Journal of Cell Science 107, 1875-1884 (1994)
Printed in Great Britain © The Company of Biologists Limited 1994

Molecular phylogeny of the kinesin family of microtubule motor proteins

Holly V. Goodson[1], Sang Joon Kang[2] and Sharyn A. Endow[2,*]

[1]Department of Biochemistry, Stanford University Medical School, Stanford, CA 94305, USA
[2]Department of Microbiology, Duke University Medical Center, Durham, NC 27710, USA

*Author for correspondence

SUMMARY

The rapidly expanding kinesin family of microtubule motor proteins includes proteins that are involved in diverse microtubule-based functions in the cell. Phylogenetic analysis of the motor regions of the kinesin proteins reveals at least five clearly defined groups that are likely to identify kinesins with different roles in basic cellular processes. Two of the groups are consistent with overall sequence similarity, while two groups contain proteins that are related in overall structure or function but show no significant sequence similarity outside the motor domain. One of these groups consists only of kinesin proteins with predicted C-terminal motor domains; another includes only kinesins required for mitotic spindle bipolarity. *Drosophila* Nod, presently an ungrouped protein, may represent a class of kinesins that, like the myosin I proteins, function as monomers. The analysis indicates that many types of kinesin proteins exist in eukaryotic organisms. At least two of the five groups identified in this analysis are expected to be present in most, or all, eukaryotes.

Key words: kinesin, microtubule motor proteins, evolutionary relationships

INTRODUCTION

Kinesin is a mechanochemical protein that binds to microtubules and translocates on microtubules in the presence of ATP. First detected in axoplasm of the squid (Allen et al., 1982, 1985), kinesin was subsequently identified (Brady, 1985; Vale et al., 1985a), purified (Vale et al., 1985a), and demonstrated to move unidirectionally on microtubules, toward the more dynamic plus ends (Vale et al., 1985b). The native protein consists of two heavy chains and two lights chains (Kuznetsov et al., 1988), with motor activity associated with the heavy chains. The motor domain, the region of the protein that contains proposed ATP- and microtubule-binding sites, is located at the N terminus of kinesin heavy chain (Yang et al., 1989).

For some 4-5 years after its discovery, kinesin was thought to be the only member of its class, in contrast to the multimember myosin and dynein motor protein families. However, the entry into the databases of the first kinesin heavy chain DNA sequence (Yang et al., 1989) facilitated the identification of proteins related to kinesin, resulting in the emergence of a burgeoning family of kinesin proteins. The first kinesin-related proteins to be discovered, *Saccharomyces cerevisiae* Kar3p (Meluh and Rose, 1990) and *Aspergillus nidulans* bimC (Enos and Morris, 1990), were found following molecular analysis of karyogamy and cell division mutants. Kinesin-related *Drosophila melanogaster* ncd was reported after molecular analysis of a mutant defective in meiotic and mitotic chromosome segregation (Endow et al., 1990), and was also found in a polymerase chain reaction (PCR) screen for kinesin-related proteins in *Drosophila* (McDonald and Goldstein, 1990). The use of PCR (Endow and Hatsumi, 1991; Stewart et al., 1991) and antipeptide antibodies directed against conserved regions of the kinesin motor domain (Cole et al., 1992; Sawin et al., 1992b) has led to the identification of multiple kinesin-related genes and proteins in different organisms, greatly expanding the kinesin family.

Members of the emerging family of kinesin proteins have been found in organisms that represent evolutionarily distant phyla in the protist, fungal, plant and animal kingdoms. The proteins contain a region with striking sequence similarity to the motor domain of kinesin heavy chain. Outside of this region, many of the proteins differ from kinesin heavy chain and from each other. The cellular functions of the kinesin proteins are likely to be diverse, based on available mutant phenotypes (Endow and Titus, 1992). The proteins have been implicated in vesicle transport in neuronal cells and are thought to act in organelle/vesicle transport in various cell types (Bloom, 1992). Kinesin proteins are also intimately involved in chromosome distribution and spindle function in meiosis and mitosis (Sawin and Endow, 1993), and have been found in organisms as distant as *Leishmania,* a kinetoplastid protist (Burns et al., 1993), and *Arabidopsis,* a higher plant (Mitsui et al., 1993).

The emergence of a large family of kinesin proteins with diverse cellular roles raises questions regarding the evolutionary relationships among the proteins, the distribution of the proteins, and the number of different kinesin proteins that are involved in a given microtubule-based function. In order to address these questions and provide a framework for thinking about the expanding family of kinesin proteins, we undertook a phylogenetic analysis of kinesin motor sequences. The

analysis provides evidence that the kinesin family of microtubule motor proteins consists of many groups that represent different types of kinesins, and that have members in widely divergent organisms. The ancient divergence of the groups suggests that the different types of kinesins perform distinct roles in basic cellular processes.

MATERIALS AND METHODS

Kinesin protein sequences

Current versions of the Genbank and EMBL databases were searched for protein sequences related to kinesin. Abbreviations and accession numbers for the resulting proteins are as follows: *S. cerevisiae* Kar3p (ScKar3, M31719), Cin8p (ScCin8, M90522), Kip1p (ScKip1, Z11962), Kip2p (ScKip2, Z11963), and Smy1p (ScSmy1, M69021); *Schizosaccharomyces pombe* cut7 (SpoCut7, X57513); *Aspergillus nidulans* bimC (AnBimC, M32075) and klpA (AnKlpA, X64603); *Arabidopsis thaliana* KatA (AtKatA, D11371); *Leishmania chagasi* kinesin (LcKin, S53127); *Strongylocentrotus purpuratus* kinesin heavy chain (SpKHC, X56844), KRP85 (formerly Kin2A, SpKRP85; L16993), and KRP95 (formerly Kin95, SpKRP95; U00996); *Loligo pealii* kinesin heavy chain (LpKHC, J05258); *Caenorhabditis elegans* kinesin heavy chain (CeKHC, L07144), Unc104 (CeUnc104, M58582), and Osm3 (CeOsm3, D14968); *D. melanogaster* kinesin heavy chain (DmKHC, M24441), Ncd (DmNcd, X52814), 3A7Kin (formerly Klp3A, Dm3A7Kin; L19117), Nod (DmNod, M36195), and Klp61F (formerly urchin, DmKlp61F; U01842); *Xenopus laevis* Eg5 (XlEg5, X71864); *Mus musculus* kinesin heavy chain (MmKHC, X61435), Kif2 (MmKif2, D12644), and Kif3 (MmKif3, D12645); *Homo sapiens* kinesin heavy chain (HsKHC, X65873), CENP-E (HsCENP-E, S46225), and MKLP1 (formerly CHO1, HsMklp1; X67155). These sequences can be obtained using the 'Retrieve' e-mail server operated by the United States National Center for Biotechnology Information (Henikoff, 1993). The *Cricetulus griseus* CHO2 (CgCHO2) protein sequence was generously made available by R. Kuriyama, *Chlamydomonas reinhardtii* Klp1 (CrKlp1) by M. Bernstein and J. Rosenbaum, *Rattus norvegicus* Krp2 (RnKrp2) by A. Sperry and S. Brady, *S. pombe* pkl1 (SpoPkl1) by A. Pidoux and Z. Cande, and *S. pombe* klp2 (SpoKlp2) by C. Troxell and R. McIntosh.

Nomenclature

Identification of new kinesin proteins of unknown function based on PCR and other molecular methods has resulted in similar names for proteins from different organisms. Descriptive mutant or other names have been retained where possible; however, many new kinesin proteins have been named Kip, Klp, Krp or Kif, names that do not distinguish them by organism or function. To assist in differentiating such proteins, we have appended to the protein name the first letter of the genus and species name of the organism in which the protein was found. Thus, Klp1 from *Chlamydomonas reinhardtii* is referred to as CrKlp1. SpoPkl1 from *S. pombe* contains two letters of the species name to differentiate the *S. pombe* and *S. purpuratus* kinesin proteins. The kinesin heavy chains from different organisms are denoted KHC, prefaced by genus and species initials. The *D. melanogaster* kinesin protein encoded at polytene chromosome map position 3A7, previously designated klp3A, has been renamed 3A7Kin to distinguish it from the KLP3 protein reported by others (Stewart et al., 1991). The *S. purpuratus* heterotrimeric kinesin protein, consisting of SpKRP85, SpKRP95, and a third nonkinesin protein subunit, is denoted SpKinesinII. We recommend that these conventions of nomenclature be adopted in reports of new kinesin proteins.

