BIOINFORMATICS DISCOVERY NOTE

Vol. 30 no. 7 2014, pages 899–902
doi:10.1093/bioinformatics/btt661

Structural bioinformatics

Advance Access publication November 20, 2013

A divergent calponin homology (NN–CH) domain defines a novel family: implications for evolution of ciliary IFT complex B proteins

Kenneth B. Schou[1,2,*], Jens S. Andersen[1] and Lotte B. Pedersen[2,*]

[1]Department of Biochemistry and Molecular Biology, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark and [2]Department of Biology, University of Copenhagen, Universitetsparken 13, DK-2100 Copenhagen OE, Denmark

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Microtubules are dynamic polymers of tubulin dimers that undergo continuous assembly and disassembly. A mounting number of microtubule-associated proteins (MAPs) regulate the dynamic behavior of microtubules and hence the assembly and disassembly of disparate microtubule structures within the cell. Despite recent advances in identification and functional characterization of MAPs, a substantial number of microtubule accessory factors have not been functionally annotated. Here, using profile-to-profile comparisons and structure modeling, we show that the yeast outer kinetochore components NDC80 and NUF2 share evolutionary ancestry with a novel protein family in mammals comprising, besides NDC80/HEC1 and NUF2, three Intraflagellar Transport (IFT) complex B subunits (IFT81, IFT57, CLUAP1) as well as six proteins with poorly defined function (FAM98A-C, CCDC22, CCDC93 and C14orf166). We show that these proteins consist of a divergent N-terminal calponin homology (CH)-like domain adjoined to an array of C-terminal heptad repeats predicted to form a coiled-coil arrangement. We have named the divergent CH-like domain NN–CH after the founding members NDC80 and NUF2.

Contact: kbschou@bio.ku.dk or lbpedersen@bio.ku.dk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on July 16, 2013; revised on October 11, 2013; accepted on November 10, 2013

1 INTRODUCTION

Microtubules are dynamic polymers that form a variety of cytoskeletal constellations in the eukaryotic cell. Common to all microtubule arrangements, ranging from the short centrioles of the centrosome to the elaborate structures of the mitotic spindle in mitosis or the axoneme of cilia, are a set of microtubule-associated proteins (MAPs) that influence the dynamics, stability and function of the tubulin polymer (Desai and Mitchison, 1997). Many putative MAPs and accessory components of the mitotic spindle, centrosome and cilia have not been functionally annotated. This is due in part to the high abundance of coiled-coil propensity proteins populating these subcellular compartments (e.g. Arnaiz *et al.*, 2009). Coiled-coil containing proteins are notoriously difficult to functionally annotate because of the high degree of sequence redundancy of coiled coils, which renders coiled-coil segments unsuitable for phylogenetic analysis. Some MAPs associated with e.g. kinetochores or centrosomes, however, adopt a bipartite structure comprising, besides their coiled coil hub, a globular head domain. The presence of such a conserved domain can greatly facilitate the identification of remote homologs in other species. For instance, in NDC80 (HEC1 in humans; hereafter referred to as NDC80) and NUF2 of the NDC80 kinetochore complex, or the microtubule plus end-tracking proteins EB1-3, the coiled-coil region is preceded by a conserved N-terminal calponin homology (CH) domain, which has aided in the identification of orthologs (Ciferri *et al.*, 2008; Hayashi and Ikura, 2003; Korenbaum and Rivero, 2002; Wei *et al.*, 2007). In the NDC80 kinetochore complex, formed by heterodimerization of NDC80 and NUF2, the interaction between the C-terminal coiled-coil regions produces a microtubule-binding interface consisting of two tightly interacting CH domains (Ciferri *et al.*, 2008; Wei *et al.*, 2007). Structural analysis has indicated that the NDC80 CH domain resembles that of the first CH domain of human fimbrin, which itself adopts a divergent CH domain fold (Wei *et al.*, 2007). Despite this structural similarity, previous sequence-based analyses have not identified NDC80 as a member the proposed major groups of CH domain-containing proteins (Gimona *et al.*, 2002).

Here we show by means of sequence profile-to-profile analysis that NDC80 and NUF2, three Intraflagellar Transport (IFT) complex B subunits, as well as six other poorly defined proteins constitute a novel family of bimodular proteins containing a divergent N-terminal CH-like domain followed by a C-terminal region of heptad repeats predicted to form a coiled coil arrangement. We have named the divergent CH-like domain NN–CH after the founding members NDC80 and NUF2.

