Kinesin-73 in the nervous system of *Drosophila* embryos

(kinesin-73 heavy chain cDNA/cytoskeleton-associated protein Gly-rich domain)

HSI-PING LI, ZHENG-MEI LIU, AND MARSHALL NIRENBERG*

Laboratory of Biochemical Genetics, Building 36, Room 1C-06, National Heart, Lung, and Blood Institute, National Institutes of Health, 36 Convent Drive, MSC 4036, Bethesda, MD 20892-4036

Contributed by Marshall Nirenberg, December 9, 1996

ABSTRACT Kinesin-73 cDNA was shown to encode a kinesin heavy chain protein that contains an N-terminal motor domain and a long central region that lacks extensive coiled-coils. The amino acid sequence of the motor domain of kinesin-73 protein is most closely related to the motor domains of Caenorhabditis elegans unc-104 and mouse KIF1A. The central region of kinesin-73 protein also is related to unc-104 and KIF1A, but the homology is lower than that of the motor domain. The C-terminal region of kinesin-73 protein contains a cytoskeleton associated protein Gly-rich domain, which is a putative microtubule binding site that is present in some cytoskeleton or dynein-associated proteins. Kinesin-73 mRNA was shown by in situ hybridization to be maternally expressed and widely distributed in the syncytial blastoderm embryo. However, later in Drosophila embryo development, expression of the kinesin-73 gene becomes restricted mostly to the central and peripheral nervous systems.

Kinesin heavy chain proteins are molecular motors that transport cellular organelles along microtubule pathways. In recent years many genes or cDNAs that encode proteins of the kinesin heavy chain superfamily have been identified (for reviews, see refs. 1–5). Each kinesin heavy chain protein has a motor domain ≈ 350 amino acid residues in length that contains a putative ATP binding site and a microtubule binding site. The native form of most kinesin molecules is a tetrameric complex consisting of two kinesin heavy chains and two kinesin light chains. However, some species of kinesin heavy chain proteins form trimeric complexes (for review, see ref. 4), and others, such as the unc-104 subfamily of kinesin proteins, remain as monomers (6–8).

In Drosophila more than 30 kinesin heavy chain genes have been mapped on chromosomes (9), and the nucleotide sequences of nucleic acids encoding 9 kinesin heavy chain proteins have been reported (9–15). Different kinds of kinesin proteins transport different cargoes and hence perform different functions. Among the functions of kinesin proteins that have been identified are the anterograde transport (toward the plus ends of microtubules) of synaptic vesicle precursors (7, 16), mitochondria transport (8), spindle function, and chromosome distribution (for reviews, see refs. 1–5). Some kinesins, with C-terminal rather than Nterminal motor domains, transport cargo in the opposite direction—i.e., from the periphery toward the minus ends of microtubules (9) in the cell soma.

We have used the enhancer trap method (17) to find genes that are expressed in the nervous system of *Drosophila* embryos. *Drosophila* genomic DNA fragments adjacent to P-element insertion sites were cloned from transgenic lines of *Drosophila*, and these DNA fragments then were used as

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked *"advertisement"* in accordance with 18 U.S.C. §1734 solely to indicate this fact.

probes to clone the corresponding species of cDNA. In this report kinesin-73 cDNA is described, which encodes a novel kinesin heavy chain protein that becomes restricted primarily to the central and peripheral nervous systems during the course of embryonic development. Kinesin-73 is a member of the *Caenorhabditis elegans* unc-104 (6–8) subfamily of kinesin heavy chain proteins.

MATERIALS AND METHODS

Generation of Transgenic Fly Lines. Transgenic lines of *Drosophila* were generated using P-*lacW* and the P-element mobilization method described by Bier *et al.* (18), and balanced using appropriate second or third chromosome balancers.

Plasmid Rescue and Library Screening. Isolation of genomic DNA from transgenic flies and plasmid rescue were performed as described by Pirrotta (19) with modifications. DNA was digested with EcoRI, and DNA fragments were circularized by ligation and used to transform Escherichia coli XL-1 Blue competent cells. Genomic DNA fragments flanking P-element insertion sites were excised from plasmid DNA by digestion with EcoRI and PstI simultaneously. These fragments were used to screen λ Charon 4 and λ GEM-11 (Promega) *Drosophila* genomic DNA libraries. Genomic DNA fragments were excised from positive λ phages and used as probes to screen a 3- to 24-hr Drosophila embryo cDNA library in $\lambda gt10$. cDNA fragments from positive clones then were subcloned into pBluescript vectors, and both strands of DNA were sequenced using Sequenase Version 2.0 DNA sequencing kits (United States Biochemical). Wisconsin Sequence Analysis programs (Genetics Computer Group, Madison WI) were used for sequence analysis and DNA segment assembly.