Alignment

The protein sequence alignment used for the analysis presented here was based on the multiple alignment program, Tulla (version 2.1) (Subbiah and Harrison, 1989). Tulla optimizes sequence alignments by performing all possible pairwise comparisons, finds the most related pair, fine-tunes the alignment, and then repeats this process to add sequences sequentially to the most related pair. The sequence alignments obtained are intended to be those that best reflect the 3-dimensional structural alignment of the proteins. The final alignment, based on initial Tulla alignments of overlapping sets of 10 sequences, was adjusted and optimized manually. The final protein sequences were added by hand, in some cases aided by the Wisconsin GCG Program Bestfit. The alignment begins with DmKHC sequence A E D S I K V and ends with K N V V C V N, corresponding to amino acids 9-340. The sequence alignment, with the omission of sequences not yet entered into the databases, is available upon request.

Treebuilding

Trees were built using the distance matrix program, CLUSTAL V (Higgins et al., 1992), and the maximum parsimony program, PAUP (version 3.1.1) (Swofford, 1993). Robustness of the resulting topologies was estimated by 'bootstrap resampling' (Felsenstein, 1985), a procedure in which positions of the alignment are randomly sampled until an alignment equivalent in size to the original alignment is created. A tree is then built from the sampled alignment. The alignment sampling and treebuilding are repeated for each bootstrap trial. Topologies well supported by the data should be found in a large fraction of the bootstrapped trees.

CLUSTAL V builds trees by first calculating a distance between each pair of sequences in the multiple alignment, creating a distance matrix. Distance is defined as the per cent nonidentity between protein sequences after correction for multiple substitutions by the method of Kimura (1983). CLUSTAL V then uses the neighbor-joining method of Saitou and Nei (1987) to build a tree in which the distances measured along the branches are as close as possible to the distances in the distance matrix. The neighbor-joining method starts with a star phylogeny (one node with n arms radiating to the n sequences) and chooses a pair of sequences to be designated neighbors (sequences joined by a single node) based on the total length of the tree in its partially resolved state. The tree length is the sum of all branch lengths. The neighbor-joining process is continued iteratively until a network containing only bifurcating nodes is obtained. Positions with gaps were included in the analysis and treated in the default manner by CLUSTAL V.

The maximum parsimony program, PAUP, searches for trees that minimize the number of evolutionary steps needed to explain the differences in protein sequence. Trees were built using the PAUP heuristic search method with random stepwise addition and 'tree bisection-reconnection' (TBR) branch swapping. PAUP provides exact methods for finding optimal trees, but the large amount of computational time required to search all possible trees for large datasets makes these methods impractical for use in analysis of >10-12 proteins. Heuristic methods do not guarantee finding the most parsimonious tree, but have been found to be very effective when used in combination with TBR branch swapping and randomized sequence addition (Swofford, 1993). Sequence input was in alphabetical order by protein or species name, or grouped by subfamily with ungrouped proteins randomly ordered. 'Constraint' trees, that defined the five groups of proteins well-supported by bootstrap analysis, were used as starting trees in some searches. Bootstrap analysis was carried out using the heuristic search method with 'simple' stepwise addition (Farris, 1970). DmKHC was used as the reference taxon for the stepwise addition. In all PAUP analyses, equal weight was assigned to each character in the alignment and gaps were treated as missing data.

Topological structure within groups well-supported by bootstrap analysis was refined using an exact PAUP method. Subsets of the sequences that included the proteins in the group together with three proteins outside the group (defined as 'outgroup' taxa) were analyzed

using the PAUP 'branch and bound' search option (Stewart, 1993). The proteins used as outgroup taxa were as follows: DmNcd group (DmKlp61F, Dm3A7Kin, HsCENP-E), AnBimC group (ScKip2, SpKRP85, MmKif2), SpKRP85 group (DmNod, XlEg5, LcKin), MmKif2 group (CeUnc104, DmKHC, MmKif3), KHC group (AnBimC, ScSmy1, DmNcd).

Analysis of partial motor sequences

Analysis of a partial motor domain sequence with PAUP was carried out using as input 29 complete and a partial DmUnc104 sequence (S. A. Endow, unpublished).

RESULTS AND DISCUSSION

Kinesin phylogenetic trees

Two different methods of analysis were used to generate phylogenetic trees for 34 kinesin proteins from 15 organisms, representing nonmetazoans and metazoans from the protist, fungal, plant and animal kingdoms. The proteins that were analyzed are shown in Table 1. Amino acid sequences from only the motor domains of the proteins were analyzed. Trees were built using the distance matrix analysis package, CLUSTAL V (Higgins et al., 1992), and the maximum parsimony program, PAUP (Phylogenic Analysis Using Parsimony) (Swofford, 1993).

The trees of the kinesin family inferred from these analyses are shown in Fig. 1. Kinesin protein names are prefaced by genus and species initials, as described in Materials and Methods. Nodes or branchpoints on the trees may represent protein divergence (gene duplication events) or species divergence; interpretations must take this ambiguity into account. Numbers adjacent to nodes on the trees refer to the percentage