2 METHODS

pBLAST and Iterative BLAST searches with PSI-BLAST were performed at NCBI (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and MPI bioinformatics toolkit (http://toolkit.tuebingen.mpg.de) servers (Biegert and Soding, 2009; Soding *et al.*, 2005), respectively, using the non-redundant (NR) protein sequence database at NCBI. Default settings were utilized in the searches. Unless otherwise stated, profile-to-profile hidden Markov model (HMM)–HMM searches were performed using HHpred (Soding *et al.*, 2005) with default settings against the PFAMA database (http://pfam.sanger.ac.uk). For Figure 1, multiple sequence alignments (MSA) were built by MAFFT (http://myhits.isb-sib.ch/cgi-bin/mafft) (Katoh *et al.*, 2002) and the resulting alignment edited in

*To whom correspondence should be addressed.

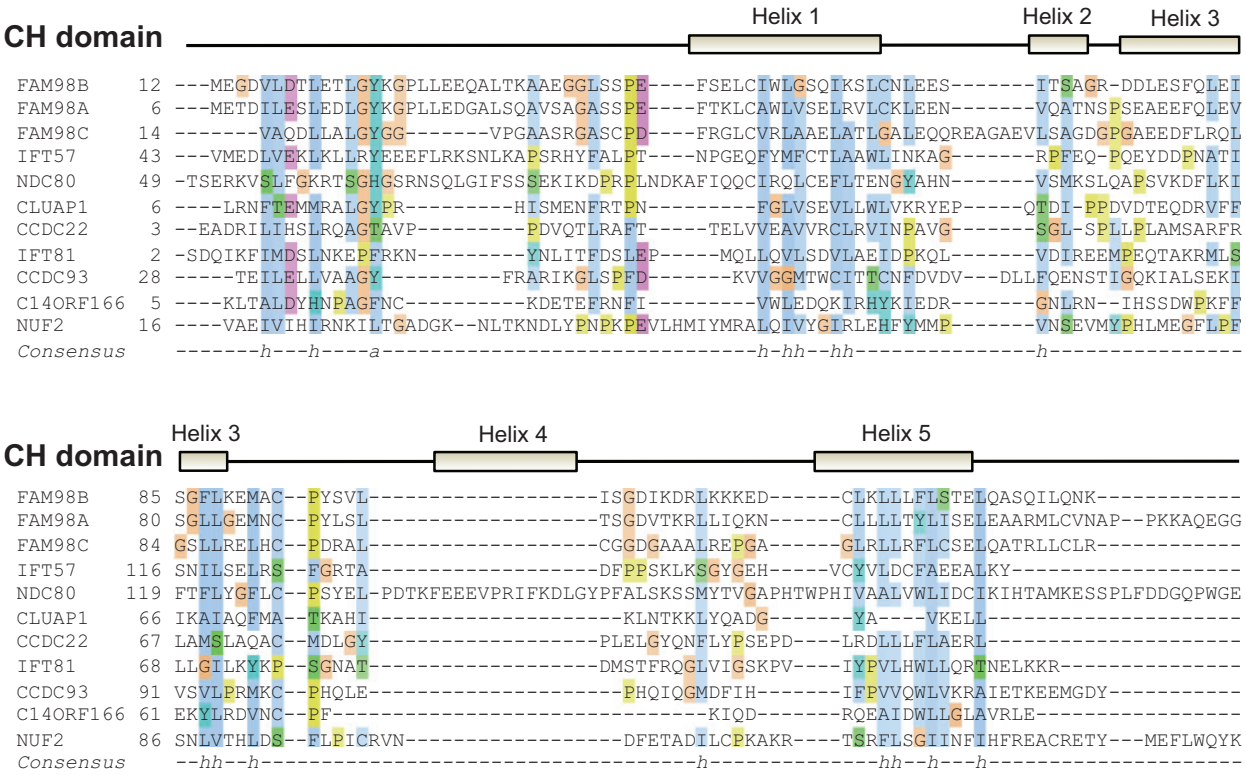

Fig. 1. Multiple alignment of the NN–CH domains from the 11 identified human proteins, designated by their UniProt identifiers. The limits (amino acid numbers) of the NN–CH domains are indicated. Coloring schemes as per Clustalx parametres in Jalview. Positions where hydrophobicity is conserved, which participate in the alpha-helices of the CH fold, are blue and designated *h* in the consensus sequence shown below the alignment. Aromatic residues are abbreviated *a*. Hydrophobic (h, blue), histidine/tyrosine (turquoise), acidic (violet), glycine (brown), proline (light green) and serine/threonine (dark green). Regions corresponding to alpha helices, predicted by HHpred and based on NDC80, are shown above the alignment (see also Supplementary Fig. S2). For accession numbers, see Supplementary Table S1

