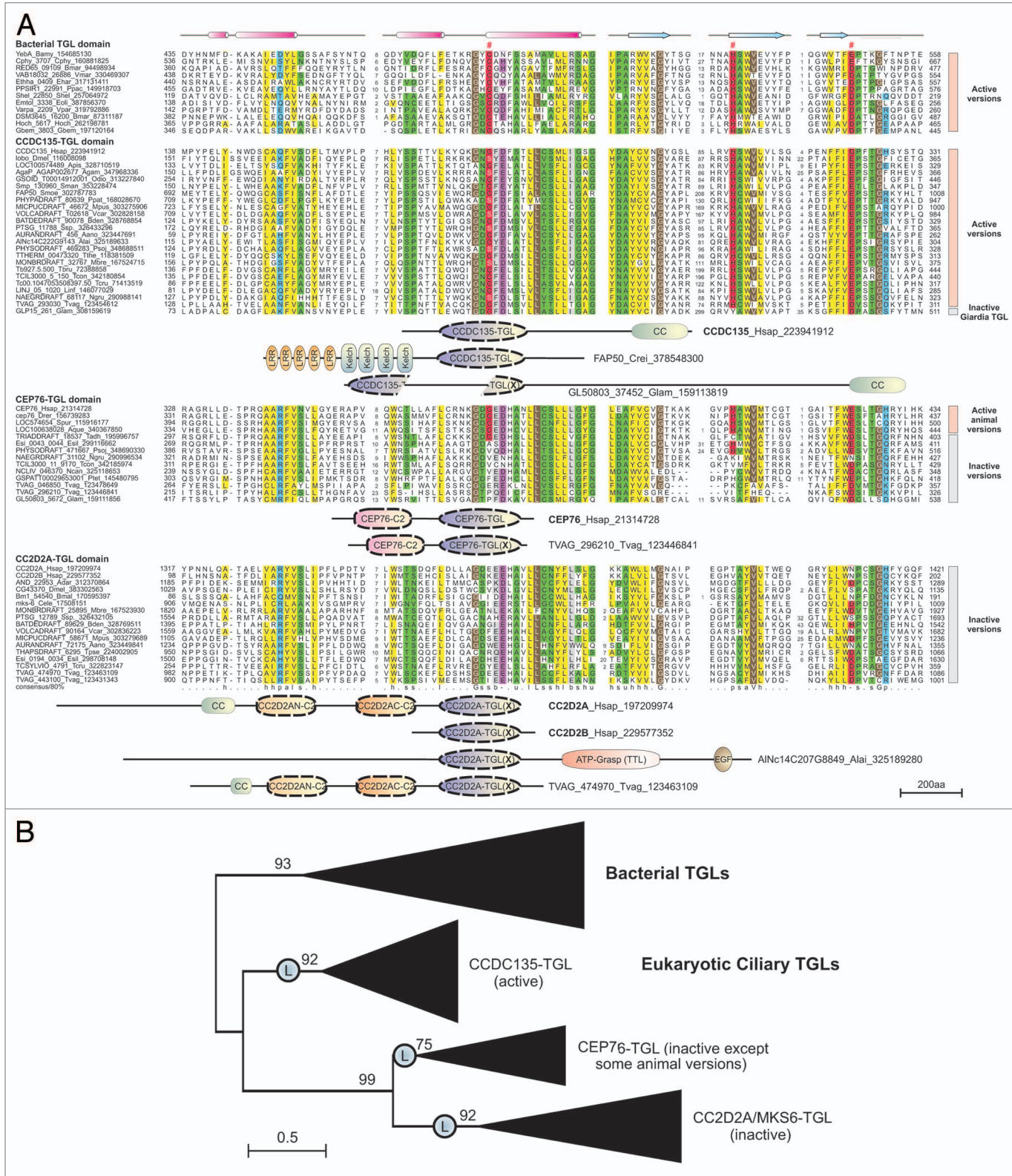


Figure 1A and B. (A) Multiple sequence alignment of the core region of TGL domains from bacterial and eukaryotic ciliary proteins. The catalytic triad residues (C/S, H, E/D) are labeled with number sign (#) and highlighted in red background. Secondary structures are shown with α -helices in pink and β -strands in light blue. Domain architectures of each ciliary TGL domains are shown below the alignment. Proteins are denoted by their gene name, species abbreviations and GI (GenBank Index) numbers separated by underscores. For species abbreviations, refer to the Materials and Methods section. (B) Divergent evolution of ciliary TGL domains prior to LECA. Capital L in the circle indicates the presence of the domain in the LECA. The tree was reconstructed using an approximately maximum-likelihood method implemented in the FastTree 2.1 program under default parameters. Bootstrap values are shown at each node.

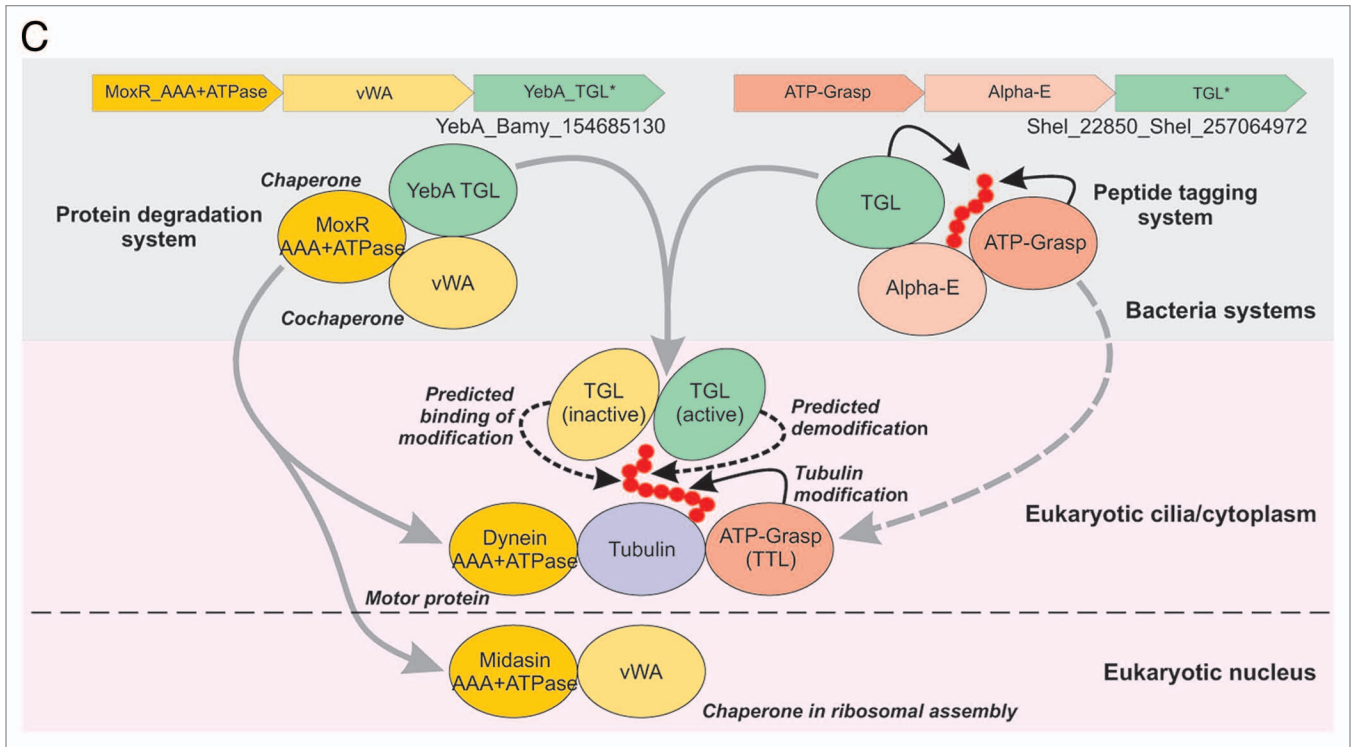


Figure 1C. A predicted model of eukaryotic TGLs involved in ciliary tubulin modifications and the evolutionary links to the bacterial protein degradation and peptide tagging systems. Canonical operons representing both bacterial protein degradation and peptide tagging systems are shown on the top. Light gray arrowed lines indicate gene transfer events and dark arrowed lines (and dashed ones) indicate known (and predicted) biochemical actions.

