Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 10114–10118, November 1991
Cell Biology

Molecular genetics of kinesin light chains: Generation of isoforms by alternative splicing

(fast axonal transport/organelle motility/molecular motors)

JANET L. CYR*, K. KEVIN PFISTER*†, GEORGE S. BLOOM*, CLIVE A. SLAUGHTER‡, AND SCOTT T. BRADY*§

*Department of Cell Biology and Neuroscience and ‡Department of Biochemistry and Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX 75235

Communicated by A. J. Hudspeth, July 24, 1991 (received for review June 18, 1991)

ABSTRACT **Movement of membrane-bounded organelles to intracellular destinations requires properly oriented microtubules and force-generating enzymes, such as the microtubule-stimulated ATPase kinesin. Kinesin is a heterotetramer with two heavy chain (≈124-kDa) and two light chain (≈64-kDa) subunits. Kinesin heavy chains contain both ATP- and microtubule-binding domains and are capable of force generation *in vitro*. Functions of the light chains are undetermined, although evidence suggests they interact with membrane surfaces. We have used molecular genetic approaches to dissect the kinesin light chain structure. Three distinct kinesin light chain cDNAs were cloned and sequenced from rat brain, and they were found to result from alternative splicing of a single gene. Polypeptides encoded by these cDNAs are identical except for their carboxyl ends. Synthesis of multiple light chains, differing from one another in primary structure, could provide a means of generating multiple, functionally specialized forms of the kinesin holoenzyme.**

Movement of membrane-bounded organelles (MBOs) from the cell body, down the length of the axon, and finally to the synaptic terminals is critical to neuronal function. This intracellular movement, termed fast axonal transport, requires properly oriented, linear arrays of microtubules (MTs) and the motor protein complex kinesin (1–10). Although originally characterized from neuronal tissues, kinesin has been localized to a variety of cell types (10) and is thought to play a role in MT-based motility in all cells.

Kinesin is a rod-shaped structure, ≈80 nm in length (9, 11–13), consisting of two heavy chain and two light chain subunits (14, 15). Genetic and electron microscopic (EM) analyses indicate that the heavy chains interact in parallel such that two amino-terminal globular domains containing MT- and ATP-binding sites (11–13, 15–19) are at one end and the carboxyl termini are at the opposite end. The two ends are connected by an α-helical, coiled-coil shaft which facilitates heavy chain dimerization (9, 11–13, 20). Less information is available about light chain architecture. Light chains are localized to the heavy chain carboxyl termini, forming a fan-like structure thought to interact with MBOs (9). At the cellular level, monoclonal antibodies against light and heavy chains have been localized to structures consistent with MBOs (10). In addition, immuno-EM studies localize kinesin on isolated mitochondria and synaptic vesicles (21).

Purification of bovine brain kinesin yields multiple forms of the heavy and light chains on both one- and two-dimensional gel electrophoresis (22). The functional significance of these variants has not yet been determined. To understand both light chain structure and the basis for their diversity, we have isolated a number of rat brain cDNAs encoding the kinesin light chains. Sequence analysis of these clones reveals at least three distinct mRNA species, which yield slightly different polypeptides.¶

MATERIALS AND METHODS

cDNA Clone Isolation and Sequencing. A λgt11 rat brain cDNA library (23) was screened with two monoclonal antibodies to kinesin light chains (10), using standard techniques (24, 25). Two partial clones immunoreactive with the L2 antibody were obtained. cDNA inserts were subcloned in Bluescript plasmids (Stratagene) and utilized for double-stranded dideoxy sequencing (26) with Sequenase (United States Biochemical). The smaller insert was labeled by random priming (Boehringer Mannheim) and used to probe two additional rat brain cDNA libraries (pHG327 and pGEM4). Sixteen additional clones were obtained. All clones isolated were sequenced at the open reading frame (ORF) 3′ end. Several isolates for each isoform contained the same sized inserts. Totals of three, five, and four independent clones for light chains A, B, and C, respectively, were obtained from screening three independent libraries; therefore the sequence differences do not represent incompletely processed mRNAs. Full-length clones for light chains A and C were recovered from the pHG327 library, while the light chain B clone was isolated from the pGEM4 library. The three full-length clones were sequenced in their entirety on both strands. Sequence manipulations utilized MICROGENIE (Beckman) and University of Wisconsin Genetics Computer Group PEPTIDESTRUCTURE and PLOTSTRUCTURE programs (27).

Southern Blot Analysis. Rat genomic DNA was isolated by standard techniques (25) and cut with *Bam*HI, *Eco*RI, or *Xba* I. DNA digests were loaded in triplicate, electrophoresed on a 0.7% agarose gel, and transferred to Nytran (Schleicher & Schuell). Blots were rinsed in 2× SSC (1× SSC = 0.15 M NaCl/0.015 M sodium citrate, pH 7.0), baked, and prewashed at 65°C in 5× SSC/0.5% SDS/1 mM EDTA, pH 8.0, for ≥15 min. Membranes were cut for use with three separate probes, prehybridized ≥1 hr in hybridization buffer (28) with denatured herring sperm DNA and yeast tRNA each at 250 μg/ml. The 5′ probe, extending from nucleotide position 345 to position 456, and 3′ probe, from position 2100 to position 2250, were generated by polymerase chain reaction (PCR) with appropriate primers and 1.5 ng of light chain A cDNA as template. PCR was performed with an [α-^{32}P]dCTP (Amersham) mixture: 100 μCi of 800 Ci/mmol plus 40 μCi of 3000 Ci/mmol (1 Ci = 37 GBq) (29). Products were resolved on

---

Abbreviations: MBOs, membrane-bounded organelles; MT, microtubule; ORF, open reading frame; EM, electron microscopic.