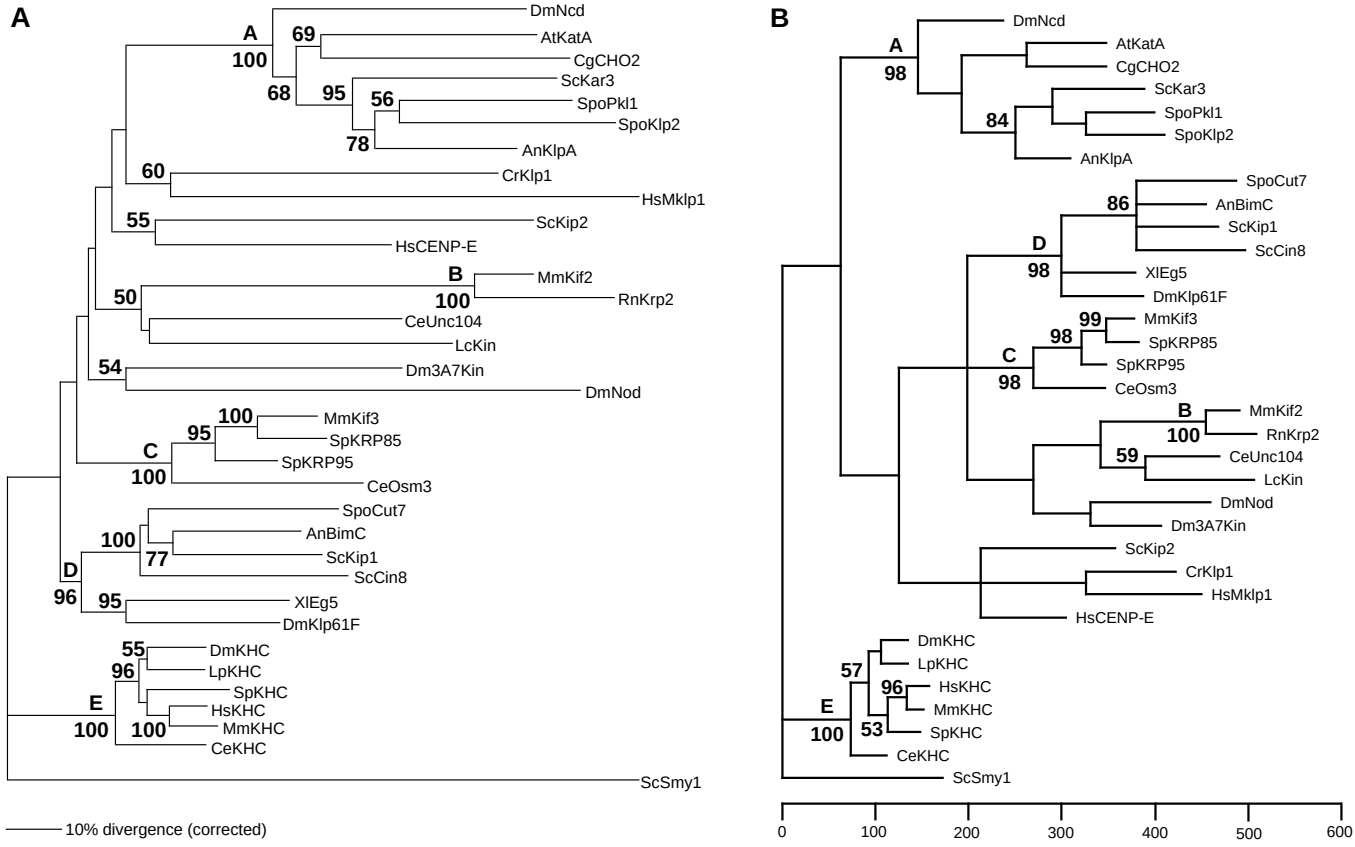

Fig. 1. Phylogenetic trees of kinesin motor domain sequences obtained by distance matrix and maximum parsimony analysis. Numbers adjacent to nodes indicate the percentage of bootstrap trials in which the proteins to the right of the node appeared as a group. Only nodes found in >50% of trials are shown. For purposes of comparison, the trees are rooted with ScSmy1 defined as the 'outgroup' kinesin protein. ScSmy1 is the most divergent of the kinesin proteins and is therefore likely to be the product of an early gene duplication, but is not known to be the ancestral kinesin. Nodes A-E mark the positions of five clearly defined groups of kinesins that correspond to similarly marked nodes in the two trees. (A) The distance matrix tree shown is a majority rule consensus tree resulting from 300 bootstrap replications with CLUSTAL V. Horizontal branch lengths are drawn to scale in units of percentage divergence after correction by the method of Kimura (1983). Vertical branch lengths have no meaning with respect to time or relatedness of sequences. The line at the bottom shows 10% divergence. (B) The maximum parsimony tree shown is a strict consensus of six optimal trees that resulted from 100 treebuilding trials performed using the PAUP heuristic method with random stepwise addition and TBR branch swapping. The same six trees were found in treebuilding trials of 100 trees each that were performed using the same alignment, but with input sequences reordered. Horizontal branch lengths are proportional to the relative number of amino acid changes needed to explain the differences in protein sequences. Vertical branch lengths have no meaning. The tree has a length of 4206 and a consistency index (CI), excluding uninformative characters, of 0.584. The scale at the bottom shows relative branch length.

Table 1. Kinesin proteins analysed and their properties

Organism	Protein	Motor position	Motor polarity	Subcellular location	References
C. reinhardtii	Klp1	N terminus	ND	Central pair C2 MT	1
S. cerevisiae	Kar3p	C terminus	Minus-end	Spindle, poles	2
	Cin8p	N terminus	ND	Spindle	3
	Kip1p	N terminus	ND	Spindle, poles	4
	Kip2p	Central	ND	ND	5
	Smy1p	N terminus	ND	Secretory vesicles?	6
S. pombe	cut7	N terminus	ND	Spindle, poles	7
	pkl1	C terminus	ND	Spindle, poles	8
	klp2	C terminus	ND	ND	9
A. nidulans	bimC	N terminus	ND	Spindle? Poles?	10
	klpA	C terminus	ND	Spindle?	11
A. thaliana	KatA	C terminus	ND	ND	12
L. chagasi	LcKin	N terminus	ND	Amastigotes	13
S. purpuratus	KHC	N terminus	Plus-end	Membranous vesicles	14
	KRP85	N terminus	Plus-end	ND	15
	KRP95	N terminus	Plus-end	ND	16
L. pealii	KHC	N terminus	Plus-end	Membrane-bounded vesicles	17
C. elegans	KHC	N terminus	(Plus-end)	(Vesicles, organelles)	18
	Osm3	N terminus	ND	Vesicles?	19
	Unc104	N terminus	ND	Synaptic vesicles?	20
D. melanogaster	KHC	N terminus	Plus-end	(Vesicles, organelles)	21
	Klp61F	N terminus	ND	Spindle?	22
	Ncd	C terminus	Minus-end	Spindle, centrosomes	23
	3A7Kin	N terminus	ND	Spindle, midbody	24
	Nod	N terminus	ND	ND	25
X. laevis	Eg5	N terminus	Plus-end	Spindle, poles	26
M. musculus	KHC	N terminus	(Plus-end)	Vesicles, organelles	27
	Kif2	Central	ND	Brain	28
	Kif3	N terminus	Plus-end	Membrane-bounded organelles	29
R. norvegicus	Krp2	?	ND	?	30
C. griseus	CHO2	C terminus	ND	Spindle, centrosomes	31
H. sapiens	KHC	N terminus	(Plus-end)	(Vesicles, organelles)	32
	CENP-E	N terminus	ND	Kinetochores, midzone	33
	MKLP1 (CHO1)	N terminus	Plus-end	Spindle midzone	34

ND, not determined; MT, microtubule.

References: 1 (Bernstein et al., 1994); 2 (Meluh and Rose, 1990; Endow et al., 1994b); 3 (Hoyt et al., 1992); 4 (Hoyt et al., 1992; Roof et al., 1992); 5 (Roof et al., 1992); 6 (Lillie and Brown, 1992, 1993); 7 (Hagan and Yanagida, 1990, 1992); 8 (Pidoux and Cande, 1993); 9 (Troxell and McIntosh, 1993); 10 (Enos and Morris, 1990); 11 (O'Connell et al., 1993); 12 (Mitsui et al., 1993); 13 (Burns et al., 1993); 14 (Wright et al., 1991; Henson et al., 1992); 15 (Cole et al., 1992, 1993); 16 (Cole et al., 1993); 17 (Kosik et al., 1990; Pfister et al., 1989); 18 (Sulston et al., 1992; Hall et al., 1991); 19 (Perkins et al., 1986; Shakir et al., 1993); 20 (Otsuka et al., 1991; Hall and Hedgecock, 1991); 21 (Yang et al., 1989, 1990; Saxton et al., 1991); 22 (Heck et al., 1993; Wilson and Fuller, 1993); 23 (Endow et al., 1990; McDonald and Goldstein, 1990; Walker et al., 1990; McDonald et al., 1990; Hatsumi and Endow, 1992; Endow et al., 1994a); 24 (B. Williams, M. Riedy, E. Williams and M. Goldberg, unpublished); 25 (Zhang et al., 1990); 26 (Le Guellec et al., 1991; Sawin et al., 1992a); 27 (K. Kato, unpublished) (Hirokawa et al., 1991; Aizawa et al., 1992); 28 (Aizawa et al., 1992); 29 (Aizawa et al., 1992; Kondo et al., 1993); 30 (Sperry and Brady, 1993); 31 (Dragas-Granoic et al., 1993); 32 (Navone et al., 1992); 33 (Yen et al., 1991, 1992); 34 (Nislow et al., 1992).

of treebuilding trials on bootstrapped datasets in which the group represented by the node was found. Bootstrap resampling provides a measure of the strength of the data supporting a given group. Nodes found in >90% of trials can be regarded as strongly supported, while nodes with low (<50%) bootstrap values should be regarded as indeterminate. Only bootstrap values of >50% are indicated. For purposes of comparison, both trees have been rooted along the branch of the most divergent kinesin protein, ScSmy1. The actual position of the 'root' or ancestral kinesin is not known and has not been determined by our analysis.

Fig. 1A shows a majority rule consensus of the trees built by CLUSTAL V in 300 bootstrap resamplings of the alignment. Branch lengths were corrected for multiple substitutions by the method of Kimura (1983). The branch length for ScSmy1 is much longer in the CLUSTAL V than in the PAUP tree because of the branch length correction. PAUP does not correct for multiple substitutions.