better characterize the functional association of the homologous prokaryotic TGL proteins. YebA-like TGL proteins are found in most members of the Gram-positive clades of bacteria, the lentisphaera-verrucomicrobia-planctomycetes clade, more sporadically in all other major bacterial lineages and in euryarchaea. This phyletic pattern is generally suggestive of an early origin in bacteria, followed by widespread dispersion through lateral transfer between different bacterial lineages, and also probably into the euryarchaea, early in their evolution. We found that the YebA-like proteins typically contain eight TM regions with the TGL domain occurring in the predicted cytoplasmic loop between transmembrane (TM) 7 and TM 8 (not shown). This suggests that the peptidase activity of the YebA protein operates in close proximity to the membrane. Further analysis of gene neighborhoods (Fig. 1C) revealed a strict association of the YebA-like TGL genes with two other genes, respectively coding for an AAA+ ATPase of the MoxR family³⁴ and a von Willebrand factor A (vWA) protein in both archaea and bacteria. In some cases, the genes for the MoxR AAA+ ATPase and the vWA protein are fused into a single gene. Despite this three-gene operon being very common in prokaryotes, the functional significance of the encoded proteins has not been hitherto explained. Clues regarding their potential function emerge from comparisons with other AAA+ ATPases. First, on at least six independent occasions, different versions of the AAA+ domain from across this superfamily have been functionally combined with different types of peptidase domains to constitute ATP-dependent protein degradation

systems such as the proteasome, HslUV, ClpXP, FtsH, Lon and YifB.³⁴ Based on this, we propose that this version of the MoxR AAA+ ATPase functions as an ATP-dependent protein unwinding engine that facilitates protein cleavage by YebA-like TGL domains. Second, the vWA domain was earlier shown to have strong functional linkages with multiple AAA+ domains belonging to the clade that unites dynein-midasin, MoxR, YifB and the chelataes.³⁴ While based on the precedence of the chelataes, the vWA domain was proposed to be required for metal insertion in substrates;⁴⁹ the new evidence from midasin and the archaeal phage tail chaperones^{50,51} suggests that it is likely to function as metal-dependent substrate-binding co-chaperones for this clade of AAA+ domains. Thus, we infer that the conserved MoxR-vWA-YebA operon is likely to encode a membrane-proximal protein degradation system (Fig. 1C), in which the vWA domain protein recruits substrates, while the MoxR AAA+ ATPase unwinds them, and the YebA-like TGL domain cleaves the unwound polypeptide. Another group of homologous prokaryotic TGL domains recovered in the searches with the ciliary TGL domains are those which we had previously described as being combined with genes encoding predicted peptide-ligases of the ATP-grasp fold and also a conserved protein termed the α -E domain (Fig. 1C).⁵² Here, too, the TGL domains are predicted to function as peptidases that reverse the peptide tags ligated by the ATP-grasp ligase (comparable to the peptide tags such polyglutamate, polyglycine and tyrosine ligated by the tubulin-modifying TTL family of ATP-grasp ligases).⁵²

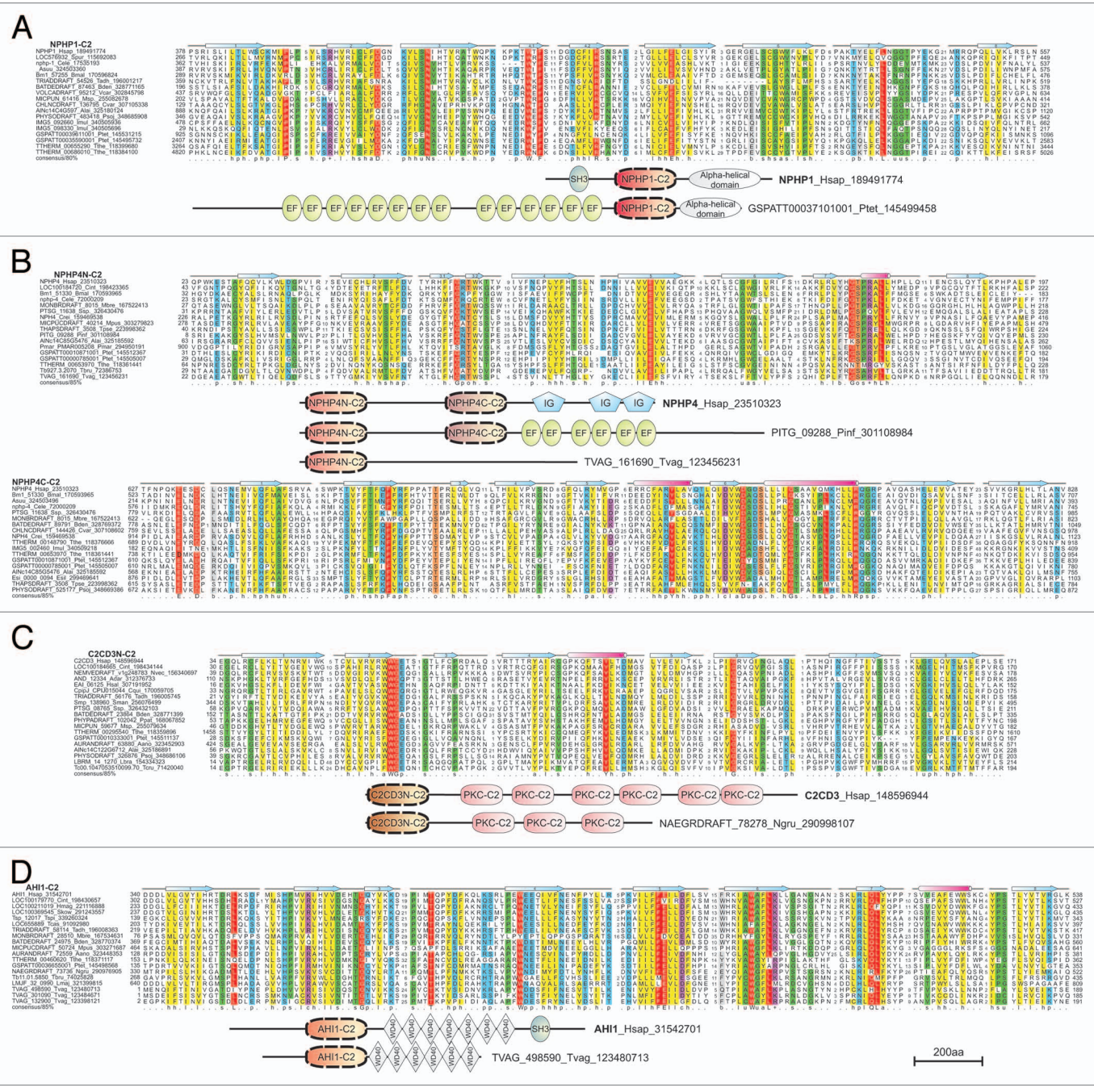
These observations have notable implications for both the function and evolution of the eukaryotic ciliary TGL proteins. Analysis of protein-protein interaction networks using the FUNCoup program⁵³ points to an interaction between the MKS and the dynein complexes with a direct edge between the latter and CC2D2A/MKS6. This suggests that at least in part the functional connection between the AAA+ ATPase and the TGL domain protein is paralleled in both prokaryotes and eukaryotes. Importantly, the MoxR AAA+ domain is the closest prokaryotic homolog of the six AAA+ domains in dynein and midasin and is specifically united to them by several sequence features.³⁴ MoxR appears to have given rise to a protein with six tandem AAA+ ATPase domains early in eukaryotic evolution, which further diverged to dynein and midasin.³⁴ In the course of this event, there appears to have been several changes in the original “prokaryotic” configuration of interacting partners. Midasin retained the vWA domain and functioned primarily as a chaperone for the removal of ribosomal assembly factors and the assembled ribosome from the nucleolus and nucleus (Fig. 1C). On the other hand, dynein appears to have lost the vWA domain while retaining its interaction with TGL domain proteins (Fig. 1C). The loss of the vWA co-chaperone probably facilitated its transition from a protein unfolding engine to a motor driving movement and trafficking. Indeed, the diversification of midasin (with a predominantly nuclear function) and dynein (with an ancient ciliary function), is likely to correspond to the formation of distinct nuclear and ciliary compartments in the eukaryotic progenitor. The functional shift in dynein relative to the ancestral MoxR ATPase probably allowed the TGL domain proteins to acquire certain independent functions, which is probably illustrated by the inactive versions (e.g., CC2D2A/MKS6) that are predicted to merely function as peptide-binding domains. The versions of the TGL found in the prokaryotic operons with the ATP-grasp peptide ligases also provide clues for the catalytically active eukaryotic ciliary TGL domains (e.g., CCDC135). Interestingly, in certain eukaryotes, such as stramenopiles, this TGL domain is fused to a related polyglutamylase-like ATP-grasp domain (Fig. 1A). This leads to the interesting possibility that the peptide tags on tubulin, such as polyglutamate and the C-terminal tyrosine, are reversed by the peptidase action of the active TGL domains, whereas the inactive versions merely bind them (Fig. 1C).