[†] Present address: Department of Anatomy and Cell Biology, University of Virginia Health Science Center, Charlottesville, VA 22908.

[§] To whom reprint requests should be addressed.

[¶] The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M75146, M75147, and M75148).

nondenaturing 12% polyacrylamide gels, stained, excised, and eluted. Light chain A *Pst* I/*Eco*RI fragment was isolated on a 0.7% low-melting-point agarose (FMC) gel and 25 ng was random primed with 50 μCi (3000 Ci/mmol) of $[\alpha\text{-}^{32}\text{P}]$dCTP. Hybridization was overnight at 65°C and final stringency washes were in 1× SSC/0.1% SDS at 65°C for 15 min.

RESULTS

Isolation of Light Chain cDNA Clones. By using monoclonal antibodies against kinesin light chains (10), two partial light chain clones were isolated from a λgt11 rat brain cDNA library (23). Both clones were immunoreactive with monoclonal antibody L2 (10). Isolates were sequenced and clone identity was confirmed by the presence of light chain tryptic peptide sequences within the ORF (Fig. 1). These tryptic peptides were isolated from purified bovine brain kinesin light chains and sequenced by automated Edman degradation. One of the clones contained 452 bp of coding sequence and 488 bp of 3′ untranslated sequences but lacked the poly(A) tail. The second clone was 309 bp in length, most of which was coding sequence. This observation indicated that full-length clones might be difficult to obtain due to *Eco*RI sites within the cDNA, thus resulting in incomplete clones during cDNA subcloning in the λgt11 *Eco*RI site. To obtain full-length clones, two additional rat brain cDNA libraries were employed that did not use a restriction digest step in construction (30). These libraries, pHG327 and pGEM4, were screened with probes generated from the smaller λgt11 isolate. A total of ≈150,000 colonies were screened and 16 additional clones were isolated, 7 of which contained inserts of sufficient size to encode a full-length ORF.

DNA and Protein Sequence Analysis. Initial sequence information from the isolated clones indicated that heterogeneity was present at the carboxyl terminus of the coding region. Thus this region of the 16 newly isolated clones was sequenced. Light chain cDNAs fell into three different classes (light chains A, B, and C) based upon the predicted carboxyl-terminal amino acid sequences (Fig. 2). Light chain B contains a 50-bp insertion at the light chain A ORF carboxyl terminus resulting in a 10 amino acid addition to the encoded polypeptide and an added 20 bp of 3′ untranslated sequences. Light chain C has a 77-bp insertion at the same location that encodes an additional 19 amino acids, 10 of which are found in light chain B. These three light chain variants may not represent the full complement of transcripts that can be produced. However, any additional transcripts must be in low abundance in mammalian brain, because they were not detected in the screening of a total of ≈200,000 cDNA clones and the subsequent partial sequencing of 18 independent isolates obtained from three different libraries.