The maximum parsimony strict consensus tree shown in Fig. 1B resulted from a search using the PAUP heuristic method with random stepwise addition and TBR branch swapping. Six optimal trees were found in 100 treebuilding trials with different random stepwise addition orders. Two further searches of 100 trees each using the same alignment, but with the input sequences reordered, resulted in the same six most parsimonious trees. The differences among the six trees are

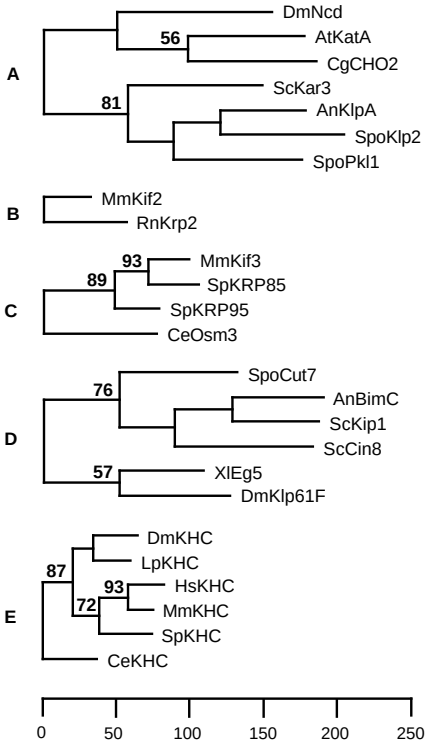

Fig. 2. Topologies of kinesin protein groups refined using a PAUP exact search method. The exact topologies were determined using subsets of the sequences that included the proteins in each group together with three proteins outside the group ('outgroup' taxa). The A-E subfamily designations correspond to the clearly defined groups in Fig. 1. The numbers adjacent to the nodes show the percentage of 100 branch and bound bootstrap trials in which the proteins to the right of the node appeared as a group. Bootstrap trials were performed using the same subsets of the data that were used to determine the exact subfamily topologies. The scale at the bottom shows relative branch lengths.

represented as polytomies with 3 or 4 branchpoints each in the strict consensus tree shown in Fig. 1B. The numbers indicate the percentage of bootstrap trials that support a given node in 300 bootstrap resamplings of the alignment performed using the PAUP heuristic search method with simple stepwise addition and TBR branch swapping.

Kinesin subfamilies

The overall topologies of the CLUSTAL V and PAUP consensus trees are very similar, with some differences in areas not well supported by bootstrap analysis. Both of the trees exhibit five well-supported groups or 'subfamilies' of kinesin proteins. These are labelled A-E and arbitrarily named the DmNcd, SpKRP85, AnBimC and KHC subfamilies, consisting of 4-7 proteins each, and the smaller MmKif2 subfamily, consisting of 2 proteins.

To obtain additional information about the relationships of proteins within each of the major groups, we used an exact PAUP method that guarantees to find the most parsimonious evolutionary relationship among the taxa. The five subfamilies were analyzed in subsets of the data that included the proteins in the group together with three proteins outside the group. The results of the exact searches, together with results of exact bootstrap trials, are shown in Fig. 2. The well-supported branchpoints within subfamilies found by the CLUSTAL V and PAUP bootstrap analyses (Fig. 1), and the subfamily topologies in the strict consensus optimal PAUP tree (Fig. 1B), are in good agreement with the exact subfamily topologies defined by PAUP.

Two of the subfamilies, DmNcd and AnBimC, contain proteins from organisms belonging to two or more kingdoms, implying that both types of kinesin existed in a common ancestor of the kingdoms. This suggests that the diversification of kinesin into multiple types occurred very early in eukaryotic evolution. The existence of multiple kinesin proteins in early eukaryotes and their retention in extremely divergent organisms suggest that they are present in many eukaryotes and probably function in basic cellular processes. Members of the KHC and SpKRP85 subfamilies are likely to be present at least throughout metazoans, based on the observation that both subfamilies already include proteins from organisms as divergent as *C. elegans* and mammals. A *Drosophila* protein represented by a partial sequence has also been reported to be a MmKif3 homologue (Aizawa et al., 1992), and therefore a member of the SpKRP85 subfamily, based on sequence similarity.

Each of the ungrouped proteins may represent a further kinesin protein subfamily. Most of these proteins differ as much from one another as the major groups differ from each other. This divergence suggests that the gene duplications leading to these proteins occurred early in eukaryotic evolution. Analysis of a partial kinesin motor sequence indicates that one of the ungrouped kinesin proteins, CeUnc104, is a representative of a larger subfamily that may exist in all metazoans: a partial *Drosophila* kinesin sequence (S. A. Endow, unpublished) analyzed using PAUP treebuilding, groups with CeUnc104. A partial murine kinesin sequence, MmKif1, has also been reported to be a CeUnc104 homolog based on sequence similarity (Aizawa et al., 1992). CeUnc104 groups with LcKin in <50% of distance matrix and ~60% of maximum parsimony bootstrap trials. *Leishmania* is a kinetoplastid protist thought to have diverged from the eukaryotic tree long before the fungal-metazoan divergence (Knoll, 1992). A subfamily that includes both LcKin and CeUnc104 would therefore suggest that CeUnc104 is an ancient form of kinesin. The bootstrap results do not permit strong conclusions, but the possibility that these proteins are related should be kept in mind as they are studied. This may also be true of HsCENP-E and ScKip2, Dm3A7Kin and DmNod, and CrKlp1 and HsMklp1, which are grouped in >50% of CLUSTAL V bootstrap trials but <50% of PAUP bootstrap trials. More kinesin protein sequences will be needed to improve distance measurements and provide points of calibration for evolutionary rates before specific predictions can be made with regard to these, as well as other potential kinesin subfamilies.

Other evidence for kinesin subfamily classifications

Evidence supporting the above subfamilies of kinesin proteins is based on analysis of the nonmotor protein sequences, overall structural characteristics, and mutant phenotypes of some of the proteins. Table 1 summarizes available information regarding the position of the conserved motor domain in the protein, the polarity of motor movement on microtubules, and the subcellular location of the proteins.

Table 2. Sequence comparisons of kinesin proteins

Sequences compared			% Identity	
			Motor region	Nonmotor region
DmKHC	and	LpKHC	81.1	53.8
		SpKHC	77.7	50.8
		HsKHC	77.1	53.7
		MmKHC	74.5	48.3
		CeKHC	74.0	40.5
SpKRP85	and	MmKif3	80.2	64.5
		SpKRP95	75.0	NA
		CeOsm3	59.5	NA
MmKif2	and	RnKrp2	71.6	NA
AnBimC	and	ScKip1	66.2	18.7
		SpoCut7	60.5	18.5
		ScCin8	57.7	20.8
		XlEg5	56.6	25.2
		DmKlp61F	51.9	21.4
DmNcd	and	AnKlpA	49.8	19.9
		SpoPkl1	48.6	15.8
		CgCHO2	47.3	21.8
		AtKatA	46.1	22.4
		ScKar3	46.0	17.6
		SpoKlp2	45.7	21.0
DmKHC	and	MmKif3	45.6	22.3
		AnBimC	45.2	17.7
		MmKif2	38.8	25.3
ScKar3	and	DmKHC	37.5	25.1
		AnBimC	36.3	21.8

Comparisons were made using the Wisconsin GCG Program Bestfit.
NA, not available.

In overall structure, the kinesin proteins resemble kinesin heavy chain, and consist of a motor domain, a 'stalk' region of predicted α-helical coiled coil, and a nonmotor 'tail' region (Yang et al., 1989; Chandra et al., 1993). DmNod, an apparent exception, is discussed below. Table 2 shows results of pairwise sequence comparisons of motor and nonmotor (stalk + tail) regions of kinesin subfamily proteins. Representative non-subfamily protein comparisons are shown at the bottom of the table. Proteins that fall into the KHC and SpKRP85 subfamilies show striking sequence similarity to other members of their group both in the motor region and the nonmotor region, which was not used in the phylogenetic analysis. Sixty to eighty per cent of amino acids in the motor region and ≥40% in the nonmotor (stalk + tail) region are identical in proteins that have been completely sequenced. The nonmotor region of RnKrp2 is not yet available for analysis to permit the comparison of nonmotor domains. Nonmotor regions of proteins in the DmNcd and AnBimC subfamilies are not significantly similar to one another (Table 2). However, other characteristics of members of these subfamilies support these groups of proteins.