In Situ Hybridization. A digoxigenin-labeled singlestranded antisense RNA probe (nucleotide residues 568-1658 shown in Fig. 2) was transcribed from a 1.1-kb kinesin-73 cDNA subclone (73C-8) using the Boehringer Mannheim Dig RNA labeling kit (SP6/T7) and the manufacturer's instructions. Wild-type Oregon R embryos were processed for hybridization by the method of Tautz and Pfeifle (20). Prehybridization, hybridization, and washes were performed as described in Mellerick and Nirenberg (21).

RESULTS

The enhancer trap method (17) was used to find genes that are expressed in the nervous system of *Drosophila* embryos.

Abbreviation: CAP-Gly domain, cytoskeleton-associated protein Glyrich domain.

The sequence reported in this paper has been deposited in the GenBank database (accession no. U81788).

^{*}To whom reprint requests should be addressed. e-mail: marshall@ codon.nih.gov.

Approximately 500 transgenic lines of *Drosophila* were generated using the P-*lacW* vector and the P-element mobilization method of Bier *et al.* (18). P-*lacW* contains the initial portion of the P-element transposase gene fused to a β -galactosidase reporter gene. The time and pattern of β -galactosidase expression during development usually are determined by nearby regulatory sequences of the gene that contains the newly inserted P-element DNA. Eighty-two transgenic fly lines were obtained that express β -galactosidase in embryos only in the nervous system, and additional lines were found that express β -galactosidase both in the nervous system and in one or more other tissues. We cloned and characterized *Drosophila* genomic DNA adjacent to the P-element insertion site and corresponding cDNA from some of these transgenic lines.

A 3.0-kb genomic DNA fragment adjacent to the P-lacW DNA inserted in chromosome 2 of *Drosophila* transgenic line 73 was cloned and used as a probe to screen a pGEM-11 *Drosophila* genomic DNA library. The genomic DNA insert from a positive pGEM-11 recombinant clone [73(1)-1, 12-kb DNA insert] then was used to screen a 3- to 24-hr *Drosophila* embryo cDNA library in λ gt10. cDNA clone 73C-1 was obtained with a DNA insert 247 nucleotide residues in length. Nucleotide sequence analysis showed that clone 73C-1 cDNA encodes a protein homologous to part of the motor domain region of kinesin heavy chain protein. Additional screening of the cDNA library yielded a set of overlapping cDNA clones that are shown in Fig. 1. Kinesin-73 cDNA is 7.1 kb in length. A partial map of restriction sites also is shown.

The nucleotide sequence of kinesin-73 cDNA and deduced amino acid sequence of kinesin-73 heavy chain protein are shown in Fig. 2. Kinesin-73 cDNA is 7107 nucleotide residues in length without the poly(A) tail and encodes a protein 1921 amino acid residues in length with a calculated M_r of 215,047 and a pI of 5.62. The 5'-untranslated region contains termination codons in each of the three reading frames. Eight of the 12 nucleotide residues flanking the ATG codon for initiation of kinesin-73 protein synthesis fit the Cavener and Ray (22) consensus sequence (CACAACCAACATGGC) flanking ATG codons initiating protein synthesis in Drosophila. Kinesin-73 protein contains a motor domain 359 amino acid residues in length in the N-terminal region of the protein. A conserved ATP binding site (23) and a putative microtubule binding site (amino acid residues 244-255) are present within the motor domain. A conserved amino acid sequence termed a CAP-Gly domain (residues 1826–1868) is present in the C-terminal region of kinesin-73 protein. CAP-Gly domains are present in some cytoskeleton- or dynein-associated proteins such as human restin (24) also termed CLIP-170 (25), Drosophila Glued (26), rat dynactin (27), C. elegans MO1A8.4 (28), and BIK1 of Saccharomyces cerevisiae (29). The amino acid sequence, Arg-Gly-Asp-Ser, near the C terminus of kinesin-73 is a putative fibronectin binding site (30). The 3'-untranslated region of kinesin-73 cDNA contains a polyadenylylation signal, AATAAA, 21



FIG. 1. Overlapping kinesin-73 cDNA clones and partial map of restriction sites of the composite kinesin-73 cDNA. B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; S, *SacI*; X, *XbaI*.

nucleotide residues upstream from the poly(A) tail [the poly(A) tail is not shown in Fig. 2].