Sequence analysis of full-length clones for each of the light chains revealed identical nucleotide sequences for all three species except for insertions present in light chains B (at 1771 bp) and C (at 1744 bp) and a single base pair change at position 523 in which the guanine of light chains B and C has been changed to an adenine in light chain A. This single base change is likely the result of an incorporation error during cDNA synthesis or plasmid replication. Guanine is believed to be the correct base at position 523. The 3′ untranslated

```
   1 GGGTCGGGGTGAGGATGCTGCGGGGCGGCTGCGGAGGCGTCGCTTGCTGCTGAGGCGGCTGGGCTCTGCGCGGCCTCGGCCCGT
  85 GACAGCGCGACGAGCGCGGCGCTTCCGATCGGACGAATGCATGACAACATGTCCACAATGGTGTACATGAAGGAAGAGAAGCTG
                                                 MetHisAspAsnMetSerThrMetValTyrMetLysGluGluLysLeu 16
 163 GAGAAGCTTACGCAGGATGAAATCATTTCTAAGACCAAGCAGGTGATCCAGGGGCTGGAGGCCCTGAAGAATGAGCACAACTCC
     GluLysLeuThrGlnAspGluIleIleSerLysThrLysGlnValIleGlnGlyLeuGluAlaLeuLysAsnGluHisAsnSer 44
 253 ATCCTGCAGAGTCTGCTGGAGACGCTGAAGTGCTTGAAGAAGGACGATGAGAGCAACCTGGTGGAGGAGAAGTCCAGCATGATC
     IleLeuGlnSerLeuLeuGluThrLeuLysCysLeuLysLysAspAspGluSerAsnLeuValGluGluLysSerSerMetIle 72
 337 CGCAAGTCCTTGGAGATGCTGGAGCTCGGCCTGAGCGAGGCGCAGGTGATGATGGCCCTGTCTAATCACCTGAATGCCGTGGAG
     ArgLysSerLeuGluMetLeuGluLeuGlyLeuSerGluAlaGlnValMetMetAlaLeuSerAsnHisLeuAsnAlaValGlu 100
 421 TCCGAGAAGCAGAAGCTCCGAGCTCAGGTTCGACGTCTGTGCCAGGAGAACCAGTGGCTGCGGGATGAACTGGCCAACACACAG
     SerGluLysGlnLysLeuArgAlaGlnValArgArgLeuCysGlnGluAsnGlnTrpLeuArgAspGluLeuAlaAsnThrGln 128
 505 CAGAAGTTACAGAAGAGCGAGCAGTCCGTGGCTCAGCTGGAGGAGGAGAAGAAGCACCTGGAGTTCATGAACCAGCTGAAGAAA
     GlnLysLeuGlnLysSerGluGlnSerValAlaGlnLeuGluGluGluLysLysHisLeuGluPheMetAsnGlnLeuLysLys 156
 589 TACGACGACGACATCTCTCCCTCGGAAGACAAAGACTCTGATTCTTCCAAAGAGCCGTTGGATGACCTCTTCCCGAATGATGAG
     TyrAspAspAspIleSerProSerGluAspLysAspSerAspSerSerLysGluProLeuAspAspLeuPheProAsnAspGlu 184
 673 GATGACCCAGGACAAGGAATCCAGCAGCAACACAGCAGTGCTGCGGCCGCCGCCCAGCAGGGCGGCTACGAGATCCCTGCCCGG
     AspAspProGlyGlnGlyIleGlnGlnGlnHisSerSerAlaAlaAlaAlaAlaGlnGlnGlyGlyTyrGluIleProAlaArg 212
 757 CTGCGCACGCTCCACAACCTGGTGATCCAGTATGCCTCGCAGGGGCGTTACGAGGTGGCGGTGCCACTCTGCAAGCAGGCCCTG
     LeuArgThrLeuHisAsnLeuValIleGlnTyrAlaSerGlnGlyArgTyrGluValAlaValProLeuCysLysGlnAlaLeu 240
 841 GAGGACCTGGAGAAGACTTCCGGCCACGACCACCCTGATGTGGCCACCATGCTCAACATCCTGGCCCTGGTGTACAGGGATCAG
     GluAspLeuGluLysThrSerGlyHisAspHisProAspValAlaThrMetLeuAsnIleLeuAlaLeuValTyrArgAspGln 268
 925 AACAAGTATAAAGACGCAGCAAACCTCCTGAACGATGCCCTGGCTATCCGTGAGAAAACCCTGGGCCGAGATCACCCTGCGGTG
     AsnLysTyrLysAspAlaAlaAsnLeuLeuAsnAspAlaLeuAlaIleArgGluLysThrLeuGlyArgAspHisProAlaVal 296
1009 GCTGCGACCCTCAACAACCTAGCAGTGCTGTACGGTAAGCGTGGCAAGTACAAAGAGGCCGAGCCGCTGTGCAAACGGGCCCTG
     AlaAlaThrLeuAsnAsnLeuAlaValLeuTyrGlyLysArgGlyLysTyrLysGluAlaGluProLeuCysLysArgAlaLeu 324
1093 GAGATCAGGGAGAAGGTTCTGGGAAAGGATCACCCTGATGTTGCTAAGCAACTAAATAATCTGGCCTTGCTGTGCCAGAACCAG
     GluIleArgGluLysValLeuGlyLysAspHisProAspValAlaLysGlnLeuAsnAsnLeuAlaLeuLeuCysGlnAsnGln 352
1177 GGCAAGTACGAGGAGGTGGAGTATTACTACCAGAGGGCCCTGGAGATCTACCAGACGAAGCTCGGACCCGATGATCCCAACGTG
     GlyLysTyrGluGluValGluTyrTyrTyrGlnArgAlaLeuGluIleTyrGlnThrLysLeuGlyProAspAspProAsnVal 380
1261 GCCAAGACGAAGAATAACCTGGCTTCCTGTTATCTGAAACAAGGGAAGTTCAAACAAGCAGAAACATTGTACAAGGAGATTCTC
     AlaLysThrLysAsnAsnLeuAlaSerCysTyrLeuLysGlnGlyLysPheLysGlnAlaGluThrLeuTyrLysGluIleLeu 408
1345 ACCCGCGCACATGAACGGGAGTTTGGATCTGTGGATGATGAAAACAAGCCCATATGGATGCACGCGGAGGAGAGAGAGGAGTGC
     ThrArgAlaHisGluArgGluPheGlySerValAspAspGluAsnLysProIleTrpMetHisAlaGluGluArgGluGluCys 436
1429 AAAGGAAAGCAGAAGGACGGGTCGTCTTTTGGAGAGTATGGCGGCTGGTATAAGGCCTGCAAAGTGGACAGTCCCACCGTCACA
     LysGlyLysGlnLysAspGlySerSerPheGlyGluTyrGlyGlyTrpTyrLysAlaCysLysValAspSerProThrValThr 464
1513 ACCACCTTGAAAAACCTGGGAGCACTTTACCGCCGCCAGGGGAAGTTTGAAGCTGCGGAGACACTGGAGGAAGCTGCCCTGAGG
     ThrThrLeuLysAsnLeuGlyAlaLeuTyrArgArgGlnGlyLysPheGluAlaAlaGluThrLeuGluGluAlaAlaLeuArg 492
1597 TCACGTAAGCAGGGTCTTGACAATGTTCACAAACAGAGAGTGGCTGAAGTGCTAAATGACCCTGAGAACGTGGAGAAGCGGAGG
     SerArgLysGlnGlyLeuAspAsnValHisLysGlnArgValAlaGluValLeuAsnAspProGluAsnValGluLysArgArg 520
1681 AGCCGGGAGAGCCTCAATGTGGACGTGGTCAAGTACGAGAGTGGCCCTGACGGAGGGGAGGAAGTGAGTATGAGCGTAGAGTGG
     SerArgGluSerLeuAsnValAspValValLysTyrGluSerGlyProAspGlyGlyGluGlu.................. 541 A.
     SerArgGluSerLeuAsnValAspValValLysTyrGluSerGlyProAspGlyGlyGluGlu.................. 541 B.
     SerArgGluSerLeuAsnValAspValValLysTyrGluSerGlyProAspGlyGlyGluGluValSerMetSerValGluTrp 548 C.
1765 AATGGATGAGAAAGATGAAGCTCGGGCTGGTTAAATGACTTGCTCAGCGTCCATGGCCTAGCCACCCGGTGCCTCCCCTCTGC
     ....................................................AlaEnd                           542 A.
     ......MetArgLysMetLysLeuGlyLeuValLysEnd                                              551 B.
     AsnGlyMetArgLysMetLysLeuGlyLeuValLysEnd                                              560 C.
1849 TCCTGCATGACGGGCAGCGCCTCCTCCGGCCCCTCCAGTGCTGTCCGCTGTCTAGCAGCCCTAGGAGCTTGTCAGCCGCACCTC
1933 TGGGACTGGCCCCTCCTTTCGGTGCTGTCGTCCTATGGGGTCTTGATTTCTGGATACATGTAGCTTTGCCAGATATGTACTTAG
2017 AAATATCAACCGTGTTAATAAGCCCATTAATGTGTAACACACAAGTTTAACAGTTAAGAGAGAGTAGCTGTTCATCTTGAGGAC
2101 AGGCGTTAGTGTCACACAGGTTTATTACAGTGGCATGATAGTGACCTGTGCAGTCACTGAGCTGTAGGAGTGAGGCATGTGCTG
2185 CACGTCCCTGTCGTACTTAGAGACCTGGCAGTGCCCTGCTCCGGGCCTCCCCAGGACACACTGACGTAGATTTGCTGTGTAGGT
2269 CTTGGTGTGGAAATTTGAATTCTGCTTTTCTCTACTAATCTGAAGTGCTCCGTGATGTGTGTATGTGTCTTTCTAACTTCTAAT
                                                                                     ^^^
2353 AAACTCACTCCAGCTGAAAAAAAAAAAAAAAAAAAn
     ^^^
```