A subfamily with C-terminal motor domains

A striking characteristic shared by the members of the DmNcd subfamily is overall structure: the mechanochemical domain of the proteins in this subfamily is at the C terminus of the protein rather than the N terminus, as in all of the other kinesin proteins analyzed. All proteins with C-terminal motor sequences that were analyzed fall into this group, suggesting that this arrangement of kinesin motor and nonmotor regions arose and was retained only once in the evolution of the kinesin proteins. The finding of this group indicates that sequence characteristics of the motor region distinguish kinesin proteins with C-terminal motors from those with N-terminal motors. Two divergent members of the DmNcd subfamily, DmNcd and ScKar3, exhibit a polarity of movement on microtubules opposite to that of KHC (Walker et al., 1990; McDonald et al., 1990; Endow et al., 1994b), suggesting that all members of the group are reversed in polarity. The identification of the molecular determinants of motor polarity is likely to be important in understanding the mechanism of function of the kinesin motors. Several proteins in the group, including DmNcd (Endow et al., 1994a), ScKar3 (Page et al., 1994), CgCHO2 (Dragas-Granoic et al., 1993) and SpoPkl1 (Pidoux and Cande, 1993), have been localized to centrosomes and spindle poles, as well as mitotic spindle fibers, suggesting a role in spindle pole function in mitosis. This would be consistent with the proposed minus-end polarity of proteins in this group, and is likely to differ from the role of the AnBimC proteins in mitosis, discussed below.

A phylogenetic relationship comparable to that of the kinesins with C-terminal motor domains does not appear to distinguish the kinesin proteins with centrally located mechanochemical domains. These proteins now include *S. cerevisiae* Kip2p and *M. musculus* Kif2. Since only a partial sequence is available for RnKrp2, it is not certain that the MmKif2 group will include only kinesin proteins with centrally located motors. ScKip2 falls outside the group, however, indicating that not all kinesin proteins with central motor domains are in the same group. The heptad repeat regions in both ScKip2 and MmKif2 are C terminal to the conserved motor domain. Thus, these proteins may more closely resemble 'conventional' KHC with added N-terminal sequences, rather than representing a new structural type of kinesin protein.

A subfamily of kinesin proteins required for mitotic spindle bipolarity

The nonmotor regions of the AnBimC group proteins do not show significant sequence similarity to one another (Table 2), however, the members of this subfamily are distinguished by their involvement in spindle pole separation, required for formation of a bipolar spindle. This proposed function is based on mutant phenotype (Enos and Morris, 1990; Hagan and Yanagida, 1990; Heck et al., 1993; Wilson and Fuller, 1993), genetic interactions (Hoyt et al., 1992; Roof et al., 1992; Saunders and Hoyt, 1992), and antibody localization experiments (Hagan and Yanagida, 1992; Hoyt et al., 1992; Roof et al., 1992; Sawin et al., 1992a). Immunolocalization and results of in vitro spindle assembly experiments are also consistent with a spindle pole function for XlEg5 (Sawin et al., 1992a). Although XlEg5 has been interpreted to function in spindle pole organization, the findings are not inconsistent with a role in spindle pole separation. Further, the disruption of spindles formed in oocyte extracts upon addition of antibody (Sawin et al., 1992a) parallels the spindle collapse observed in vivo following temperature-induced inactivation of ScCin8 in ScKip1 null cells (Saunders and Hoyt, 1992). In both cases, the collapsed spindles suggest that the kinesin proteins are required to maintain spindle bipolarity. Not all of the kinesin spindle motors are members of the AnBimC subfamily; members of the DmNcd group also function in the spindle and localize to spindle poles, but do not appear to be required for

mitotic spindle pole separation. Two further, ungrouped proteins, HsMklp1 and HsCENP-E, are found in the spindle midbody in anaphase and telophase, and are thought to perform functions in the spindle that differ from those of the AnBimC subfamily.

A heterotrimeric kinesin

Based on evidence from biochemical and electron microscopic studies, the native form of 'conventional' kinesin is a tetramer consisting of two heavy chains and two light chains (Kuznetsov et al., 1988). The predicted coiled-coil region of KHC and the kinesin-related proteins is thought to mediate assembly of the proteins into homodimers. This has been found for KHC, based on electron microscopy (Scholey et al., 1989; Hirokawa et al., 1989), and for DmNcd, based on electron microscopy, gel filtration and sedimentation analysis (Chandra et al., 1993), and is expected to be true of other kinesin proteins. Recent work indicates that two members of the SpKRP85 subfamily, SpKRP85 and SpKRP95, can be isolated as a complex with a third, nonkinesin protein (Cole et al., 1993). The heterotrimeric complex is thought to function as a kinesin motor protein, denoted SpKinesinII, consisting of three different polypeptide subunits. Such an association of two distinct motor-bearing kinesin proteins in a heterotrimeric complex has not been observed previously. The complex might represent a way of regulating or coordinating two different, but related microtubule motor activities. Other members of the SpKRP85 subfamily, which now includes CeOsm3, MmKif3 and a *Drosophila* protein, may also function as heterotrimeric kinesin proteins.

A kinesin I motor protein?

As noted above, most of the kinesin proteins contain a region of predicted α-helical coiled coil that lies adjacent to the mechanochemical domain and is thought to mediate dimerization of the proteins. DmNod is the only kinesin protein reported to date that lacks a region of heptad repeats of hydrophobic amino acids (Zhang et al., 1990). Analysis of DmNod using COILS, a computer program that predicts coiled coil structure (Lupas et al., 1991), resulted in overall probabilities of <0.05 (T. G. Oas and S. A. Endow, unpublished). The low predicted probabilities, together with the high proline content (~6%) of the nonmotor region, support the absence of coiled-coil structure. DmNod may therefore represent a class of kinesins that, like the myosin I proteins, function as monomers. A schematic representation of the different forms of kinesin proteins for which evidence now exists is shown in Fig. 3.

Relationships within groups

In addition to identifying major groups, analysis of sequence phylogenies can assist in determining whether proteins within a group are products of genes that diverged as a result of a speciation event (orthologues) or as a result of a gene duplication (paralogues). Definition of proteins as orthologues or paralogues can be useful for predicting function. Orthologous proteins often have similar functions, while paralogous proteins have often diverged in function. Incongruity between protein and species phylogenies indicates paralogy. Recent molecular phylogenies indicate that insects and nematodes diverged after the separation of the line leading to chordates (Lake, 1990; Christen et al., 1991). Both the clustal V and the

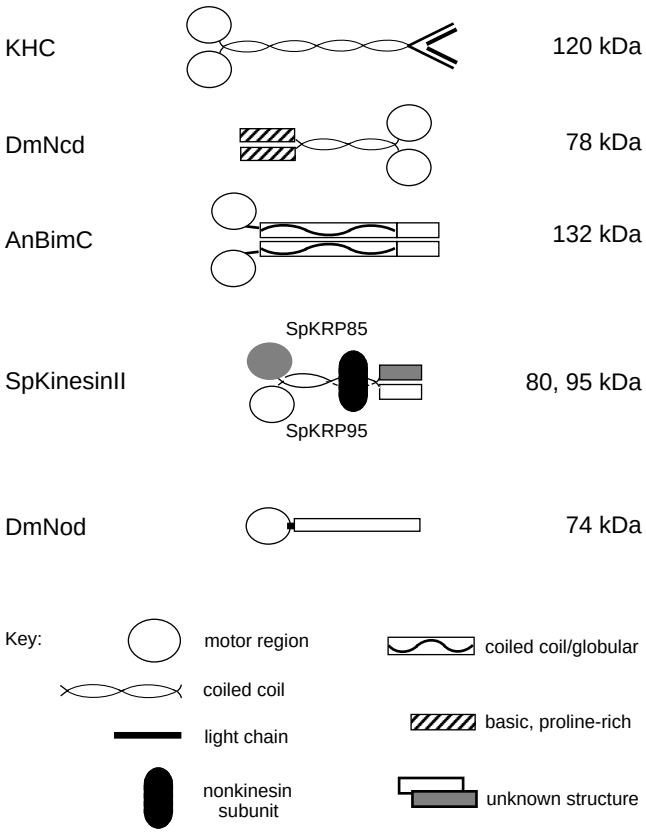

Fig. 3. Schematic depiction of kinesin protein structures based on biochemical and electron microscopy studies (KHC and DmNcd) (Kuznetsov et al., 1988; Scholey et al., 1989; Hirokawa et al., 1989; Chandra et al., 1993) or predicted from analysis of amino acid sequences (all others). The KHC molecular mass shown is for DmKHC. SpKinesinII is a heterotrimeric complex of SpKRP85, SpKRP95 and a third, nonkinesin protein of 115 kDa (Cole et al., 1993). The nature of the association of the three proteins is not yet known - the coiled coil structure of the two motor protein subunits is proposed from the presence of a region of heptad repeats of hydrophobic amino acids in SpKRP85. The coiled-coil/globular domain in AnBimC and the absence of coiled-coil structure in DmNod are predicted by analysis of the protein sequences using COILS, a coiled-coil prediction program (Lupas et al., 1991) (T. G. Oas and S. A. Endow, unpublished). Lengths are not to scale.