Most kinesin heavy chain proteins are composed of three domains: a globular motor domain ≈ 350 amino acid residues in length with binding sites for ATP and microtubules, a central stalk domain with many α -helical heptad repeats that form extensive coiled–coils that enable some kinesin heavy chain molecules to form dimers, and a globular tail domain that is thought to bind to organelles that are transported along microtubule tracks.

Coiled-coils predicted for kinesin-73 protein by the algorithm of Lupas *et al.* (31) are shown in Fig. 3. Kinesin-73 protein does not contain a central region with long α -helices and extensive coiled-coils, but does contain two or three short α -helices that probably form stable coiled-coils. Further work is required to determine whether the putative short coiledcoils interact with other protein molecules. The central region of kinesin-73 protein therefore resembles the central regions of mouse kinesin heavy chain proteins KIF1A (6) and KIF1B (7), which also have several short putative coiled-coils in the central region and *C. elegans* unc-104 (8) that are thought to function as monomers.

The amino acid sequence of kinesin-73 protein is most closely related to the amino acid sequences of C. elegans unc-104 (8), mouse KIF1A and KIF1B (7), and human H-ATSV (axonal transporter of synaptic vesicles) (32), which is 97% homologous to KIF1A. Regions of homology between kinesin-73 and unc-104, KIF1A, and KIF1B, revealed by dot matrix analysis, are shown in Fig. 4 A-C, respectively. The amino acid sequence of the motor domain of kinesin-73 protein (residues 1-359) is 59%, 62%, and 50% identical to the motor domains of unc-104, KIF1A, and KIF1B, respectively. Kinesin-73 amino acid residues 360 to \approx 607 also exhibit strong homology to the corresponding regions of unc-104, KIF1A, and KIF1B although some gaps appear. Short segments of homology separated by larger segments that are not homologous are present between amino acid residues 608 to 1246 of kinesin-73 and the corresponding regions of unc-104 and KIF1A, but these segments of homology are not present in KIF1B. Some additional regions of similarity were detected between kinesin-73 amino acid residues 1246 to 1767 and the corresponding regions of unc-104 and KIF1A that are not shown here. Amino acid sequence homology between kinesin-73 and Drosophila KHC (10) or KLP68D (14) is restricted to the motor domain (data not shown). These results suggest that kinesin-73 protein is a member of the unc-104 subfamily of kinesin heavy chain proteins.

A search of the GenBank database for additional homology revealed a 43 amino acid residue CAP-Gly domain (33) in the C-terminal region of kinesin-73. A comparison of the amino acid sequence of the CAP-Gly domain of kinesin-73 with five other proteins that contain CAP-Gly domains is shown in Fig. 5. Human restin (24), also termed cytoplasmic linker protein-170 or CLIP-170, contains two CAP-Gly domains that bind to microtubules (25). Restin links endocytic vesicles to microtubules (24). Both Drosophila Glued (26) and rat dynactin (27) are associated with 20S complexes involved in cytoplasmic dynein-mediated vesicle transport. A CAP-Gly domain also is present in MO1A8.4 protein of C. elegans (28) and BIK1 protein of S. cerevisiae (29). BIK1 apparently is required for the formation or stabilization of microtubules during mitosis and for spindle pole body fusion during conjugation (29). The identity between the amino acid sequence of the CAP-Gly domain of kinesin-73 and those of the other proteins shown in Fig. 5 ranges from 60% to 41%. Twelve of the 43 CAP-Gly amino acid residues are invariant, and only 1 or 2 amino acid replacements are present at 8 additional amino acid positions. Conservative amino acid replacements are present at 12 positions.