FIG. 1. DNA sequence and predicted amino acid sequence (three-letter notation) of kinesin light chains A, B, and C. Single underlined amino acids delineate region of homology with the bovine kinesin light chain tryptic peptides. Two conservative changes were seen between the bovine peptides and predicted rat sequences, resulting in 97% amino acid identity. These differences were at residues 444 and 491, where the bovine peptides contained Thr and Met, respectively. ORFs beginning at the first in-frame Met (multiple Met residues at the ORF beginning are in bold) extend for 1623, 1653, and 1680 bases for light chains A, B, and C, respectively. Sequence gaps are denoted by dotted lines. Beginning at nucleotide 1681, the predicted amino acid sequences for each of the light chains are shown. Relevant in-frame stop codons are doubly underlined. Light chain B clone terminated at nucleotide position 1930 and poly(A) tail lengths for light chains A and C consisted of 27 or 43 base pairs (bp), respectively. Polyadenylylation signal has carets below the line. No evidence for homology between the encoded polypeptides and any proteins contained in the European Molecular Biology Laboratory or GenBank data bases was found as of October 1990 and July 1991, respectively.

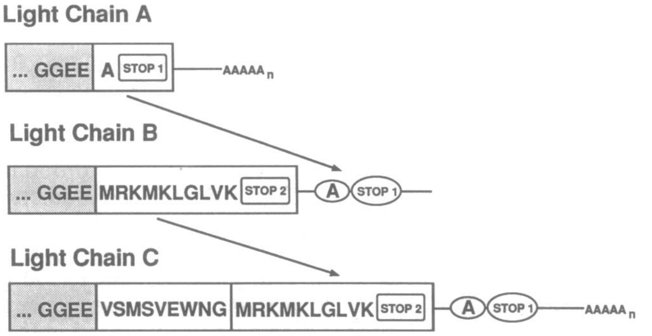
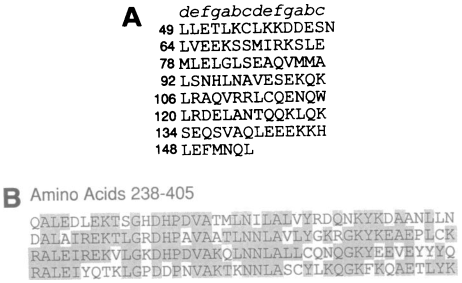