PAUP analysis indicate that CeKHC diverged from the KHC proteins *before* the separation of lines leading to *Drosophila* and vertebrate KHC. This suggests that CeKHC is a member of a KHC subclass that separated from the other present-day KHC proteins prior to the divergence of insects and chordates. The CeKHC nonmotor region is less similar to other KHC nonmotor regions (~40% identity) than other KHC nonmotor regions are to one another (≥~50% identity) (Table 2 and data not shown). These results provide evidence that CeKHC has diverged in function from the other KHC proteins and suggest that some metazoans may have more than one type of KHC.

The analysis indicates that diversification of other kinesin protein groups is ancient as well. Two *S. purpuratus* proteins (SpKRP85, SpKRP95) are present in the SpKRP85 group, two *S. cerevisiae* proteins (ScCin8, ScKip1) are in the AnBimC group, and two *S. pombe* proteins (SpoPkl1, SpoKlp2) are in

the DmNcd group. The bootstrap values and exact subfamily topologies indicate that SpKRP85 and SpKRP95 arose as a gene duplication prior to the divergence of echinoderms and chordates. Other deuterostomes are therefore likely to have proteins corresponding to both SpKRP85 and SpKRP95. The bootstrap data and exact subfamily topologies are consistent with the interpretation that MmKif3 and SpKRP85 are orthologues, or closely related paralogues, that carry out the same function in different organisms. This is also likely to be true of AnBimC and ScKip1. The bootstrap values for the DmNcd subfamily are too low to permit strong conclusions; however, the separation of DmNcd from CgCHO2, and the pairing of the mammalian CgCHO2 protein with the plant protein, AtKatA, suggest the existence of at least two types of C-terminal motor kinesin proteins prior to the divergence of plants and animals. Plants and animals might therefore be expected to have two, and possibly more, kinesin proteins belonging to the DmNcd group. Identification of further subclasses and prediction of the organisms in which they exist will require the discovery of more kinesin proteins and analysis of the protein sequences.

Structure-function correlations

The relationships inferred from phylogenetic analysis have a potential use in prediction of biochemical characteristics, or functions, of otherwise uncharacterized proteins. Protein sequences conserved within classes can allow structure/function correlations to be made once class-specific characteristics become known. One result of our analysis is the observation of a group containing all C-terminal kinesin motor proteins. Two divergent members of the group, DmNcd and ScKar3, exhibit a polarity of translocation on microtubules opposite to that of KHC (Walker et al., 1990; McDonald et al., 1990; Endow et al., 1994b), a property that is expected to map to the motor domain (Chandra et al., 1993; Stewart et al., 1993). The finding that both proteins are minus-end microtubule motors suggests that other members of the group are also reversed in polarity compared with KHC. If so, class-specific characteristics in protein sequence may allow identification of motor regions that specify polarity of motor translocation. Other biochemical characteristics of kinesin motor proteins, such as velocity and properties of the ATP hydrolysis cycle, may also be amenable to this type of analysis.

How many kinesins per cellular function?

How many kinesin proteins are involved in a given cellular function? Although further work is required to answer this question with certainty, it is now clear that at least two major groups of kinesin proteins, the DmNcd and AnBimC groups, contain kinesin motor proteins that function in the meiotic and/or mitotic spindle (Sawin and Endow, 1993). Based on the present study, members of these two subfamilies are predicted to be present in all eukaryotes (possibly excluding the earliest diverging lineages). Thus, spindle function, and consequently chromosome distribution in meiosis and mitosis in most eukaryotes is probably dependent on the activity of at least two kinesin motor proteins and may require two further, presently ungrouped proteins, HsCENP-E and HsMklp1, for spindle midzone function. Dm3A7Kin may also be involved in spindle function, but its role has not yet been established.

Multiple kinesin proteins have also been implicated in organelle and vesicle transport. The kinesin proteins involved in transport that have been identified so far include members of the KHC group, CeUnc104 (Otsuka et al., 1991) and CeOsm3 (Perkins et al., 1986). These proteins represent three different subfamilies, based on our present analysis, implying that a minimum of three distinct kinesins function in intracellular organelle/vesicle transport. The finding that *C. elegans* contains members of all three subfamilies suggests that multiple kinesins are needed to carry out this function in other metazoans as well. ScSmy1 (Lillie and Brown, 1992) has recently been implicated in actin-based vesicle transport (Lillie and Brown, 1993) and may represent a fourth class of kinesin proteins involved in organelle/vesicle transport.

Conclusions

Phylogenetic analysis of kinesin motor domain sequences identifies five clearly defined groups that are consistent with nonmotor sequence similarity, overall structure of the proteins, or cellular function. Present evidence suggests that members of at least two of these groups will be found in most, or all, eukaryotes. As many as 14 different types of kinesins are represented by the five groups or single proteins on the trees. Among the ungrouped kinesin proteins, *Drosophila* Nod may represent a class that functions as monomers, rather than dimers. The acquisition of more sequences will reveal further groups and improve distance measurements, allowing more specific predictions to be made with regard to the existence and distribution of types within the groups.

We thank M. Bernstein and J. Rosenbaum, A. Sperry and S. Brady, A. Pidoux and Z. Cande, C. Troxell and R. McIntosh, and R. Kuriyama, for generously providing unpublished sequences for this analysis, D. Kenan for help with the Tulla alignment program, J. Scholey for helpful comments, and T.G. Oas for analysis of DmNod, AnBimC and Dm3A7Kin with COILS. R. Vilgalys provided help in getting started with the PAUP analysis and comments on the manuscript, and M. Uyenoyama made valuable suggestions for improvement of the manuscript. H.V.G. gratefully acknowledges the support, advice and encouragement of J. A. Spudich. The work was supported by grants from the NIH to J. A. Spudich, and from the NIH and ACS to S.A.E.

REFERENCES

Aizawa, H., Sekine, Y., Takemura, R., Zhang, Z., Nangaku, M. and Hirokawa, N. (1992). Kinesin family in murine central nervous system. *J. Cell Biol.* **119**, 1287-1296.

Allen, R. D., Metuzals, J., Tasaki, I., Brady, S. T. and Gilbert, S. P. (1982). Fast axonal transport in squid giant axon. *Science* **218**, 1127-1129.

Allen, R. D., Weiss, D. G., Hayden, J. H., Brown, D. T., Fujiwake, H. and Simpson, M. (1985). Gliding movement of and bidirectional transport along single native microtubules from squid axoplasm: evidence for an active role of microtubules in cytoplasmic transport. *J. Cell Biol.* **100**, 1736-1752.

Bernstein, M., Beech, P. L., Katz, S. G. and Rosenbaum, J. L. (1994). A new kinesin-like protein, Klp1, localized to a single microtubule in the *Chlamydomonas* flagellum. *J. Cell Biol.* (in press).

Bloom, G. S. (1992). Motor proteins for cytoplasmic microtubules. *Curr. Opin. Cell Biol.* **4**, 66-73.

Brady, S. T. (1985). A novel brain ATPase with properties expected for the fast axonal transport motor. *Nature* **317**, 73-75.

Burns, J. M., Jr, Shreffler, W. G., Benson, D. R., Ghalib, H. W., Badaro, R. and Reed, S. G. (1993). Molecular characterization of a kinesin-related antigen of *Leishmania chagasi* that detects specific antibody in African and American visceral leishmaniasis. *Proc. Nat Acad. Sci. USA* **90**, 775-779.

Chandra, R., Salmon, E. D., Erickson, H. P., Lockhart, A. and Endow, S. A.