Identification of novel C2 domains in multiple components of the NPHP1-4-8 and MKS complexes. The functional parallel between the prokaryotic system, which combines MoxR with a membrane-linked TGL domain, and the eukaryotic dynein complex and ciliary TGLs suggested that membrane association of the axonemal and motor proteins might have been an early event in the origin of the cilia. Whereas the prokaryotic TGL protein associates with the membrane using TM helices, the CC2D2A/MKS6 is predicted to associate with membranes via a C2 domain. The presence of the C2 domain in CC2D2A also parallels the divergent C2 domains (called B9-C2 domains), which we had earlier described in the three paralogous components of the MKS complex, namely MKS1/BBS13, MKS9/MKSR-1/B9D1 and MKS10/MKSR-2/B9D2/Stumpy.³⁷ The prediction that these B9-C2 domains are critical for the membrane-association of the

ciliary basal body has been supported by multiple recent studies.^{16,54} First, several lines of genetic and biochemical evidence suggest that the B9-C2 domain proteins act early in ciliogenesis, in the period between the maturation of the basal body and its docking with the membrane. Second, mutations in the B9-C2 domains of MKS1 and MKS9, which are predicted to disrupt their structure, result in large scale loss of cilia in mutant cells.^{16,54} The presence of two distinct types of C2 domains in different components of the MKS complex hinted to us that diversification of the C2 domain might have been an important event in the evolution of these complexes. Given the extreme divergence of the different C2 domains, we used both sequence-profile searches and profile-profile comparison to search for additional C2 domains in the components of the MKS and NPHP complexes. Consequently, we have identified 10 new versions of C2 domains in several ciliary protein families.

Sequence comparisons revealed that the previously known C2 domain in CC2D2A/MKS6 belongs to a distinct clade of C2 domains (CC2D2AC-C2, Fig. 2A; see below). Further, we found that the N-terminal uncharacterized globular domain found in CC2D2A/MKS6 recovers C2 domains as the best hits in a HHpred search (e.g., protein kinase C-epsilon C2 domain, PDB: 1gmi; probability 95%; $p = 10^{-10}$). An alignment against a panel of diverse, known C2 domains showed that it indeed bears definitive features typical of them, confirming it to be a previously unknown version of the C2 domain superfamily (Fig. 2A). Similarly, we found a previously uncharacterized globular domain N terminal to the TGL domain in CEP76 (Fig. 2B). Profile-profile searches with this domain recovered C2 domains with significant scores (e.g., the myoferlin C2 domain, 2dmh; probability 97%; $p = 10^{-8}$), indicating the presence of yet another undetected C2 domain. Thus, taken together, we were able to identify four distinct varieties of the C2 domain (B9-C2, CC2D2AN-C2, CC2D2AC-C2, CEP76-C2) in the MKS complex and associated centriolar proteins.

Double mutants with disruptions of one component each, respectively from the MKS and NPHP1-4-8 complexes, display loss of the Y-shaped linkers in the ciliary transitional zone, which link the outer microtubule doublets to the ciliary membrane.^{6,16} This suggests that the membrane-anchoring function of the MKS complex acts in conjunction with the NPHP 1-4-8 complex.¹⁶ Interestingly, even RPGRIP1L/NPHP8/MKS5 shows a classical PKC-C2 domain in its central region,⁵⁵ suggesting that direct membrane-association via C2 domains might be a feature that extends to the NPHP 1-4-8 complex. Given that there are several uncharacterized globular domains in components of the NPHP 1-4-8 complex, we analyzed them further with sequence profile, HMM and profile-profile comparison searches. As a result, we recovered two additional distinctive, divergent C2 domains flanking the central PKC-C2 domain in NPHP8 and its paralog RPGRIP1 (RPGRIP1N-C2 and RPGRIP1C-C2; Fig. 2C). By means of further searches, we found that NPHP1⁵⁶ had one novel C2 domain in the central region of the protein (NPHP1-C2; Fig. 3A), whereas NPHP4⁵⁷ had two previously undetected, divergent C2 domains in the N terminus and the central region of the protein (NPHP4N-C2 and NPHP4C-C2; Fig. 3B). Thus,



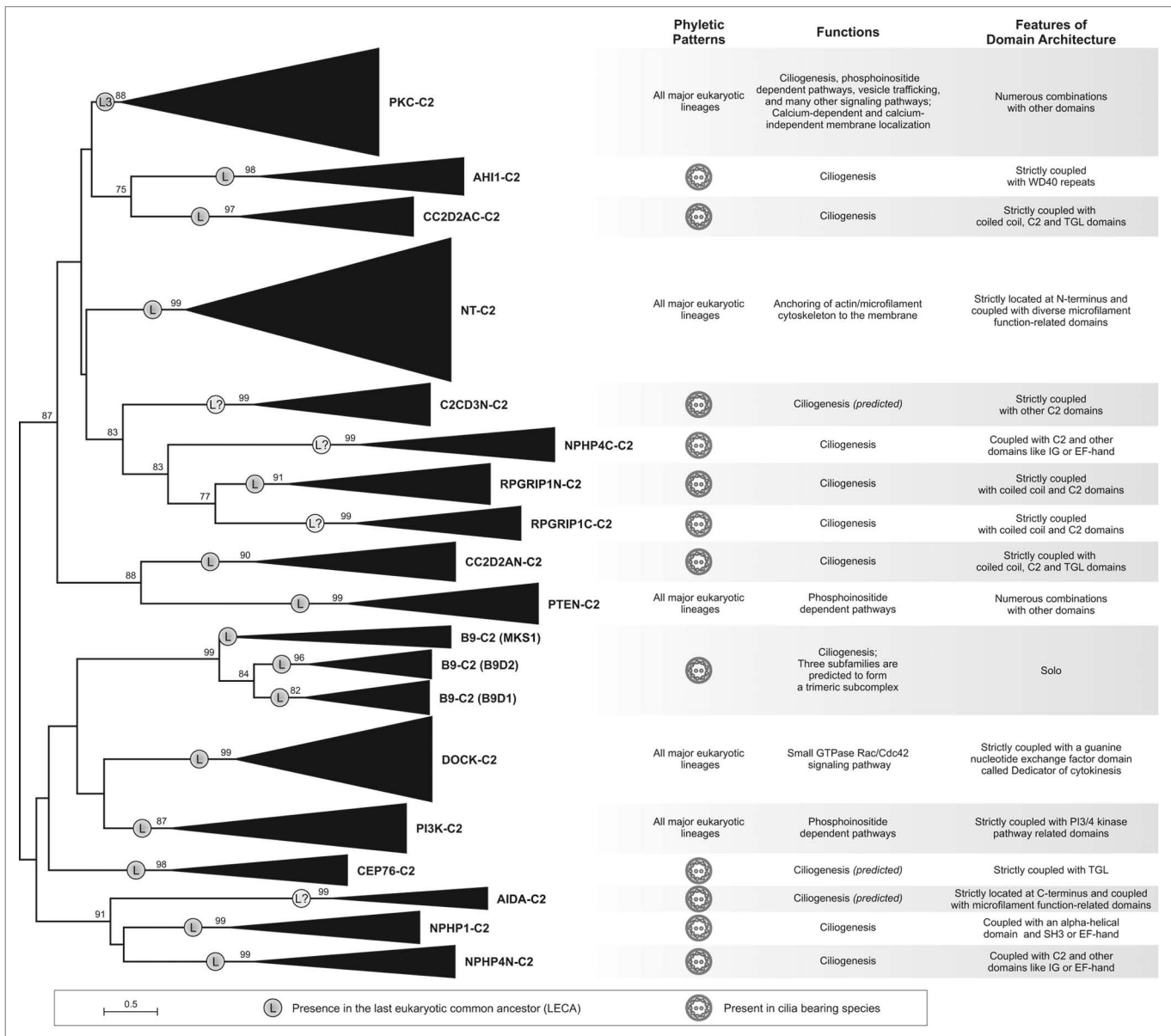


Figure 4. Evolutionary relationship of the different C2 domain families and the comparison of their phyletic patterns, functions and features of domain architecture. The tree was reconstructed using an approximately maximum-likelihood method implemented in the FastTree 2.1 program under default parameters. Nodes supported with bootstrap values greater than 75% are shown.