In situ hybridization with an antisense RNA probe showed that kinesin-73 is a maternal mRNA that is evenly distributed

GGGCTGAGAAAAGCGAAAAAGGGAAAAATGAAAAAGGCAATACAGAAAATCAAAATTGAAGTGAAACCAAGTGGGAAAACGAAGGAGGTAAAGATTGGCGTGCTCATAACCCACGGAGAG	120
GACAACTACGGCTAACAACTAACGGTCATAAAAGTCAAAAGCAACGAAGCAACGAAGGCGGGGAGGAGGAGGCACGGCCCCGCCTGCAGGCCCCGCCTTAGGGAGCCCCGCCATCCGCATATC	240
CACTCCGCTTCAACTAACAGAGATTCCAAAAACAACACCCCCGCTGCATCCCACATCCCGCGTGGTGGTGGTGATAACATCCAAGTCGCGCCGCGTAGGGGTGCGACCCTTCAATCGCCGAGATCGA	360
M & C D K T K V B V P V P P F N D P F T	21
ACTAGATACGAAATGTATCGTGGAAATGGAAAAACAGCAGACGATACTGCAGAATGCGCGCCGCCACTGGAAAAAATGGAGAAAAACAACCAAAGGACAAAGCACTGCATTCGATCGA	480 61
CTCATTGAACCCCGAGGACGAGAACTTTGCGTCCCAGGAGACAGTGTTCGATTGCGTGGGACGTGGAACTTCGGATAATGCATTGCAGGCGATAATGCGTGCATATTCGCTTACGGCCA	600
S L N P E D E N F A S Q E T V F D C V G R G I L D N A F Q G Y N A C I F A Y G Q	101
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	720 141
CAAGGTGGAGGTGTCCTACATGGAGATCTATAACGAGAAGGTCCACGATCTGGTCGATCCCAAGCGAACAAACA	840 181
CGGACUGTCGCAGCIGCCIGTGACATCCTATICAGGACATCGATGACCGAAGGACAACAACTCACGAACGAGGGGGCGCCGCCACCAATATAGATGCCGGAGTCGTCGCCC	960
G L S Q L A V T S Y Q D I D N L M T E G N K S R T V A A T N M N A E S S R S H A	221
V F S V V L T Q I L T D Q A T G V S G E K V S R M S L V D L A G S E R A V K T G G S G CALCACTTGGCCTGTCATCTCCAAGTTGGCCGATCAATCGAAGAAGACGGCTAACGACAAATCTCTGAACGACAACTTGGCCTGGTCATCTCCCAAGTTGGCCGATCAATCGACGACAAATCTCTGACGACAAATCTCTGACGACGACGACGACGACGACGACGACGACGACGACGACG	261
A V G D R L K E G S N I N K S L T T L G L V I S K L A D Q S N G K K S G N D K F	301
CGTCCCTACCGCGATTCCGTGCTCACCGGCGTGCAAAGAAAAATTTGGTGGGCAATTCCAGAACTGTTATGGTAGCGACAATCTCGGCACATAACTACGAGGAGAACGCTTTC	1320
V P Y R D S V L T W L L K D N L G G N S R T V M V A T I S P S A D N Y E E T L S	341
CACGCTGCGTTATGCAGATCGAGCCAAGCGCATTGTTAACCACGCTGTTGTCAACGAAGATCCCAATGCCCGCATCATTGGTGAGCTGCGACACGAGGTGGAGACCCTTAGAAGTATGCT	1440
T L R Y A D R A K R I V N H A V V N E D P N A R I I R E L R H E V E T L R S M L GAAACAGCCACTGGGTGGGGGGATGTGCGAGGAGGAGGAGGAGGGGGAGAGCGGGTAGGAAGGA	381 1560 421
ANACGAACGGCAGGCAGGCAGGCAGGCAGGCAGGAGAAATGGGCATTAGTGTACAGGGCAGGCA	1680 461
GCTGGTCTATTATCTCAAGGATCGCACCAGTATTGGCGGACGCCATATCAGTGGCCAGCAGCGGATATACAACTTTCGGGTCTGGGAATCCAGCCCGAGCACTGTGTGATCACTATCGA	1800
L V Y L K D R T L I G G R T I S G Q Q P D I Q L S G L G I Q P E H C V I T I E	501