Jalview (http://www.jalview.org/). The consensus of the alignment was calculated and colored using the Clustalx color scheme implemented in Jalview. Secondary structure information and structural alignment were predicted using HHpred (Soding *et al.*, 2005). For Supplementary Figure S2 amino acid sequences were retrieved at NCBI (http://www.ncbi.nlm.nih.gov/protein) and aligned with the COBALT server (http://www.ncbi.nlm.nih.gov/tools/cobalt). Sequence alignments were analyzed for the occurrence of secondary protein structure in the Ali2D program (http://toolkit.tuebingen.mpg.de/ali2d). MODELLER (http://toolkit.tuebingen.mpg.de/modeller) (Sali *et al.*, 1995) was employed for homology modeling of 3D structures using templates chosen based on highest probability and significantly low *E*-value to the target sequence, as obtained in HHpred by mining the PDB and SCOP databases. Resulting 3D model coordinates were analyzed in Discovery Studio 3.5 Visualizer. Coiled-coil propensities were predicted as high propensity heptad repeats in Coils (http://embnet.vital-it.ch/software/COILS_form.html) (Lupas *et al.*, 1991) using weighted and unweighted search algorithms. Identified coiled-coil regions were further validated based on the occurrence of alpha-helical content (>50% helical) using the HNN server (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hnn.html).

3 RESULTS

In an effort to functionally annotate uncharacterized members of the core centrosome as well as accessory factors, we surveyed the protein architecture of previously identified and putative centrosomal proteins (Andersen *et al.*, 2003; Howng *et al.*, 2004; Jakobsen *et al.*, 2011). We noted that proteins such as C14orf166/CGI-99 and FAM98B bear homology by sequence to the N-terminus of kinetochore protein NDC80. High probability sequence similarities between, e.g. the N-terminus of human FAM98B (aa 9–128) with the N-terminus of NDC80 are apparent by HMM-based profile-to-profile searches using either local or global search algorithms in HHpred (local/global probabilities 92.46/90.60 and *E*-values 0.049/0.017) (Soding *et al.*, 2005). Interestingly, as with NDC80, the C-terminus of human FAM98B (aa 195–231) contains disparate arrays of heptad repeats (Fig. 2) as predicted by weighted and unweighted matrices in Coils (Lupas *et al.*, 1991), indicating that these proteins are members of a common evolutionary family of coiled-coil proteins.

To identify additional NDC80 paralog candidates, we first searched the PFAM database in HHpred using a HMM based on five reiterative PSI-BLAST searches initiated with the NN–CH domain of human NDC80 (aa 98–195). This search identified the DUF2465 domain (PFAM entry PF10239) of the FAM98 protein family as the highest scoring match (probability 95.14, $E = 0.018$), suggesting that FAM98 (A–C) carry a NN–CH domain. The alignment of DUF2465 was merged with the NDC80 alignment to initiate a subsequent HMM–HMM search

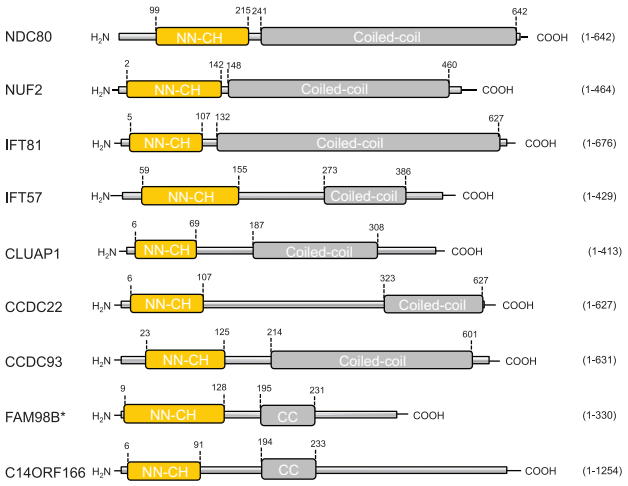

Fig. 2. Schematic domain organization of the identified NN–CH proteins. UniProt identifiers are indicated. CC, coiled-coil region. FAM98B* designates all three FAM98 isoforms in humans