versions formed well-supported branches that were distinct from the classical PKC-C2 domains and other ancient versions, such as NT-C2 (Fig. 4), which was previously implicated in anchoring the actin cytoskeleton.³⁷ The monophyly of the distinct C2 clades found among the ciliogenesis components was also supported by the unique sequence signatures of each of the groups (Figs. 2 and 3). Together, these observations indicated that a notable part of the early radiation of the C2 domains, which resulted in the emergence of the key ciliogenesis complexes, probably happened between the time of the first eukaryotic common ancestor (FECA) and the LECA. Interestingly, this reconstruction also showed that the majority of the C2 domains, which can be potentially traced back to the LECA, are specifically associated

with ciliogenesis. Hence, it raises the possibility that the original radiation the C2 domains in eukaryotes happened in the context of the membrane-association of the ciliary cytoskeleton.

Functional implications of diverse C2 domain proteins in ciliogenesis. While the C2 domains are a strong predictor of membrane interactions,³⁷ the versions that are potentially traceable to the LECA show considerable sequence diversification and also diversity of domain architectures (Figs. 2 and 3). Hence, there is likely to have been a functional diversification of the C2 domains and the proteins containing them even at the time of the LECA. Of these, the pre-LECA triplication of proteins with B9-C2 domains and strict maintenance of these three distinct representatives throughout eukaryotes was interpreted

as implying that the three proteins form a trimeric subcomplex within the MKS complex, whose subunit stoichiometry is strongly conserved. This complex might be compared with the other complexes in eukaryotes with multiple paralogous subunits forming symmetrical toroidal or spherical structures with fixed subunit stoichiometry,⁶⁰ e.g., the CCT/TRiC-complex, which also forms part of the ciliary BBSome,⁶¹ the Rad9–1–1 complex in DNA repair⁶² and the core proteasome.⁶³ Consistent with this, the B9–C2 proteins, as a rule, have simple domain architectures with no combinations with other domains.³⁷ Hence, we predict that the B9–C2 is likely to form a toroidal structure at the heart of the MKS complex. In contrast, the C2 domains of the CC2D2A/MKS6 and CEP76 proteins occur in multidomain architectures combined with TGL domains. Hence, these proteins are more likely to function as adaptors, with the C2 domains binding the membrane and their catalytically inactive TGL domains (e.g., those in CC2D2A/MKS6 and CC2D2B) probably binding peptides from ciliary cytoplasm or axonemal proteins (Fig. 2A and B). The presence of coiled coil segments in these proteins is also suggestive of dimerization or interaction with other structural components with coiled coil segments (e.g., components of the IFT complex).¹³ A similar scenario is likely for the RPGRIP1/NPHP8, which combine the C2 domains with N-terminal coiled-coil segments (Fig. 2C). In light of their proposed role in forming scaffolds in the ciliary transition zone,⁶⁴ we propose that they are likely to be important as membrane anchors for targets bound by the MKS complex for intra-ciliary transport or signaling via the coiled-coil regions in these proteins. AHI1/Jouberin could also function similarly with the WD40 domains present C-terminal to the C2 domain playing a role in recruiting specific target proteins to the ciliary membrane (Fig. 3D).

Interestingly, we also found evidence for considerable plasticity in the domain architectures of these proteins beyond the conserved core, which includes the C2 domain. For example, in the case of AHI1/Jouberin, the conserved core, which is traceable to the LECA, includes the C2 domain and a WD40 β -propeller. In subsequent eukaryotic evolution, SH3 domain seems to have been added at the C terminus (Fig. 3D). In the case of the NPHP4, the C-terminal region contains immunoglobulin (IG) domains in certain eukaryotes, and EF-hand domains in other eukaryotes (Fig. 3B). On the other hand, in NPHP1 the N-terminal region is variable, with certain eukaryotic clades showing N-terminal SH3 domains, while others show EF-hand domains (Fig. 3A). The presence of the EF-hand domains in multiple proteins in these complexes suggest that, at least in certain eukaryotes, they are likely to respond to the presence of Ca^{2+} . This is consistent with the recruitment of two highly conserved paralogs of the Ca^{2+} -binding EF-hand protein centrin associated with the contractile function of the microtubular skeleton right from the time of the LECA.⁶⁵ However, examination of the sequence alignments suggests that, interestingly, none of the C2 domains traceable to the ciliary complexes in the LECA have Ca^{2+} -binding ability. Thus, the Ca^{2+} -binding role of the C2 domain (PKC-C2) is likely to have arisen independently of the ciliary function among C2 domains. The above-reported domain architectural variability of these ciliary proteins is consistent with the lineage-specific

diversification of target proteins that are transported to the cilium by the action of these proteins in different eukaryotic lineages. For example, the hedgehog-signaling pathway proteins such as patched, smoothed, SuFu and Gli, which are localized to the cilium, are only found in the animal lineage. Hence, the domain architectures of the primary components of the MKS and NPHP1–4–8 complexes indicate that C2 and TGL domains constitute the ancient core, which might be further extended accretion of different domains in particular eukaryotic lineages.

Based on the above analysis and the available genetic evidence, we predict that the C2 and TGL domain-containing proteins can be visualized as constituting the Y-shaped linkers of the ciliary transitional zone, with the arms of the Y-shaped linkers formed primarily by the coiled coil segments and the heads contacting the membrane being mainly comprised of the C2 domains. The three B9–C2 domains potentially form a central toroidal core at these membrane-contacting sites. On the other hand, the active and inactive TGL domains might constitute the axoneme- and substrate-interaction interfaces of the Y-shaped linkers. Thus, the combination of the C2 and TGL domains and coiled coil segments in these proteins explains their role both as structural components and as gatekeepers of intra-ciliary transport.

Evolutionary implications for origin of eukaryotic cilia and general conclusions. Several recent studies have aimed at tracing the provenance of key eukaryotic cellular components to the prokaryotic superkingdoms.^{29–31,36,52,66,67} For example, both actin and tubulin have been infrequently found in a small number of archaeal and bacterial lineages. Phylogenetic analysis has been used to argue that the archaea might have been the source of tubulin.³⁶ Based on the presence of tubulin in thaumarchaea and related lineages, it has been argued that the eukaryotic microtubular skeleton was acquired from an archaeal progenitor of the eukaryotes that resembled a thaumarchaeon. This possibility is consistent with the presence of other eukaryote-like features in thaumarchaea, such as the ubiquitin (Ub) system^{68,69} and actin-like proteins.³⁶ However, it should be borne in mind that many of these eukaryote-like features are not found in the same thaumarchaeon, are also rather infrequently found in prokaryotes and are encoded by potentially mobile operons (as demonstrated in the case of the Ub-system⁶⁹). Hence, while a thaumarchaeon-like organism could have been the archaeal progenitor of eukaryotes, it is possible that this progenitor did not have all the eukaryote-like genes together in its genome right from the inception, but accreted them over time via lateral transfer. Indeed, such gradual accretion of numerous, diverse mobile operons is observed in certain bacteria with large genomes, such as colonial myxobacteria.⁵² This accretion then created conditions that brought together diverse systems, allowing their “mixing and matching” to generate novel systems such as the cilium. In particular, the origin of the common ancestor of dyneins and midasins from the mobile MoxR-like systems favors a scenario, where the eukaryotic progenitor acquired additional systems by lateral transfer either from other co-occurring archaea and bacteria or from the bacterial endosymbiont that gave rise to the mitochondrion. Our observation of the presence of a functional linkage between the dynein complex and the ciliary TGL domains (Fig. 1C), which

is mirrored in the mobile prokaryotic MoxR-like systems with membrane-linked TGL domains, suggests that the acquisition of the MoxR-like precursor of dynein and the functionally linked TGL domains might have been a critical event for the emergence of the eukaryote-type cilia from the microtubular cytoskeleton. This is also comparable to the earlier observation on the mobile Mgl operons of prokaryotes,⁷⁰ which encode a small GTPase (MglA) that is likely to have been the progenitor of the eukaryotic Arf-like proteins,^{29,35} and its GTPase-activating protein (MglB),⁷¹ which is a homolog of the dynein-light chain 7. Given the role of the Mgl operon in cell polarity and gliding motility in certain bacteria,⁷² it is possible that this system might have even had an initial role in ciliary localization and motility of the eukaryotic progenitor. Thus, the accretion of multiple systems derived from mobile operons, namely the MoxR-like system, perhaps with functionally linked transglutaminases, the Mgl system, the TTL enzyme (polyglutamylase) and mobile small GTPase progenitors of the Ran-Ras-Rho-Rab-like clade from bacteria (probably the mitochondrial precursor) provided the necessary raw material for the origin of the eukaryotic cilium.