FIG. 2. Diagrammatic representation of kinesin light chain heterogeneity at the carboxyl terminus of the coding region. ORFs (denoted by boxes) for light chains A, B, and C are identical except for a 50-bp (light chain B) or 77-bp (light chain C) insertion at the carboxyl terminus. These insertions result in three light chain isoforms that are identical for the first 541 amino acids (shaded boxes) but differ in their carboxyl-terminal amino acids. The ORF of light chain A terminated with an alanine. In light chain B, the alanine and subsequent stop codon (stop 1) seen in light chain A are contained within the 3′ untranslated sequences and an additional 50 bp has been alternatively spliced in the coding region. This insertion encodes for 10 amino acids, a new in-frame stop codon (stop 2), and 20 bp of additional 3′ untranslated sequences. Light chain C is a light chain B variant in which an additional 9 amino acids have been inserted 5′ to the light chain B addition, resulting in a 19-amino acid tail, 10 of which are found in light chain B. Untranslated sequences are denoted by the solid lines and ellipses. The light chain B clone is missing part of the 3′ untranslated sequences and the poly(A) tail. The portion of the 3′ untranslated region that is present is identical to light chains A and C.

sequences and poly(A) tails varied in length. In addition, the clones had differing amounts of 5′ untranslated sequences, but all three contained the in-frame stop codon designating the ORF 5′ end.

The first methionine codon encountered downstream from the stop codon in the 5′ untranslated region signifies the ORF beginning. However, the first 10 amino acids of the predicted protein include three additional methionines (Fig. 1). This high number of methionines opens the possibility that alternative translation initiation sites are used, thereby generating further diversity at the amino terminus. We are as yet unable to establish which methionine(s) initiate translation because peptide sequencing revealed a blocked amino terminus (data not shown). However, analyses of other messages containing multiple potential initiation sites (31) indicate that the 5′-most methionine is usually the major start site of translation.

Assuming translation begins at the first in-frame methionine, the cDNAs encode polypeptides of 542, 551, and 560 amino acids with deduced molecular weights of 61,640, 62,754 and 63,744, respectively (Fig. 1). Predicted molecular weights are consistent with observed migration of biochemically purified bovine brain kinesin light chains on SDS/polyacrylamide gels (15, 22). Further analysis of amino acid composition and sequences suggests that these polypeptides are largely hydrophilic with no major hydrophobic domains. The first 163 amino acids of these polypeptides include a high percentage of helix-forming residues and no prolines, consistent with an α-helical amino terminus. Within this region 15 heptad repeats [$(defgabc)_n$] extend from residues 49 to 154 (Fig. 3*A*). Heptad repeats in which core residues (*a* and *d*) are enriched in hydrophobic amino acids (Table 1) enable optimal packing of two α-helices into a coiled-coil structure (32). In addition, the distribution of apolar and charged amino acids within light chain heptad repeats corresponds well to known α-helical coiled-coil regions of various proteins including kinesin heavy chain and myosin (13, 33) (Table 1).


A *defgabcdefgabc*
49 LLETLKCLKKDDESN
64 LVEEKSSMIRKSLE
78 MLELGLSEAQVMMA
92 LSNHLNAVESEKQK
106 LRAQVRRLCQENQW
120 LRDELANTQQKLQK
134 SEQSVAQLEEEKKH
148 LEFMNQL

B Amino Acids 238-405
QALEDLEKTSGHDHPDVATMLNILALVYRDQNKYKDAANLLN
DALAIREKTLGRDHPAVAATLNNLAVLYGKRGKYKEAEPLCK
RALEIREKVLGKDHPDVAKQLNNLALLCQNQGKYEEVEYYYQ
RALEIYQTKLGPDDPNVAKTKNNLASCYLKQGKFKQAETLYK

FIG. 3. Primary sequence motifs of the kinesin light chains. (*A*) Region of heptad repeat sequence found from Leu-49 to Leu-154. Fifteen heptad repeats are present in the light chain primary sequence. The "skip" residue at position 63 is not uncommon in segments of heptad repeats and has been found in several proteins, including the nematode myosin rod (32). (*B*) Alignment of four imperfect amino acid repeats located in kinesin light chains A, B, and C. Repeats begin at Gln-238 and extend through Lys-405. Each repeat consists of 42 amino acids, which center at several conserved residues. Identical amino acids are shaded.

The secondary structure of the remainder of the polypeptide is more difficult to predict. Close examination of the primary sequence in this region reveals four imperfect tandem repeats of 42 amino acids each, extending from position 238 to 405 (Fig. 3*B*). This region may contain several helices interrupted by turns or bends. Although these repeats contain groupings of hydrophobic and charged residues, they lack the strong hydrophobic core needed to produce a tightly packed globular domain. Thus, this area may be more diffuse, consistent with appearances of the fan-like domain of kinesin (9). The significance of these tandem repeats is unknown, but their length and the high degree of amino acid identity indicate that this region is important for light chain function, possibly through interactions with a membrane surface or proteinaceous receptor.