against the PFAM database. We repeated this procedure once more, totalling three iterative HMM–HMM searches, to compile all the matched alignments that produced only high probability scores (>95). The resulting MSA was used to kick-start HMM–HMM searches in the PFAM database directed specifically against human proteins (Soding *et al.*, 2005). We thus identified 10 human proteins with high probability homology to the NDC80 N-terminus (Figs 1, 2 and Supplementary Table S1). Each protein was validated by reciprocal HMM–HMM searches for the occurrence of a conserved NN–CH domain. In each instance, the minimal N-terminal portion showing sequence homology to NDC80 was extracted and used as a query in HHpred. We confirmed the presence of the NN–CH domain in all 10 identified proteins with probability values >95, suggesting genuine homology to the NN–CH domain of NDC80. Alignment-based secondary structure predictions using the Ali2D program (http://toolkit.tuebingen.mpg.de/ali2d) supported these findings (Supplementary Fig. S2). Further, we could predict an N-terminal tertiary-structure model of each protein NN–CH domain that was compatible with the known 3D structure of the NDC80 N-terminus (Supplementary Fig. S1) (Wei *et al.*, 2007). Structural threading using MODELLER (Sali and Blundell, 1993) confirmed human NDC80 (2IGP) as the best, and statistically significant match in a search in the PDB database for similar tertiary structures (detailed list of all PDB entries found is available upon request).

To examine the NN–CH proteins for the co-occurrence of coiled-coil conformations, the C-terminal portion of each protein candidate was first probed for the presence of heptad repeats in Coils (Lupas *et al.*, 1991) and subsequently analyzed in HNN for high alpha-helical content coinciding with heptad repeat tendency. Interestingly, all newly identified NN–CH domain-containing proteins were predicted to harbor C-termini consisting of long alpha-helical matrices exhibiting a high-degree coiled-coil propensity (Fig. 2), suggesting that all the NN–CH protein derivatives structurally resemble NDC80.

Although the N-terminal domain of NDC80 was shown to adopt a CH domain fold (Wei *et al.*, 2007), our searches failed to detect any high-scoring matches to canonical CH domain-containing proteins such as EB1 (Hayashi and Ikura, 2003), consistent with previous sequence-based analyses of CH domain-containing proteins (Gimona *et al.*, 2002). Likewise, a HMM–HMM search seeded with the CH domain of human EB1 (Hayashi and Ikura, 2003) failed to recover significant matches to any of the newly identified NN–CH domain-bearing protein profiles in PFAM, indicating that the NN–CH domain represents a discrete branch of the CH domain protein family. In support of this, a comparative analysis of the history and evolution of the NN–CH family members, as assessed by pair-wise BLAST searches against representative proteomes, revealed that all NN–CH family members likely evolved from a precursor of yeast NDC80 and NUF2 (Supplementary Table S1).

The NDC80 and NUF2 proteins thus share common ancestry with all NN–CH domain derivatives. Based on the homology of NN–CH proteins to components of the yeast NDC80–NUF2 kinetochore complex, we propose that these proteins, including the ciliary IFT–B proteins identified in this survey, evolved, at least partly, as a specialized form of the kinetochore Ndc80p complex.

4 DISCUSSION AND CONCLUSIONS

For many coiled-coil proteins such as MAPs, components of centrosomes and the IFT–B complex of cilia, evolutionary annotation based on comparative analysis has been hampered by sequence redundancy of coiled coils and the lack of a conserved reference domain. This is in contrast, for example, to many IFT–A proteins, which display homology to vesicle coating and nuclear pore complex proteins (Jekely and Arendt, 2006; van Dam *et al.*, 2013). Here, using profile-to-profile searches, we show that three IFT–B proteins and several poorly characterized human proteins belong to a family of bimodular NN–CH and coiled-coil proteins that appears to share common ancestry with yeast NDC80. The NN–CH domains were identified as a novel branch of the CH domain. Structure analysis of NDC80 revealed that the NDC80 N-terminus adopts a CH domain fold (Ciferri *et al.*, 2008; Wei *et al.*, 2007). Indeed, fold searches seeded with NN–CH profiles against the PDB database recovered significant hits to the solved NDC80 structure as well as to canonical CH domains confirming that the NN–CH and CH domain adopt the same tertiary structure.