In bacteria and certain archaea, the proteins of the FtsZ-tubulin superfamily form a polymeric ring that mediates cytokinesis^{73,74} or chromosome segregation,⁷⁵ while in thaumarchaea, there is no evidence for the FtsZ-tubulin superfamily proteins participate in cell division.⁷⁶ This could have “freed” the FtsZ-tubulin proteins in the lineage leading, eukaryotes to adopt alternative functions. Our above analysis suggests that it was the coming together of the above-described components with the tubulin-like proteins that triggered the emergence of the eukaryotic ciliary system. Multiple lines of evidence point to this event being closely linked to the emergence of the nuclear compartment, which is the quintessential feature of eukaryotes. In functional terms, loss of the ancient chromosome segregation mechanisms dependent on nucleic acid pumps in eukaryotes⁷⁷ was possibly the main factor that linked the emergence of the nucleus with the origin of cilia; both eukaryotic chromosome segregation and motility came to depend on the same contractile microtubular cytoskeleton. The major event in emergence of both the nuclear and ciliary compartments was the emergence of the localization system dependent on the karyopherins and the RAN GTPase.^{30,31} As noted above, this event is likely to have also been accompanied by the divergence of midasin and dynein from their common ancestor, with only dynein retaining the functional interactions with TGL domain proteins in proximity of the membrane, a key feature for the origin of the cilium. Nevertheless, the question remains as to what differentiated the nuclear and ciliary compartments if they utilized the same localization and gating mechanisms. Our identification of novel C2 domains in the two key complexes related to transition zone morphology, ciliogenesis and intra-ciliary transport aids in answering this question. While several nuclear membrane proteins have been identified, including those which can be confidently traced back to the LECA, none of them contain C2 domains.³⁰ In contrast, the core of both the MKS and NPH 1-4-8 complexes, which are central to ciliogenesis, have multiple C2 domains that can be traced back to the LECA (Fig. 4). This suggests that the radiation of the C2

domains was central to differentiation of the nuclear and ciliary compartments, with these domains shaping unique membrane-associated structures that came to define ciliary function. Thus, in a sense, in the case of ciliogenesis, ontology, i.e., development of the cilia dependent on the MKS and NPHP 1-4-8 complexes, follows evolution. Currently, C2 domains are not known outside of eukaryotes; hence, it is unclear if they were a eukaryotic innovation or emerged from a preexisting prokaryotic version. Nevertheless, it is clear that a major part of their early radiation specifically occurred in the context of the eukaryotic ciliary apparatus (Fig. 4). Given that number of lines of evidence point toward the bacterial endosymbiont and genetic material contributed by it being critical for the origin of the nucleus in the ancestral eukaryote,⁶⁶ it is likely that the main steps in the emergence of cilia also happened after the endosymbiotic event. Indeed, in support of this, the currently available evidence points toward a bacterial origin for the TTL ATP-grasp domains that are central to microtubular modifications in ciliary function.⁵²

In conclusion, the findings reported here offer certain key testable hypotheses regarding ciliogenesis and ciliary function. First, the identification of both inactive and potentially active versions of the TGL domain help identify an important determinant for protein-protein interactions in the ciliary transitional zone and predict a potential proteolytic processing activity in the ciliary compartment that could target microtubule modifications or transported proteins. Second, our discovery of multiple new C2 domains in ciliary components helps in a conceptual reconstruction of the Y-shaped linkers in the transitional zone, with multiple membrane-contacting sites that could help both in the association of ciliary cytoskeleton with the membrane and also membrane-linked intra-ciliary trafficking. Importantly, the identification of novel C2 domains in CEP76, AHI1/Jouberin and C2CD3 proteins supports their function in close proximity to the core MKS and NPHP 1-4-8 complexes. We hope that further tests of the predictions presented here would help in a better understanding of eukaryotic ciliogenesis and also clarify the biochemical basis for a number of human ciliopathies.

Materials and Methods

Iterative profile searches with the PSI-BLAST⁷⁸ and JACKHMMER⁷⁹ programs were used to retrieve homologous sequences in the protein non-redundant (NR) database at National Center for Biotechnology Information (NCBI). For most searches, a cut-off e-value of 0.01 was used to assess significance. In each iteration, the newly detected sequences that had e-values lower than the cut-off were examined for being false positives. Similarity-based clustering was performed using the BLASTCLUST program (<ftp://ftp.ncbi.nih.gov/blast/documents/blastclust.html>) to remove the highly similar sequences. Multiple sequence alignments were built by Kalign⁸⁰ and Muscle⁸¹ programs, followed by manual adjustments based on profile-profile alignment, secondary structure prediction and structural alignment. Consensus secondary structures were predicted using the JPred program.⁸² Protein remote homology relationship was detected by sequence-profile comparisons with the

PSI-BLAST program and profile-profile comparisons with the HHpred program.⁸³ Phylogenetic analysis was conducted using an approximately maximum-likelihood method implemented in the FastTree 2.1 program under default parameters.⁸⁴ The tree was rendered using the MEGA Tree Explorer.⁸⁵ For bacterial TGL genes, their gene neighborhoods were extracted and analyzed. The protein sequences of all neighbors were clustered using the BLASTCLUST program to identify related sequences in gene neighborhoods. Each cluster of homologous proteins was then assigned an annotation based on the domain architecture or shared conserved domain. A complete list of Genbank GIs for proteins investigated in this study are provided in the Supplemental Material.

Species abbreviations: Aaeg, *Aedes aegypti*; Aano, *Aureococcus anophagefferens*; Adar, *Anopheles darlingi*; Agam, *Anopheles gambiae*; Alai, *Albugo laibachii*; Amel, *Apis mellifera*; Apis, *Acyrtosiphon pisum*; Aque, *Amphimedon queenslandica*; Asuu, *Ascaris suum*; Bamy, *Bacillus amyloliquefaciens*; Bden, *Batrachochytrium dendrobatidis*; Bmal, *Brugia malayi*; Bmar, *Bermanella marisrubri*; Bmar, *Blastopirellula marina*; Cele, *Caenorhabditis elegans*; Cint, *Ciona intestinalis*; Cphy, *Clostridium phytofermentans*; Cqui, *Culex quinquefasciatus*; Crei, *Chlamydomonas reinhardtii*; Csin, *Clonorchis sinensis*; Cvar, *Chlorella variabilis*; Dmel, *Drosophila melanogaster*; Drer, *Danio rerio*; Ehar, *Ethanoligenens harbinense*; Eoli, *Emticicia oligotrophica*; Esil, *Ectocarpus siliculosus*; Gbem, *Geobacter bemi-djiensis*; Glam, *Giardia lamblia*; Hmag, *Hydra magnipapillata*;