(1993). Structural and functional domains of the *Drosophila* ncd microtubule motor protein. *J. Biol. Chem.* **268**, 9005-9013.

Christen, R., Ratto, A., Baroin, A., Perasso, R., Grell, K. G. and Adoutte, A. (1991). An analysis of the origin of metazoans, using comparisons of partial sequences of the 28S RNA, reveals an early emergence of triploblasts. *EMBO J.* **10**, 499-503.

Cole, D. G., Cande, W. Z., Baskin, R. J., Skoufias, D. A., Hogan, C. J. and Scholey, J. M. (1992). Isolation of a sea urchin egg kinesin-related protein using peptide antibodies. *J. Cell Sci.* **101**, 291-301.

Cole, D. G., Chinn, S. W., Wedaman, K. P., Hall, K., Vuong, T. and Scholey, J. M. (1993). Novel heterotrimeric kinesin-related protein purified from sea urchin eggs. *Nature* **366**, 268-270.

Dragas-Granoic, S., Kato, T., Khodjakov, A., Kofron, M. and Kuriyama, R. (1993). Identification of a novel kinesin-like protein with C-terminal motor domain from mitotic spindles. *Mol. Biol. Cell* **4**, 243a.

Endow, S. A., Henikoff, S. and Soler-Niedziela, L. (1990). Mediation of meiotic and early mitotic chromosome segregation in *Drosophila* by a protein related to kinesin. *Nature* **345**, 81-83.

Endow, S. A. and Hatsumi, M. (1991). A multimember kinesin gene family in *Drosophila*. *Proc. Nat. Acad. Sci. USA* **88**, 4424-4427.

Endow, S. A. and Titus, M. A. (1992). Genetic approaches to molecular motors. *Annu. Rev. Cell Biol.* **8**, 29-66.

Endow, S. A., Chandra, R., Komma, D. J., Yamamoto, A. H. and Salmon, E. D. (1994a). Mutants of the *Drosophila* ncd microtubule motor protein cause centrosomal and spindle pole defects in mitosis. *J. Cell Sci.* **107**, 859-867.

Endow, S. A., Kang, S. J., Satterwhite, L. L., Rose, M. D., Skeen, V. P. and Salmon, E. D. (1994b). The yeast Kar3 is a minus-end microtubule motor protein that destabilizes microtubules preferentially at the minus ends. *EMBO J.* (In press.)

Enos, A. P. and Morris, N. R. (1990). Mutation of a gene that encodes a kinesin-like protein blocks nuclear division in A. nidulans. *Cell* **60**, 1019-1027.

Farris, J. S. (1970). Methods for computing Wagner trees. *Syst. Zool.* **19**, 83-92.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783-791.

Hagan, I. and Yanagida, M. (1990). Novel potential mitotic motor protein encoded by the fission yeast *cut7*$^{+}$ gene. *Nature* **347**, 563-566.

Hagan, I. and Yanagida, M. (1992). Kinesin-related cut7 protein associates with mitotic and meiotic spindles in fission yeast. *Nature* **356**, 74-76.

Hall, D. H. and Hedgecock, E. M. (1991). Kinesin-related gene *unc-104* is required for axonal transport of synaptic vesicles in C. elegans. *Cell* **65**, 837-847.

Hall, D. H., Plenefisch, J. and Hedgecock, E. M. (1991). Ultrastructural abnormalities of kinesin mutant unc-116. *J. Cell Biol.* **115**, 389a.

Hatsumi, M. and Endow, S. A. (1992). The *Drosophila* ncd microtubule motor protein is spindle-associated in meiotic and mitotic cells. *J. Cell Sci.* **103**, 1013-1020.

Heck, M. M. S., Pereira, A., Pesavento, P., Yannoni, Y., Spradling, A. C. and Goldstein, L. S. B. (1993). The kinesin-like protein KLP61F is essential for mitosis in *Drosophila*. *J. Cell Biol.* **123**, 665-679.

Henikoff, S. (1993). Sequence analysis by electronic mail server. *Trends Biochem. Sci.* **18**, 267-268.

Henson, J. H., Nesbitt, D., Wright, B. D. and Scholey, J. M. (1992). Immunolocalization of kinesin in sea urchin coelomocytes. Association of kinesin with intracellular organelles. *J. Cell Sci.* **103**, 309-320.

Higgins, D. G., Bleasby, A. J. and Fuchs, R. (1992). CLUSTAL V: improved software for multiple sequence alignment. *CABIOS* **8**, 189-191.

Hirokawa, N., Pfister, K. K., Yorifuji, H., Wagner, M. C., Brady, S. T. and Bloom, G. S. (1989). Submolecular domains of bovine bain kinesin identified by electron microscopy and monoclonal antibody decoration. *Cell* **56**, 867-878.

Hirokawa, N., Sato-Yoshitake, R., Kobayashi, N., Pfister, K. K., Bloom, G. S. and Brady, S. T. (1991). Kinesin associates with anterogradely transported membranous organelles in vivo. *J. Cell Biol.* **114**, 295-302.

Hoyt, M. A., He, L., Loo, K. K. and Saunders, W. S. (1992). Two *Saccharomyces cerevisiae* kinesin-related gene products required for mitotic spindle assembly. *J. Cell Biol.* **118**, 109-120.

Kimura, M. 1983. *The Neutral Theory Of Evolution.* Cambridge University Press, Cambridge, UK.

Knoll, A. H. (1992). The early evolution of eukaryotes: a geological perspective. *Science* **256**, 622-627.

Kondo, S., Sato-Yoshitake, R., Aizawa, H., Nakata, T., Matsuura, Y. and Hirokawa, N. (1993). KIF3 is a new microtubule based fast anterograde motor in the nerve axon. *Mol. Biol. Cell* **4**, 49a.

Kosik, K. S., Orecchio, L. D., Schnapp, B., Inouye, H. and Neve, R. L. (1990). The primary structure and analysis of the squid kinesin heavy chain. *J. Biol. Chem.* **265**, 3278-3283.

Kuznetsov, S. A., Vaisberg, E. A., Shanina, N. A., Magretova, N. N., Chernyak, V. Y. and Gelfand, V. I. (1988). The quaternary structure of bovine brain kinesin. *EMBO J.* **7**, 353-356.

Lake, J. A. (1990). Origin of the metazoa. *Proc. Nat. Acad. Sci. USA* **87**, 763-766.

Le Guellec, R., Paris, J., Couturier, A., Roghi, C. and Philippe, M. (1991). Cloning by differential screening of a *Xenopus* cDNA that encodes a kinesin-related protein. *Mol. Cell. Biol.* **11**, 3395-3398.

Lillie, S. H. and Brown, S. S. (1992). Suppression of a myosin defect by a kinesin-related gene. *Nature* **356**, 358-361.

Lillie, S. H. and Brown, S. S. (1993). Characterization of Myo2p, a type V myosin, and of Smy1p, a kinesin-related protein that can suppess a defect in *MYO2*. *Mol. Biol. Cell* **4**, 227a.

Lupas, A., Van Dyke, M. and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164.

McDonald, H. B. and Goldstein, L. S. B. (1990). Identification and characterization of a gene encoding a kinesin-like protein in Drosophila. *Cell* **61**, 991-1000.

McDonald, H. B., Stewart, R. J. and Goldstein, L. S. B. (1990). The kinesin-like *ncd* protein of Drosophila is a minus end-directed microtubule motor. *Cell* **63**, 1159-1165.

Meluh, P. B. and Rose, M. D. (1990). *KAR3*, a kinesin-related gene required for yeast nuclear fusion. *Cell* **60**, 1029-1041.

Mitsui, H., Yamaguchi-Shinozaki, K., Shinozaki, K., Nishikawa, K. and Takahashi, H. (1993). Identification of a gene family (*kat*) encoding kinesin-like proteins in *Arabidopsis thaliana* and the characterization of secondary structure of KatA. *Mol. Gen. Genet.* **238**, 362-368.

Navone, F., Niclas, J., Hom-Booher, N., Sparks, L., Bernstein, H. D., McCaffrey, G. and Vale, R. D. (1992). Cloning and expression of a human kinesin heavy chain gene: interaction of the COOH-terminal domain with cytoplasmic microtubules in transfected CV-1 cells. *J. Cell Biol.* **117**, 1263-1275.