Several of the identified NN–CH proteins associate with microtubule-based structures, namely kinetochores/spindle microtubules, centrosomes and cilia. First, NDC80 and NUF2 bind directly to microtubules at the outer kinetochore (Tooley and Stukenberg, 2011) and a FAM98B ortholog was identified as a meiotic MAP in *Xenopus* (Gache *et al.*, 2010). Second, CCDC93 and C14orf166 co-fractionate or interact with centrosome proteins (Howng *et al.*, 2004; Jakobsen *et al.*, 2011). Third, IFT81, IFT57 and Cluap1 are components of the IFT–B complex in flagella of *Chlamydomonas reinhardtii* (Cole *et al.*, 1998) and other organisms (Pedersen and Rosenbaum, 2008), and are required for ciliogenesis (Botilde *et al.*, 2013; Li and Sun, 2011; Pasek *et al.*, 2012; Perkins *et al.*, 1986; Tsujikawa and Malicki, 2004). Since the NN–CH

domains of the NDC80–NUF2 dimer directly bind to kinetochore microtubules (Ciferri *et al*., 2008; Wei *et al*., 2007), it is likely that the other NN–CH family members similarly interact directly with microtubules, although association with other known CH domain interactors, such as F-actin, cannot be ruled out (Sjoblom *et al*., 2008). This has interesting implications for the three IFT–B proteins IFT57, IFT81 and Cluap1. IFT is required for the assembly and maintenance of virtually all eukaryotic cilia and flagella (Pedersen and Rosenbaum, 2008) and consists of large IFT protein complexes that fractionate as two biochemically distinct IFT–A and IFT–B sub-complexes. These are transported anterogradely and retrogradely along the ciliary axoneme, together with associated ciliary cargo proteins/building blocks, by kinesin-2 and cytoplasmic dynein-2 motors, respectively (Pedersen and Rosenbaum, 2008). In contrast to IFT–A proteins the origin of the majority of the IFT–B polypeptides is unknown (van Dam *et al*., 2013). Our finding that IFT81, IFT7 and Cluap1 share overall protein architecture with NDC80 and NUF2 indicates that they may have microtubule binding properties. Indeed, while this article was in review, an elegant study was published demonstrating that the amino terminus of IFT81 contains an NDC80-like CH domain (PDB entry 4LVP) that directly binds tubulin (Bhogaraju *et al*., 2013). Moreover, IFT81 was found to interact directly with IFT74 (Bhogaraju *et al*., 2013; Lucker *et al*., 2005), forming a dimeric complex that binds purified tubulin and microtubules with high affinity (Bhogaraju *et al*., 2013). Another recent study also indicated the presence of a CH domain in the N-terminus of IFT57 and IFT81 (Taschner *et al*., 2012), but a phylogenetic relationship with NDC80, NUF2 and other NN–CH domain proteins was not reported.

Collectively, we identified a novel protein family of 11 proteins in the human proteome that contain a novel CH domain, which we have named the NN–CH domain. The conserved bimodular protein composition of the NN–CH members, comprising a NN–CH domain and a coiled-coil region, and the association of these members with the kinetochore, centrosome and cilia suggest that the novel NN–CH protein family members function in various microtubule-associated activities.

Funding: Lundbeck Foundation; Novo Nordisk Foundation; Danish Council for Independent Research, Natural Sciences [grant number 10-085373]; Nordforsk.

Conflict of Interest: none declared

REFERENCES

Andersen,J.S. *et al*. (2003) Proteomic characterization of the human centrosome by protein correlation profiling. *Nature*, **426**, 570–574.

Arnaiz,O. *et al*. (2009) Cildb: a knowledgebase for centrosomes and cilia. *Database (Oxford)*, **2009**, bap022.

Bhogaraju,S. *et al*. (2013) Molecular basis of tubulin transport within the cilium by IFT74 and IFT81. *Science*, **341**, 1009–1012.

Biegert,A. and Soding,J. (2009) Sequence context-specific profiles for homology searching. *Proc. Natl Acad. Sci. USA*, **106**, 3770–3775.