Southern Blot Analysis. Nucleotide identity throughout untranslated regions and ORFs (except as noted in Fig. 1) suggested that all three light chain cDNAs were products of a single gene. Southern blots of rat genomic DNA were analyzed by using a *Pst* I/*Eco*RI fragment of light chain A cDNA as a probe (Fig. 4). Although this *Pst* I/*Eco*RI fragment hybridized to multiple bands, small PCR-generated 5′

Table 1. Percentage of hydrophobic and charged residues in each of the three positions of the heptad repeats in kinesin and α-helical coiled-coil proteins

	% of residues			
Protein	Core (*a*, *d*)	Inner (*e*, *g*)	Outer (*b*, *c*, *f*)	Total
		Hydrophobic		
Kinesin light chain	18	7	6	30
Kinesin heavy chain	17	5	4	26
Myosin	16	3	2	21
		Positively charged		
Kinesin light chain	3	6	8	17
Kinesin heavy chain	5	5	7	17
Myosin	4	4	12	20
		Negatively charged		
Kinesin light chain	2	9	7	18
Kinesin heavy chain	2	9	12	22
Myosin	1	8	13	23

Comparison is with the known α-helical coiled-coil regions of kinesin heavy chain (371 amino acids) and myosin (1064 amino acids). Hydrophobic residues are defined as F, I, L, M, W, V, and Y. Positively charged residues are taken as K, H, and R. Negatively charged residues are D and E. Data for the known coiled-coil regions are from refs. 13 and 33.

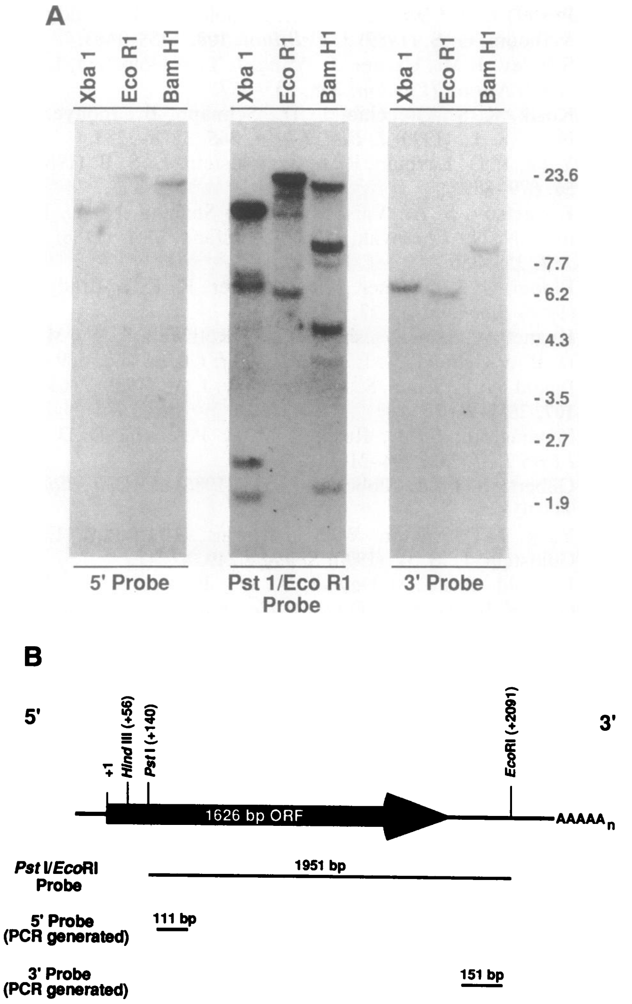


FIG. 4. Determination of the number of kinesin light chain genes by Southern blot analyses. (*A*) Southern analysis of rat genomic DNA digests (15 μg per lane) were analyzed with three probes generated from light chain A which are also present in light chains B and C. Sizes of molecular weight standards are given on the right in kilobases. (*B*) Diagrammatic representation of the probes generated from light chain A cDNA and utilized for Southern analysis. The probes used were (*i*) a 111-bp 5′ probe, (*ii*) a 1951-bp *Pst* I/*Eco*RI probe consisting of 92% ORF and 536 bp of 3′ untranslated sequences, or (*iii*) a 151-bp 3′ probe.

and 3′ probes (Fig. 4*B*) each hybridized to single bands in a lane of digested DNA (Fig. 4*A*), as expected for a single gene.

Multiple Transcript Generation by Alternative Splicing. Several lines of evidence suggest that the multiple kinesin light chain transcripts are generated by differential splicing of a single gene. The strongest evidence is the nucleotide identity of the three cDNAs except in areas of insertions (Fig. 1). This identity includes the stop codon used to terminate the light chain A ORF (stop 1 in Fig. 2). If multiple genes code these isoforms there would be no evolutionary pressure to conserve this stop codon in the 3′ untranslated region of light chains B and C. In addition, Southern blots detect only one gene coding for these mRNA species. If multiple genes code these mRNAs one would expect the small 5′ and 3′ probes to hybridize to more than one band per lane (Fig. 4*A*). Since neither *Bam*HI nor *Xba* I cuts within the cDNA sequence and introns of multiple genes are unlikely to be conserved, the hybridization patterns of the 5′ and 3′ probes indicate that one gene encodes these mRNAs. From these data we conclude that the observed light chain mRNA heterogeneity results from alternative splicing of a single gene.

DISCUSSION

Although a substantial amount of information has been elucidated concerning the kinesin heavy chains, little is known about the light chains. Purification of kinesin from bovine brain yields at least five electrophoretic species for the light chains (22). The biochemical basis for this diversity is not completely understood, but the presence of multiple transcripts in combination with posttranslational modifications (refs. 39 and 40; R. G. Elluru, G.S.B., and S.T.B., unpublished data) may account for this heterogeneity.