GGATAGTGGACTGTACATGGAGCAGTGCAGTGACGCGTTGCTTTGTGAACGGATCTGCTGGGAAAAGACCACCACTCCGGAACGGCGACGTATCCTGTGGGGCAACCACCAC	1920
D S G L Y M E P V Q G A R C F V N G S A A V E K T P L Q N G D R I L W G N H H F	541
TTTCCCGCTCAACTGCCGAAGACCAATAACAACTAGTATGTGCGCTTCGGAGCCCCAGACGAGCCGACGCGACGCAACGACGACGAC	2040
CAACGATECTATECTAGACGGCCATTGCTCGGCTGGGAACGTCAGCACGGAGGAGAGAAGAACGACGGCGGGGGGGG	2160 621
TCTGTCGCCTAGTACACCATATIGCACCTTATGGTCGTTATGATCCACTGGGCATGGGGTAGGGTTACCCCCAATATCTCCTACTTCGCAATGGGAGGGGGAAAAAGTGGGGCACAGAAGA	2280
L S P S T P Y A P Y A P Y D P L R M G K I T P N T P T S Q M R V E K W A Q E R D	661
TGGATGTTTCGTCGCAGTTTGGGTCAGCTGAAAAGGGATATTATGCGTGCG	2400
E M F R R S L G Q L K T D I M R A N S L V Q E A N F L A E E M E K K T K F S V T	701
TCTGCAGATTCCGCCAGCTAATCTAACAGAAGCGCTGGGCATTTGTCAGCGAACCCGCCATTCTGGCGAAACCCGCAAACTCTGGTGAGCAAGATCTGGACGAAGAACCT	2520
CGAAACAGCTGATAGCCAGATGCGACGAGCAGCAGAGGAGCGCGCTTGTGAAGGGCGATTGCTGATGAGCGAGC	2640 781
TGCCAAGCCACAGGATCCGTTFTACGAGTCGCAAGAGAACCACAATCTTATTGGCGTGGCCAATATTATTCTTGGAGGTTCTGTTTCATGACGTCAAGCTGGACTACCACGCCCATCAT	2760
A K P Q D P F Y E S Q E N H N L I G V A N I F L E V L F H D V K L D Y H T P I I	821
TACCTAACCAAGCAAGCTAGTCGATCTAAGTCGACTGACGCGACTGCCCGCCC	2880
S Q Q G E V A G R L Q V E I E R I A G Q M P Q D R M C E S V S E S S G D S R D E	861
GTACGACGACCCCTGTGGATCCCACATGGATCACGATTACCTGCGTGTGGCGATGGCCATTGGTCGCCTCCCAACTTTGTCGCTTTTGCCGTTTTTGGGG	3000
I D D F V D F T S N Q I T C K V T I K C A S G L P L S L S N F V F C Q I T F W GA TCACCAAGAGATGGTGTTGTCCGGTTATCAATGCGATGACCAGGACCAGGACATGGTGTTTAGGTTTGAACACCACCCCCGGACTTACCATAACGAAGAGTTTTTGGA H Q E M V V P V I N A E S T A H D Q N M V F K F E H T Q D F T V T I N E E F L E	3120 941
GCACTGCATAGAAGGAGCTCTGTCCATTGAGGGGGGTGTGGGGCATCGCAGCGGCGGCGGCGGAGTGGAACGCGAACAGCGAGGCGGAGCGGAGGCGAGGCGGAGGGGGG	3240 981 3360
W A E L S R K I E L W V E I H E L N D N G E Y S P V E V T N R N E V L T G G I Y	1021
CCASTTACGTCAGGGGCGAGGGGGGGGAATGTAGGGGGGGGAAAACTCTGGCACTTGGCCATTGGCAGTGGCAGTTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGTGGCAGT	3480
CCALCTCCGTTGCAGCCACCACTGGATTCTTACCAGGAGGAGGAGGAGCGAGGAGCGAGGAGCGAGGAGCGAGGAG	3600
GCTTATAAAGAAAGAGGAGAAAGAACAGAACAGGAGCGCGAGGGAGGGAGCTAAGCTTGGTTCATCAGTGGGTCTCGTTGACGGGGGAACGCAACGCGGTGTTGGTGCCAGCTCGGGG	3720
L I K K E E K N E Q E R E R E L S L V H Q W V S L T E E R N A V L V P A P G S G	1141