Hoch, *Haliangium ochraceum*; Hsal, *Harpegnathos saltator*; Hsap, *Homo sapiens*; Imul, *Ichthyophthirius multifiliis*; Lbra, *Leishmania braziliensis*; Ldon, *Leishmania donovani*; Linf, *Leishmania infantum*; Lloa, *Loa loa*; Lmaj, *Leishmania major*; Mbre, *Monosiga brevicollis*; Mocc, *Metaseiulus occidentalis*; Mpus, *Micromonas pusilla*; Mrot, *Megachile rotundata*; Msp., *Micromonas sp*; Ncan, *Neospora caninum*; Ngru, *Naegleria gruberi*; Nvec, *Nematostella vectensis*; Odio, *Oikopleura dioica*; Pinf, *Phytophthora infestans*; Pmar, *Perkinsus marinus*; Ppac, *Plesiocystis pacifica*; Ppat, *Physcomitrella patens*; Psoj, *Phytophthora sojae*; Ptet, *Paramecium tetraurelia*; Rory, *Rhizopus oryzae*; Shel, *Slackia heliotrinireducens*; Skow, *Saccoglossus kowalevskii*; Sman, *Schistosoma mansoni*; Smoe, *Selaginella moellendorffii*; Spur, *Strongylocentrotus purpuratus*; Ssp., *Salpingoeca sp*; Tadh, *Trichoplax adhaerens*; Tbru, *Trypanosoma brucei*; Tcas, *Tribolium castaneum*; Tcon, *Trypanosoma congolense*; Tcru, *Trypanosoma cruzi*; Tpsc, *Thalassiosira pseudonana*; Tspi, *Trichinella spiralis*; Tthe, *Tetrahymena thermophila*; Tvag, *Trichomonas vaginalis*; Vcar, *Volvox carteri*; Vmar, *Verrucosispora maris*; Vpar, *Variovorax paradoxus*.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Supplemental Material

Supplemental material may be downloaded here:
www.landesbioscience.com/journals/cc/article/22068/

References

- Dentler WL. Microtubule-membrane interactions in cilia and flagella. *Int Rev Cytol* 1981; 72:1-47; PMID:7019129; [http://dx.doi.org/10.1016/S0074-7696\(08\)61193-6](http://dx.doi.org/10.1016/S0074-7696(08)61193-6).
- Singla V, Reiter JF. The primary cilium as the cell's antenna: signaling at a sensory organelle. *Science* 2006; 313:629-33; PMID:16888132; <http://dx.doi.org/10.1126/science.1124534>.
- Goetz SC, Anderson KV. The primary cilium: a signaling centre during vertebrate development. *Nat Rev Genet* 2010; 11:331-44; PMID:20395968; <http://dx.doi.org/10.1038/nrg2774>.
- Li JB, Gerdes JM, Haycraft CJ, Fan Y, Teslovich TM, May-Simera H, et al. Comparative genomics identifies a flagellar and basal body proteome that includes the BBS5 human disease gene. *Cell* 2004; 117:541-52; PMID:15137946; [http://dx.doi.org/10.1016/S0092-8674\(04\)00450-7](http://dx.doi.org/10.1016/S0092-8674(04)00450-7).
- Kim J, Lee JE, Heynen-Genel S, Suyama E, Ono K, Lee K, et al. Functional genomic screen for modulators of ciliogenesis and cilium length. *Nature* 2010; 464:1048-51; PMID:20393563; <http://dx.doi.org/10.1038/nature08895>.
- Czarnecki PG, Shah JV. The ciliary transition zone: from morphology and molecules to medicine. *Trends Cell Biol* 2012; 22:201-10; PMID:22401885; <http://dx.doi.org/10.1016/j.tcb.2012.02.001>.
- Heuser T, Raytchev M, Krell J, Porter ME, Nicastro D. The dynein regulatory complex is the nexin link and a major regulatory node in cilia and flagella. *J Cell Biol* 2009; 187:921-33; PMID:20008568; <http://dx.doi.org/10.1083/jcb.200908067>.
- Hu Q, Milenkovic L, Jin H, Scott MP, Nachury MV, Spiliotis ET, et al. A septin diffusion barrier at the base of the primary cilium maintains ciliary membrane protein distribution. *Science* 2010; 329:436-9; PMID:20558667; <http://dx.doi.org/10.1126/science.1191054>.
- Kim SK, Shindo A, Park TJ, Oh EC, Ghosh S, Gray RS, et al. Planar cell polarity acts through septins to control collective cell movement and ciliogenesis. *Science* 2010; 329:1337-40; PMID:20671153; <http://dx.doi.org/10.1126/science.1191184>.
- Janke C, Rogowski K, Wloga D, Regnard C, Kajava AV, Strub JM, et al. Tubulin polyglutamylase enzymes are members of the TTL domain protein family. *Science* 2005; 308:1758-62; PMID:15890843; <http://dx.doi.org/10.1126/science.1113010>.
- Jin H, Nachury MV. The BBSome. *Curr Biol* 2009; 19:R472-3; PMID:19549489.
- Rosenbaum JL, Witman GB. Intraflagellar transport. *Nat Rev Mol Cell Biol* 2002; 3:813-25; PMID:12415299; <http://dx.doi.org/10.1038/nrm952>.
- Taschner M, Bhogaraju S, Lorentzen E. Architecture and function of IFT complex proteins in ciliogenesis. *Differentiation* 2012; 83:S12-22; PMID:22118932.
- Sang L, Miller JJ, Corbit KC, Giles RH, Brauer MJ, Otto EA, et al. Mapping the NPHP-JBTS-MKS protein network reveals ciliopathy disease genes and pathways. *Cell* 2011; 145:513-28; PMID:21565611; <http://dx.doi.org/10.1016/j.cell.2011.04.019>.
- Williams CL, Li C, Kida K, Inglis PN, Mohan S, Semence L, et al. MKS and NPHP modules cooperate to establish basal body/transition zone membrane associations and ciliary gate function during ciliogenesis. *J Cell Biol* 2011; 192:1023-41; PMID:21422230; <http://dx.doi.org/10.1083/jcb.201012116>.
- Chih B, Liu P, Chinn Y, Chalouni C, Komuves LG, Hass PE, et al. A ciliopathy complex at the transition zone protects the cilia as a privileged membrane domain. *Nat Cell Biol* 2012; 14:61-72; PMID:22179047; <http://dx.doi.org/10.1038/ncb2410>.
- Garcia-Gonzalo FR, Corbit KC, Sirerol-Piquer MS, Ramaswami G, Otto EA, Noriega TR, et al. A transition zone complex regulates mammalian ciliogenesis and ciliary membrane composition. *Nat Genet* 2011; 43:776-84; PMID:21725307; <http://dx.doi.org/10.1038/ng.891>.
- Reiter JF, Blacque OE, Leroux MR. The base of the cilium: roles for transition fibres and the transition zone in ciliary formation, maintenance and compartmentalization. *EMBO Rep* 2012; 13:608-18; PMID:22653444; <http://dx.doi.org/10.1038/embo.2012.73>.
- Silverman MA, Leroux MR. Intraflagellar transport and the generation of dynamic, structurally and functionally diverse cilia. *Trends Cell Biol* 2009; 19:306-16; PMID:19560357; <http://dx.doi.org/10.1016/j.tcb.2009.04.002>.
- Zhao C, Malicki J. Nephrocystins and MKS proteins interact with IFT particle and facilitate transport of selected ciliary cargos. *EMBO J* 2011; 30:2532-44; PMID:21602787; <http://dx.doi.org/10.1038/emboj.2011.165>.
- Dawe HR, Smith UM, Cullinane AR, Gerrelli D, Cox P, Badano JL, et al. The Meckel-Gruber Syndrome proteins MKS1 and meckelin interact and are required for primary cilium formation. *Hum Mol Genet* 2007; 16:173-86; PMID:17185389; <http://dx.doi.org/10.1093/hmg/ddl459>.
- Jauregui AR, Nguyen KC, Hall DH, Barr MM. The *Caenorhabditis elegans* nephrocystins act as global modifiers of cilium structure. *J Cell Biol* 2008; 180:973-88; PMID:18316409; <http://dx.doi.org/10.1083/jcb.200707090>.
- Ou G, Blacque OE, Snow JJ, Leroux MR, Scholey JM. Functional coordination of intraflagellar transport motors. *Nature* 2005; 436:583-7; PMID:16049494; <http://dx.doi.org/10.1038/nature03818>.
- Lehtreck KF, Johnson EC, Sakai T, Cochran D, Ballif BA, Rush J, et al. The *Chlamydomonas reinhardtii* BBSome is an IFT cargo required for export of specific signaling proteins from flagella. *J Cell Biol* 2009; 187:1117-32; PMID:20038682; <http://dx.doi.org/10.1083/jcb.200909183>.