Nislow, C., Lombillo, V. A., Kuriyama, R. and McIntosh, J. R. (1992). A plus-end-directed motor enzyme that moves antiparallel microtubules *in vitro* localizes to the interzone of mitotic spindles. *Nature* **359**, 543-547.

O'Connell, M. J., Meluh, P. B., Rose, M. D. and Morris, N. R. (1993). Suppression of the *bimC*4 mitotic spindle defect by deletion of *klpA*, a gene encoding a KAR3-related kinesin-like protein in *Aspergillus nidulans*. *J. Cell Biol.* **120**, 153-162.

Otsuka, A. J., Jeyaprakash, A., García-Añoveros, J., Tang, L. Z., Fisk, G., Hartshorne, T., Franco, R. and Born, T. (1991). The C. elegans *unc-104* gene encodes a putative kinesin heavy chain-like protein. *Neuron* **6**, 113-122.

Page, B. D., Satterwhite, L. L., Rose, M. D. and Snyder, M. (1994). Localization of the Kar3 kinesin heavy chain related protein requires the Cik1 interacting protein. *J. Cell Biol.* **124**, 507-519.

Perkins, L. A., Hedgecock, E. M., Thomson, J. N. and Culotti, J. G. (1986). Mutant sensory cilia in the nematode *Caenorhabditis elegans*. *Dev. Biol.* **117**, 456-487.

Pfister, K. K., Wagner, M. C., Stenoien, D. L., Brady, S. T. and Bloom, G. S. (1989). Monoclonal antibodies to kinesin heavy and light chains stain vesicle-like structures, but not microtubules in cultured cells. *J. Cell Biol.* **108**, 1453-1463.

Pidoux, A. L. and Cande, W. Z. (1993). pkl1: a kinesin-like protein from *Schizosaccharomyces pombe* which localises to the mitotic spindle. *Mol. Biol. Cell* **4**, 243a.

Roof, D. M., Meluh, P. B. and Rose, M. D. (1992). Kinesin-related proteins required for assembly of the mitotic spindle. *J. Cell Biol.* **118**, 95-108.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406-425.

Saunders, W. S. and Hoyt, M. A. (1992). Kinesin-related proteins required for structural integrity of the mitotic spindle. *Cell* **70**, 451-458.

Sawin, K. E., LeGuellec, K., Philippe, M. and Mitchison, T. J. (1992a). Mitotic spindle organization by a plus-end-directed microtubule motor. *Nature* **359**, 540-543.

Sawin, K. E., Mitchison, T. J. and Wordeman, L. G. (1992b). Evidence for kinesin-related proteins in the mitotic apparatus using peptide antibodies. *J. Cell Sci.* **101**, 303-313.

Sawin, K. E. and Endow, S. A. (1993). Meiosis, mitosis and microtubule motors. *BioEssays* **15**, 399-407.

Saxton, W. M., Hicks, J., Goldstein, L. S. B. and Raff, E. C. (1991). Kinesin heavy chain is essential for viability and neuromuscular functions in Drosophila, but mutants show no defects in mitosis. *Cell* **64**, 1093-1102.

Scholey, J. M., Heuser, J., Yang, J. T. and Goldstein, L. S. B. (1989). Identification of globular mechanochemical heads of kinesin. *Nature* **338**, 355-357.

Shakir, M. A., Fukushige, T., Yasuda, H., Miwa, J. and Siddiqui, S. S. (1993). *C. elegans osm-3* gene mediating osmotic avoidance behaviour encodes a kinesin-like protein. *Mol. Neurosci.* **4**, 891-894.

Sperry, A. O. and Brady, S. T. (1993). Identification and characterization of new members of the kinesin superfamily. *Mol. Biol. Cell* **4**, 228a.

Stewart, C.-B. (1993). The powers and pitfalls of parsimony. *Nature* **361**, 603-607.

Stewart, R. J., Pesavento, P. A., Woerpel, D. N. and Goldstein, L. S. B. (1991). Identification and partial characterization of six new members of the kinesin superfamily in *Drosophila*. *Proc. Nat. Acad. Sci. USA* **88**, 8470-8474.

Stewart, R. J., Thaler, J. P. and Goldstein, L. S. B. (1993). Direction of microtubule movement is an intrinsic property of the motor domains of kinesin heavy chain and *Drosophila* ncd protein. *Proc. Nat. Acad. Sci. USA* **90**, 5209-5213.

Subbiah, S. and Harrison, S. C. (1989). A method for multiple sequence alignment with gaps. *J. Mol. Biol.* **209**, 539-548.

Sulston, J., et al. (1992). The *C. elegans* genome sequencing project: a beginning. *Nature* **356**, 37-41.

Swofford, D. L. (1993). *PAUP: Phylogenetic Analysis Using Parsimony*. Computer program distributed by the Illinois Natural History Survey, Champaign, IL 61820.

Troxell, C. L. and McIntosh, J. R. (1993). Cloning of genes encoding kinesin-like proteins from Schizosaccharomyces pombe. *Mol. Biol. Cell* **4**, 243a.

Vale, R. D., Reese, T. S. and Sheetz, M. P. (1985a). Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell* **42**, 39-50.

Vale, R. D., Schnapp, B. J., Mitchison, T., Steuer, E., Reese, T. S. and Sheetz, M. P. (1985b). Different axoplasmic proteins generate movement in opposite directions along microtubules in vitro. *Cell* **43**, 623-632.

Walker, R. A., Salmon, E. D. and Endow, S. A. (1990). The *Drosophila claret* segregation protein is a minus-end directed motor molecule. *Nature* **347**, 780-782.

Wilson, P. G. and Fuller, M. T. (1993). Centrosome uncoupling is mediated by *urcn*, a bimC homolog in *Drosophila*. *Mol. Biol. Cell* **4**, 228a.

Wright, B. D., Henson, J. H., Wedaman, K. P., Willy, P. J., Morand, J. N. and Scholey, J. M. (1991). Subcellular localization and sequence of sea urchin kinesin heavy chain: evidence for its association with membranes in the mitotic apparatus and interphase cytoplasm. *J. Cell Biol.* **113**, 817-833.

Yang, J. T., Laymon, R. A. and Goldstein, L. S. B. (1989). A three-domain structure of kinesin heavy chain revealed by DNA sequence and microtubule binding analyses. *Cell* **56**, 879-889.

Yang, J. T., Saxton, W. M., Stewart, R. J., Raff, E. C. and Goldstein, L. S. B. (1990). Evidence that the head of kinesin is sufficient for force generation and motility in vitro. *Science* **249**, 42-47.

Yen, T. J., Compton, D. A., Wise, D., Zinkowski, R. P., Brinkley, B. R., Earnshaw, W. C. and Cleveland, D. W. (1991). CENP-E, a novel human centromere-associated protein required for progression from metaphase to anaphase. *EMBO J.* **10**, 1245-1254.

Yen, T. J., Li, G., Schaar, B. T., Szilak, I. and Cleveland, D. W. (1992). CENP-E is a putative kinetochore motor that accumulates just before mitosis. *Nature* **359**, 536-539.

Zhang, P., Knowles, B. A., Goldstein, L. S. B. and Hawley, R. S. (1990). A kinesin-like protein required for distributive chromosome segregation in Drosophila. *Cell* **62**, 1053-1062.

(Received 24 January 1994 - Accepted 17 March 1994)

Note added in proof

A new kinesin protein, DSK1, recently identified in the diatom, *Cylindrotheca fusiformis*, by M. Foss, H. Wein, L. Wordeman and W. Z. Cande (in preparation) groups with MmKif2 and RnKrp2 in 99% of PAUP bootstrap trials. This suggests that the MmKif2 subfamily of kinesin proteins, like the DmNcd and AnBimC subfamilies, is widespread among eukaryotes. CfDSK1 localizes to mitotic spindles of the diatom (Foss, M., Wein, H. and Cande, W. Z. (1993). Identifying kinesins and other proteins in diatom mitotic spindles. *Mol. Biol. Cell* **4**, 243a), implicating members of the MmKif2 subfamily in mitotic spindle function.