The cloning and sequencing of kinesin light chain cDNAs provide the most substantial information about the light chains to date. Structurally, the light chains contain an α-helical amino terminus containing 15 heptad repeats. Due to physical constraints, the number of heptad repeats in any non-α-helical structure is limited to two or three (32); therefore this region of the light chains is predicted to form an α-helical coiled-coil structure. The involvement of this domain in the interaction of two polypeptides is clear but the identity of the subunit to which each light chain binds is less obvious. Although this structural motif may be involved in light chain dimerization, to date no evidence supports this hypothesis. This area may facilitate heavy chain/light chain interactions. Biochemically, separation of light chains from heavy chains has required the use of denaturing agents. This high-affinity association does not result from interchain disulfide bonds (34). An α-helical coiled-coil structure involving a heavy chain and light chain may account for this tight interaction between the two constituents. Localization of light chain antibody epitopes to the heavy chain "fan-shaped" tail (9) does not preclude the possibility that the heavy chain/light chain interaction site is within the holoenzyme stalk. We propose that the light chain amino terminus interacts with heavy chain stalk domains and the light chain carboxyl terminus contributes to the kinesin fan-like tail.

Four imperfect tandem repeats extending from residues 238 to 405 form another repeat motif. These repeats consist of 42 amino acids each containing 11 conserved amino acids and 27 residues found in at least three of the four repeats. This high degree of identity between repeats suggests a role in maintaining secondary and tertiary structure of the polypeptide and that this motif is necessary for proper function.

In addition to providing structural information, isolation of light chain cDNA clones revealed light chain diversity in the form of multiple mRNAs encoding slightly different polypeptides. Several plausible hypotheses can be proposed to explain the existence of multiple light chain isoforms: (*i*) one or more isoforms may be cell-type specific, (*ii*) individual light chain variants may target kinesin to specific organelle classes, or (*iii*) light chain diversity may underlie kinesin mechanochemical diversity. The first hypothesis has many precedents; specific expression of isoforms in brain is common (clathrin light chains, tubulin, etc.) (35, 36). Multiple light chain mRNAs could result from expression of one or more isoforms by specific brain cell types. If these isoforms prove to be widespread, light chain isoforms are more likely to be involved in interactions of kinesin with specific cellular components or regulation of kinesin function.

Second, the potential function of light chains in targeting and/or binding kinesin to specific organelles is consistent with EM and immunofluorescence data. EM analyses demonstrate that light chains are at the end of the holoenzyme that binds organelles (9) and that kinesin is associated with a variety of MBOs (21). Immunofluorescence microscopy of cultured cells with two monoclonal antibodies to light chains yields a punctate pattern consistent with MBOs. The individual staining patterns have subtle differences, suggesting that isoforms may be associated with different MBOs (10). This possibility is attractive because secondary structure predictions of the 10-amino acid tail of light chains B and C (Figs. 1 and 2) suggest an amphipathic helix (data not shown). The role of amphipathic helices for proper localization of mitochondrial proteins is well established (37). Light chains are clearly not imported into mitochondria, but this structural motif could be important for mitochondrial targeting. Generation of isoform-specific antibodies against carboxyl termini of light chains A, B, and C may unambiguously determine organelle distribution for different light chains.

Finally, light chain isoforms may generate complexity of kinesin function. Precedents can be seen in myosin-based contractility, where myosin light chains are capable of modulating myosin function (for review, see ref. 38). Potential regulation of kinesin by light chains may be more complex than observed in myosin. Electrophoresis of bovine kinesin reveals multiple isoforms of heavy chains as well as light chains (22). Heavy chain heterogeneity may involve post-translational modifications (39, 40) (R. G. Elluru, G.S.B., and S.T.B., unpublished data) but there may also be primary sequence differences. The existence of numerous light chain isoforms, and perhaps heavy chain isoforms as well, creates the potential for multiple combinatorials with different properties including MBO specificity, regulation of MT-stimulated ATPase activities, or directionality of movement.

In determining kinesin function, roles played by both heavy and light chain subunits must be ascertained. A careful assessment of light chain function has not been possible by standard biochemical techniques, because free light chains have not been purified without detergents. Isolation of mammalian kinesin light chain cDNAs will enable us to dissect functions of these subunits and to obtain a clearer understanding of their role in movement of MBOs.

We are grateful to David Coleman for the λgt11 library and Gabriel Travis for pGH327 and pGEM4 libraries. We are indebted to L. Gierasch, J. Goldstein, L. Miranda, J. Norton, D. Russell, J. Sambrook, and G. Travis for their thoughtful comments and suggestions. We thank H. Clendening and B. Cantu for DNA sequencing and C. R. Moomaw and K. Orth for amino acid analysis. This work was supported by National Institutes of Health Grants NS23320 to S.T.B. and NS23868 to S.T.B. and G.S.B., National Institutes of Health Training Fellowship GM08203 to J.L.C., and Robert Welch Foundation Grant I-1077 to G.S.B. and S.T.B.