25. Hampl V, Hug L, Leigh JW, Dacks JB, Lang BF, Simpson AG, et al. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc Natl Acad Sci USA* 2009; 106:3859-64; PMID:19237557; <http://dx.doi.org/10.1073/pnas.0807880106>.
26. Hodges ME, Scheumann N, Wickstead B, Langdale JA, Gull K. Reconstructing the evolutionary history of the centriole from protein components. *J Cell Sci* 2010; 123:1407-13; PMID:20388734; <http://dx.doi.org/10.1242/jcs.064873>.
27. Carvalho-Santos Z, Azimzadeh J, Pereira-Leal JB, Betrencourt-Dias M. Evolution: Tracing the origins of centrioles, cilia, and flagella. *J Cell Biol* 2011; 194:165-75; PMID:21788366; <http://dx.doi.org/10.1083/jcb.201011152>.
28. Jin H, White SR, Shida T, Schulz S, Aguiar M, Gygi SP, et al. The conserved Bardet-Biedl syndrome proteins assemble a coat that traffics membrane proteins to cilia. *Cell* 2010; 141:1208-19; PMID:20603001; <http://dx.doi.org/10.1016/j.cell.2010.05.015>.
29. Jékely G. Small GTPases and the evolution of the eukaryotic cell. *Bioessays* 2003; 25:1129-38; PMID:14579253.
30. Mans BJ, Anantharaman V, Aravind L, Koonin EV. Comparative genomics, evolution and origins of the nuclear envelope and nuclear pore complex. *Cell Cycle* 2004; 3:1612-37; PMID:15611647; <http://dx.doi.org/10.4161/cc.3.12.1316>.
31. Jékely G, Arendt D. Evolution of intraflagellar transport from coated vesicles and autogenous origin of the eukaryotic cilium. *Bioessays* 2006; 28:191-8; PMID:16435301.
32. Kee HL, Dishinger JF, Blasius TL, Liu CJ, Margolis B, Verhey KJ. A size-exclusion permeability barrier and nucleoporins characterize a ciliary pore complex that regulates transport into cilia. *Nat Cell Biol* 2012; 14:431-7; PMID:22388888; <http://dx.doi.org/10.1038/ncb2450>.
33. Obado SO, Rout MP. Ciliary and nuclear transport: different places, similar routes? *Dev Cell* 2012; 22:693-4; PMID:22516195; <http://dx.doi.org/10.1016/j.devcel.2012.04.002>.
34. Iyer LM, Leipe DD, Koonin EV, Aravind L. Evolutionary history and higher order classification of AAA+ ATPases. *J Struct Biol* 2004; 146:11-31; PMID:15037234; <http://dx.doi.org/10.1016/j.jsb.2003.10.010>.
35. Leipe DD, Wolf YI, Koonin EV, Aravind L. Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol* 2002; 317:41-72; PMID:11916378; <http://dx.doi.org/10.1006/jmbi.2001.5378>.
36. Yutin N, Koonin EV. Archaeal origin of tubulin. *Biol Direct* 2012; 7:10; PMID:22458654; <http://dx.doi.org/10.1186/1745-6150-7-10>.
37. Zhang D, Aravind L. Identification of novel families and classification of the C2 domain superfamily elucidate the origin and evolution of membrane targeting activities in eukaryotes. *Gene* 2010; 469:18-30; PMID:20713135; <http://dx.doi.org/10.1016/j.gene.2010.08.006>.
38. Beatson S, Ponting CP. GIFT domains: linking eukaryotic intraflagellar transport and glycosylation to bacterial gliding. *Trends Biochem Sci* 2004; 29:396-9; PMID:15288869; <http://dx.doi.org/10.1016/j.tibs.2004.06.002>.
39. Wan H, Li L, Federhen S, Wootton JC. Discovering simple regions in biological sequences associated with scoring schemes. *J Comput Biol* 2003; 10:171-85; PMID:12804090; <http://dx.doi.org/10.1089/106652703321825955>.
40. Tallila J, Jakkula E, Peltonen L, Salonen R, Kestilä M. Identification of CC2D2A as a Meckel syndrome gene adds an important piece to the ciliopathy puzzle. *Am J Hum Genet* 2008; 82:1361-7; PMID:18513680; <http://dx.doi.org/10.1016/j.ajhg.2008.05.004>.
41. Tsang WY, Spektor A, Vijayakumar S, Bista BR, Li J, Sanchez I, et al. Cep76, a centrosomal protein that specifically restrains centriole reduplication. *Dev Cell* 2009; 16:649-60; PMID:19460342; <http://dx.doi.org/10.1016/j.devcel.2009.03.004>.
42. Yang Y, Cochran DA, Gargano MD, King I, Samhat NK, Burger BP, et al. Regulation of flagellar motility by the conserved flagellar protein CG34110/Ccdc135/FAP50. *Mol Biol Cell* 2011; 22:976-87; PMID:21289096; <http://dx.doi.org/10.1091/mbc.E10-04-0331>.
43. Anantharaman V, Aravind L. Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes. *Genome Biol* 2003; 4:R11; PMID:12620121; <http://dx.doi.org/10.1186/gb-2003-4-2-r11>.
44. Anantharaman V, Koonin EV, Aravind L. Peptide-N-glycanases and DNA repair proteins, Xp-C/Rad4, are, respectively, active and inactivated enzymes sharing a common transglutaminase fold. *Hum Mol Genet* 2001; 10:1627-30; PMID:11487565; <http://dx.doi.org/10.1093/hmg/10.16.1627>.
45. Iyer LM, Anantharaman V, Wolf MY, Aravind L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int J Parasitol* 2008; 38:1-31; PMID:17949725; <http://dx.doi.org/10.1016/j.ijpara.2007.07.018>.
46. Huang J, Gurung B, Wan B, Matkar S, Veniaminova NA, Wan K, et al. The same pocket in menin binds both MLL and JUND but has opposite effects on transcription. *Nature* 2012; 482:542-6; PMID:22327296; <http://dx.doi.org/10.1038/nature10806>.
47. Murai MJ, Chruszcz M, Reddy G, Grembecka J, Cierpicki T. Crystal structure of menin reveals binding site for mixed lineage leukemia (MLL) protein. *J Biol Chem* 2011; 286:31742-8; PMID:21757704; <http://dx.doi.org/10.1074/jbc.M111.258186>.
48. Hodder AN, Drew DR, Epa VC, Delorenzi M, Bourgon R, Miller SK, et al. Enzymic, phylogenetic, and structural characterization of the unusual papain-like protease domain of Plasmodium falciparum SERA5. *J Biol Chem* 2003; 278:48169-77; PMID:13679369; <http://dx.doi.org/10.1074/jbc.M306755200>.
49. Snider J, Houry WA. MoxR AAA+ ATPases: a novel family of molecular chaperones? *J Struct Biol* 2006; 156:200-9; PMID:16677824; <http://dx.doi.org/10.1016/j.jsb.2006.02.009>.
50. Ulbrich C, Diepholz M, Bassler J, Kressler D, Pertschy B, Galani K, et al. Mechanochemical removal of ribosome biogenesis factors from nascent 60S ribosomal subunits. *Cell* 2009; 138:911-22; PMID:19737519; <http://dx.doi.org/10.1016/j.cell.2009.06.045>.
51. Scheele U, Erdmann S, Ungewickell EJ, Felisberto-Rodrigues C, Ortiz-Lombardia M, Garrett RA. Chaperone role for proteins p618 and p892 in the extracellular tail development of Acidianus two-tailed virus. *J Virol* 2011; 85:4812-21; PMID:21367903; <http://dx.doi.org/10.1128/JVI.00072-11>.
52. Iyer LM, Abhiman S, Maxwell Burroughs A, Aravind L. Amidoligases with ATP-grasp, glutamine synthetase-like and acetyltransferase-like domains: synthesis of novel metabolites and peptide modifications of proteins. *Mol Biosyst* 2009; 5:1636-60; PMID:20023723; <http://dx.doi.org/10.1039/b917682a>.
53. Alexeyenko A, Schmitt T, Tjårnberg A, Guala D, Frings O, Sonhammer EL. Comparative interactomics with Funcoup 2.0. *Nucleic Acids Res* 2012; 40(Database issue):D821-8; PMID:22110034; <http://dx.doi.org/10.1093/nar/gkr1062>.
54. Dowdle WE, Robinson JF, Kneist A, Sierol-Piquer MS, Frints SG, Corbit KC, et al. Disruption of a ciliary B9 protein complex causes Meckel syndrome. *Am J Hum Genet* 2011; 89:94-110; PMID:21763481; <http://dx.doi.org/10.1016/j.ajhg.2011.06.003>.
55. Arts HH, Doherty D, van Beersum SE, Parisi MA, Letteboer SJ, Gorden NT, et al. Mutations in the gene encoding the basal body protein RPGRIPL1, a nephrocystin-4 interactor, cause Joubert syndrome. *Nat Genet* 2007; 39:882-8; PMID:17558407; <http://dx.doi.org/10.1038/ng2069>.
56. Parisi MA, Bennett CL, Eckert ML, Dobyns WB, Gleeson JG, Shaw DW, et al. The NPHP1 gene deletion associated with juvenile nephronophthisis is present in a subset of individuals with Joubert syndrome. *Am J Hum Genet* 2004; 75:82-91; PMID:15138899; <http://dx.doi.org/10.1086/421846>.
57. Wiik AC, Wade C, Biagi T, Ropstad EO, Bjerkås E, Lindblad-Toh K, et al. A deletion in nephronophthisis 4 (NPHP4) is associated with recessive cone-rod dystrophy in standard wire-haired dachshund. *Genome Res* 2008; 18:1415-21; PMID:18687878; <http://dx.doi.org/10.1101/gr.074302.107>.
58. Hoover AN, Wynkoop A, Zeng H, Jia J, Niswander LA, Liu A. C2cd3 is required for cilia formation and Hedgehog signaling in mouse. *Development* 2008; 135:4049-58; PMID:19004860; <http://dx.doi.org/10.1242/dev.029835>.
59. Dixon-Salazar T, Silhavy JL, Marsh SE, Louie CM, Scott LC, Gururaj A, et al. Mutations in the AHI1 gene, encoding joubertin, cause Joubert syndrome with cortical polymicrogyria. *Am J Hum Genet* 2004; 75:979-87; PMID:15467982; <http://dx.doi.org/10.1086/425985>.
60. Anantharaman V, Iyer LM, Aravind L. Comparative genomics of protists: new insights into the evolution of eukaryotic signal transduction and gene regulation. *Annu Rev Microbiol* 2007; 61:453-75; PMID:17506670; <http://dx.doi.org/10.1146/annurev.micro.61.080706.093309>.
61. Seo S, Baye LM, Schulz NP, Beck JS, Zhang Q, Slusarski DC, et al. BBS6, BBS10, and BBS12 form a complex with CCT/TRiC family chaperonins and mediate BBSome assembly. *Proc Natl Acad Sci USA* 2010; 107:1488-93; PMID:20080638; <http://dx.doi.org/10.1073/pnas.0910268107>.
62. Doré AS, Kilkenny ML, Rzechorzek NJ, Pearl LH. Crystal structure of the rad9-rad1-hus1 DNA damage checkpoint complex—implications for clamp loading and regulation. *Mol Cell* 2009; 34:735-45; PMID:19446481; <http://dx.doi.org/10.1016/j.molcel.2009.04.027>.
63. Smith DM, Benaroudj N, Goldberg A. Proteasomes and their associated ATPases: a destructive combination. *J Struct Biol* 2006; 156:72-83; PMID:16919475; <http://dx.doi.org/10.1016/j.jsb.2006.04.012>.
64. Coene KL, Mans DA, Boldt K, Gloeckner CJ, van Reeuwijk J, Bolat E, et al. The ciliopathy-associated protein homologs RPGRI1 and RPGRIPL1 are linked to cilium integrity through interaction with Nek4 serine/threonine kinase. *Hum Mol Genet* 2011; 20:3592-605; PMID:21685204; <http://dx.doi.org/10.1093/hmg/ddr280>.
65. Mahajan B, Selvapandian A, Gerald NJ, Majam V, Zheng H, Wickramarachchi T, et al. Centrin, cell cycle regulation proteins in human malaria parasite Plasmodium falciparum. *J Biol Chem* 2008; 283:31871-83; PMID:18693242; <http://dx.doi.org/10.1074/jbc.M800028200>.
66. Aravind L, Anantharaman V, Zhang D, De Souza RF, Iyer LM. Gene flow and biological conflict systems in the origin and evolution of eukaryotes. *Frontiers in Cellular and Infection Microbiology* 2012; 2.
67. Zhang D, de Souza RF, Anantharaman V, Iyer LM, Aravind L. Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biol Direct* 2012; 7:18; PMID:22731697.

68. Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H, et al. Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* 2011; 39:3204-23; PMID:21169198; <http://dx.doi.org/10.1093/nar/gkq1228>.
69. Burroughs AM, Iyer LM, Aravind L. Functional diversification of the RING finger and other binuclear treble clef domains in prokaryotes and the early evolution of the ubiquitin system. *Mol Biosyst* 2011; 7:2261-77; PMID:21547297; <http://dx.doi.org/10.1039/c1mb05061c>.
70. Koonin EV, Aravind L. Dynein light chains of the Roadblock/LC7 group belong to an ancient protein superfamily implicated in NTPase regulation. *Curr Biol* 2000; 10:R774-6; PMID:11084347.
71. Miertzschke M, Koerner C, Vetter IR, Keilberg D, Hot E, Leonardy S, et al. Structural analysis of the Ras-like G protein MglA and its cognate GAP MglB and implications for bacterial polarity. *EMBO J* 2011; 30:4185-97; PMID:21847100; <http://dx.doi.org/10.1038/emboj.2011.291>.
72. Zhang Y, Franco M, Ducret A, Mignot T. A bacterial Ras-like small GTP-binding protein and its cognate GAP establish a dynamic spatial polarity axis to control directed motility. *PLoS Biol* 2010; 8:e1000430; PMID:20652021; <http://dx.doi.org/10.1371/journal.pbio.1000430>.
73. Oliva MA, Martin-Galiano AJ, Sakaguchi Y, Andreu JM. Tubulin homolog TubZ in a phage-encoded partition system. *Proc Natl Acad Sci USA* 2012; 109:7711-6; PMID:22538818; <http://dx.doi.org/10.1073/pnas.1121546109>.
74. Busiek KK, Margolin W. Split decision: a thaumarchaeon encoding both FtsZ and Cdv cell division proteins chooses Cdv for cytokinesis. *Mol Microbiol* 2011; 82:535-8; PMID:21895799; <http://dx.doi.org/10.1111/j.1365-2958.2011.07833.x>.
75. Ni L, Xu W, Kumaraswami M, Schumacher MA. Plasmid protein TubR uses a distinct mode of HTH-DNA binding and recruits the prokaryotic tubulin homolog TubZ to effect DNA partition. *Proc Natl Acad Sci USA* 2010; 107:11763-8; PMID:20534443; <http://dx.doi.org/10.1073/pnas.1003817107>.
76. Pelve EA, Lindås AC, Martens-Habbena W, de la Torre JR, Stahl DA, Bernander R. Cdv-based cell division and cell cycle organization in the thaumarchaeon *Nitrosopumilus maritimus*. *Mol Microbiol* 2011; 82:555-66; PMID:21923770; <http://dx.doi.org/10.1111/j.1365-2958.2011.07834.x>.
77. Burroughs AM, Iyer LM, Aravind L. Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn* 2007; 3:48-65; PMID:18753784; <http://dx.doi.org/10.1159/000107603>.
78. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997; 25:3389-402; PMID:9254694; <http://dx.doi.org/10.1093/nar/25.17.3389>.
79. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* 2010; 11:431; PMID:20718988; <http://dx.doi.org/10.1186/1471-2105-11-431>.
80. Lassmann T, Sonnhammer EL. Kalign—an accurate and fast multiple sequence alignment algorithm. *BMC Bioinformatics* 2005; 6:298; PMID:16343337; <http://dx.doi.org/10.1186/1471-2105-6-298>.
81. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 2004; 32:1792-7; PMID:15034147; <http://dx.doi.org/10.1093/nar/gkh340>.
82. Cuff JA, Clamp ME, Siddiqui AS, Finlay M, Barton GJ. JPred: a consensus secondary structure prediction server. *Bioinformatics* 1998; 14:892-3; PMID:9927721; <http://dx.doi.org/10.1093/bioinformatics/14.10.892>.
83. Söding J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* 2005; 21:951-60; PMID:15531603; <http://dx.doi.org/10.1093/bioinformatics/bti125>.
84. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 2009; 26:1641-50; PMID:19377059; <http://dx.doi.org/10.1093/molbev/msp077>.
85. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007; 24:1596-9; PMID:17488738; <http://dx.doi.org/10.1093/molbev/msm092>.