Botilde,Y. *et al*. (2013) Cluap1 localizes preferentially to the base and tip of cilia and is required for ciliogenesis in the mouse embryo. *Dev. Biol*., **301**, 203–212.

Ciferri,C. *et al*. (2008) Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. *Cell*, **133**, 427–439.

Cole,D.G. *et al*. (1998) *Chlamydomonas* kinesin-II-dependent intraflagellar transport (IFT): IFT particles contain proteins required for ciliary assembly in *Caenorhabditis elegans* sensory neurons. *J. Cell Biol*., **141**, 993–1008.

Desai,A. and Mitchison,T.J. (1997) Microtubule polymerization dynamics. *Ann. Rev. Cell Dev. Biol*., **13**, 83–117.

Gache,V. *et al*. (2010) Xenopus meiotic microtubule-associated interactome. *PloS one*, **5**, e9248.

Gimona,M. *et al*. (2002) Functional plasticity of CH domains. *FEBS Lett*., **513**, 98–106.

Hayashi,I. and Ikura,M. (2003) Crystal structure of the amino-terminal microtubule-binding domain of end-binding protein 1 (EB1). *J. Biol. Chem*., **278**, 36430–36434.

Howng,S.L. *et al*. (2004) A novel ninein-interaction protein, CGI-99, blocks ninein phosphorylation by GSK3beta and is highly expressed in brain tumors. *FEBS lett*., **566**, 162–168.

Jakobsen,L. *et al*. (2011) Novel asymmetrically localizing components of human centrosomes identified by complementary proteomics methods. *Embo J*., **30**, 1520–1535.

Jekely,G. and Arendt,D. (2006) Evolution of intraflagellar transport from coated vesicles and autogenous origin of the eukaryotic cilium. *Bioessays*, **28**, 191–198.

Katoh,K. *et al*. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*., **30**, 3059–3066.

Korenbaum,E. and Rivero,F. (2002) Calponin homology domains at a glance. *J. Cell Sci*., **115**, 3543–3545.

Li,J. and Sun,Z. (2011) Qilin is essential for cilia assembly and normal kidney development in zebrafish. *PloS one*, **6**, e27365.

Lucker,B.F. *et al*. (2005) Characterization of the intraflagellar transport complex B core: direct interaction of the IFT81 and IFT74/72 subunits. *J. Biol. Chem*., **280**, 27688–27696.

Lupas,A. *et al*. (1991) Predicting coiled coils from protein sequences. *Science*, **252**, 1162–1164.

Pasek,R.C. *et al*. (2012) Mammalian Clusterin associated protein 1 is an evolutionarily conserved protein required for ciliogenesis. *Cilia*, **1**, 20.

Pedersen,L.B. and Rosenbaum,J.L. (2008) Intraflagellar transport (IFT) role in ciliary assembly, resorption and signalling. *Curr. Top. Dev. Biol*., **85**, 23–61.

Perkins,L.A. *et al*. (1986) Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol*., **117**, 456–487.

Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol*., **234**, 779–815.

Sali,A. *et al*. (1995) Evaluation of comparative protein modeling by MODELLER. *Proteins*, **23**, 318–326.

Sjoblom,B. *et al*. (2008) Novel structural insights into F-actin-binding and novel functions of calponin homology domains. *Curr. Opin. Struct. Biol*., **18**, 702–708.

Soding,J. *et al*. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*., **33**, W244–W248.

Taschner,M. *et al*. (2012) Architecture and function of IFT complex proteins in ciliogenesis. *Differentiation*, **83**, S12–S22.

Tooley,J. and Stukenberg,P.T. (2011) The Ndc80 complex: integrating the kinetochore's many movements. *Chromosome Res*., **19**, 377–391.

Tsujikawa,M. and Malicki,J. (2004) Intraflagellar transport genes are essential for differentiation and survival of vertebrate sensory neurons. *Neuron*, **42**, 703–716.

van Dam,T.J. *et al*. (2013) Evolution of modular intraflagellar transport from a coatomer-like progenitor. *Proc. Natl Acad. Sci. USA*, **110**, 6943–6948.

Wei,R.R. *et al*. (2007) The Ndc80/HEC1 complex is a contact point for kinetochore-microtubule attachment. *Nat. Struct. Mol. Biol*., **14**, 54–59.