1. Lasek, R. J. & Brady, S. T. (1985) *Nature (London)* **316**, 645–647.
2. Brady, S. T. (1985) *Nature (London)* **317**, 73–75.
3. Vale, R. D., Reese, T. S. & Sheetz, M. P. (1985) *Cell* **42**, 39–50.
4. Brady, S. T., Pfister, K. K. & Bloom, G. S. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1061–1065.
5. Gilbert, S. P., Allen, R. D. & Sloboda, R. D. (1985) *Nature (London)* **315**, 245–247.
6. Hollenbeck, P. J. (1989) *J. Cell Biol.* **108**, 2335–2342.
7. Leopold, P. L., Synder, R., Bloom, G. S. & Brady, S. T. (1990) *Cell Motil. Cytoskeleton* **15**, 210–219.
8. Paschal, B. M. & Vallee, R. B. (1987) *Nature (London)* **330**, 181–183.
9. Hirokawa, N., Pfister, K. K., Yorifuji, H., Wagner, M. C., Brady, S. T. & Bloom, G. S. (1989) *Cell* **56**, 867–878.
10. Pfister, K. K., Wagner, M. C., Stenoien, D. L., Brady, S. T. & Bloom, G. S. (1989) *J. Cell Biol.* **108**, 1453–1463.
11. Scholey, J. M., Heuser, J., Yang, J. T. & Goldstein, L. S. B. (1989) *Nature (London)* **338**, 355–357.
12. Kosik, K. S., Orecchio, L. D., Schnapp, B., Inouye, H. & Neve, R. L. (1990) *J. Biol. Chem.* **265**, 3278–3283.
13. Yang, J. T., Laymon, R. A. & Goldstein, L. S. B. (1989) *Cell* **56**, 879–889.
14. Kuznetsov, S. A., Vaisberg, E. A., Shanina, N. A., Magretova, N. N., Chernyak, V. Y. & Gelfand, V. I. (1988) *EMBO J.* **7**, 353–356.
15. Bloom, G. S., Wagner, M. C., Pfister, K. K. & Brady, S. T. (1988) *Biochemistry* **27**, 3409–3416.
16. Kuznetsov, S. A., Vaisberg, Y. A., Rothwell, S. W., Murphy, D. B. & Gelfand, V. I. (1989) *J. Biol. Chem.* **264**, 589–595.
17. Ingold, A. L., Cohn, S. A. & Scholey, J. M. (1988) *J. Cell Biol.* **107**, 2657–2667.
18. Penningroth, S. M., Rose, P. M. & Peterson, D. D. (1987) *FEBS Lett.* **222**, 204–210.
19. Gilbert, S. P. & Sloboda, R. D. (1986) *J. Cell Biol.* **103**, 947–956.
20. Yang, J. T., Saxton, W. M., Stewart, R. J., Raff, E. C. & Goldstein, L. S. B. (1990) *Science* **249**, 42–47.
21. Leopold, P. L., McDowall, A. W., Pfister, K. K., Bloom, G. S. & Brady, S. T. (1990) *J. Cell Biol.* **111**, 417a (abstr.).
22. Wagner, M. C., Pfister, K. K., Bloom, G. S. & Brady, S. T. (1989) *Cell Motil. Cytoskeleton* **12**, 195–215.
23. Salzer, J. L., Holmes, W. P. & Colman, D. R. (1987) *J. Cell Biol.* **104**, 957–965.
24. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
25. Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidmen, J., Smith, J. & Struhl, K. (1987) *Current Protocols in Molecular Biology* (Wiley-Interscience, New York).
26. Sanger, F., Nicklen, S. & Coulson, A. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
27. Jameson, B. A. & Wolf, H. (1988) *CABIOS* **4**, 181–186.
28. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
29. Schowalter, D. B. & Sommer, S. S. (1989) *Anal. Biochem.* **177**, 90–94.
30. Tsou, A.-P., Lai, C., Danielson, P., Noonan, D. J. & Sutcliffe, J. G. (1986) *Mol. Cell. Biol.* **6**, 768–778.
31. Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872.
32. Cohen, C. & Parry, D. A. D. (1986) *Trends Biochem. Sci.* **11**, 245–248.
33. McLachlan, A. D. & Karn, J. (1983) *J. Mol. Biol.* **164**, 605–626.
34. Bloom, G. S., Wagner, M. C., Pfister, K. K., Leopold, P. L. & Brady, S. T. (1989) in *Cell Movement*, eds. Warner, F. D. & McIntosh, J. R. (Liss, New York), Vol. 2, pp. 321–333.
35. Jackson, A. P., Soew, H.-F., Holmes, N., Drickamer, K. & Parham, P. (1987) *Nature (London)* **326**, 154–159.
36. Burgoyne, R. D., Cambry-Deakin, M. A., Lewis, S. A., Sarkar, S. & Cowan, N. J. (1988) *EMBO J.* **7**, 2311–2319.
37. Douglas, M. G., McCammon, M. T. & Vassarotti, A. (1986) *Microbiol. Rev.* **50**, 166–178.
38. Warrick, H. M. & Spudich, J. A. (1987) *Annu. Rev. Cell Biol.* **3**, 379–421.
39. Buster, D., Lohka, M. & Scholey, J. M. (1990) *J. Cell Biol.* **111**, 418a (abstr.).
40. Murphy, D. B., McNiven, M. A., Wallis, K. T., Kutznetsov, S. A. & Gelfand, V. I. (1989) *J. Cell Biol.* **109**, 80a (abstr.).