CATTCCCGGAGCACCTCATGGGAGCGCGCCTCGGAATGGACCACATGTGCCGTCTTCTTCTTCAACCTAAAGGTCGGCGACAAACACAACAACAACAACAACGACGACT	3840
I P G A P A S W E P P S G M E P H V P V L F L N L N G D D L S A Q N T N D E L S	1181
CGTCGCCGGCATCCATTCCTCTCCCAAGAGAGCACGAACATAAGTTTTACACGCTGCAAATTCTGCAGCACCTAGATAAGGATGTTTGCTGTGTGGCAGCCAGC	3960
VAGINSILSKEHGHKFYTLQHLDKDVCCVASWDSSMH TGACAGTCAGGCCGTGAACCGTGTGAGCCGAGGGCCATGAGCGGGTCTATCTCATTCTGCGCACCAACGGACCCACGCACCTATGGATCTCGTCCGCCAGGGAGCGAAGCGACTAAG DSQALNRVTEANERVYLIRTVRLSHPAPMDLVLRKRLS	4080 1261
CATTAACATCAAGAAGGGACAGACGGCTAACCGACCGCCTGAAGAAATTCCGTCTTGTGCGGGGAGAGAGCGCCATCTGGCAGGGGGGTGTCACCTATGAGGTGGTCTTAACATTCCAAA	4200
INIKKGQTLTDAACATTCCAAA	1301
ACCTTCCGACGACTTCGACGACTGCCGTCGTTGTGCGCGGGAGAGGACGCCACTGGCGGAGACGGCGGTGTCACCTATGAGGGGCGCGCGTGTTCTTACATTCCAACATTCCAACA	4320
A S E E L E D R E S L A Q L A A S G D D C S A S D G E T Y I E K Y T R G V S A V	1341
GGAGAGCATACTGACTCTGGATAGACTGCGGCAGAATGTGGCGGTCAAGGAGCTGGAAACGGCACATGGACAGCCGTCAGCATGCGCAAGACCGTCAGCGTGCCGAACTTCTCACAGCA	4440
ESILTLDRLRQNVKELETAHGQPLSMRKTVSVPNFSQQ	4560
ACTCATAAATAAACTTACGCAGATTTAGGCCATCGATGGAGGGGGGGG	1421
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	4680 1461 4800
S K L L G M R M T T L H E E P L G G H R S L D E E P E D S Y S D S E Y A A E Y E	1501
ACAGGAGCGGCAGCAGCAGCAACAAAAGCATGGCCACGCGATCCCGCGCTCCAGGCCTTCCAAGACCATGGACCCAGCAGCCGCCAGCAGCACCAAAGCTACTTGAGCTACCGCGC	4920
Q E R Q Q N K S M A T R S R L T A S K T M D S F M D V S S H S N Q S Y L S Y T S	1541
CAGTGCCAACGCTAACATGAAGCACCTGACCGCCGCCGCCGCCGCCGCCACGCGGCCCCGGGTCCCCGGCGCGCCACGAAAATCTGAGCAACGAGGAATAT	5040
S A N A N M K H L T G L A T L S M S S S T S S G Y G S Q A V S C N N L S N E D I	1581
TGCTTCAATGCGTTCCATGAGCATTGATGAGACACCAGAATTCGATCGGGTCAACTCGAATTCGCCACCGAACCGGCAGGCA	5160 1621 5280
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	1661 5400
K N N N N N V D L V N E P K P L S G Q M E L E E P M S Q P E S Q T E F A T D N Q	1701
GAACGGCAACAGGTCCTCCGACGAGTTAAGCCACAGTTCCCGAGGAGGATGCGCTGGAGGCGGAGGAAGGTGCCCAATAC	5520
N G N R S S D E L S H S S E D L L E G D G I V R E L P A G K V V R R K K S N T	1741
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	5640 1781 5760
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	1821 5880
I R F V <u>G T T E F O P G A W I G V E L D T P T G K N D G S V K G V O Y F O C K P</u>	1861
CAAGCACGGCATGTTTGTGCGCTGCTGCGACAAGGGTGACCAGGCGATGCGAGCCTACAAGGCCGCCGAGAGAGCAACAGCATCAGCAATGAGTACCTCAAT	6000
<u>E BERLEYK</u> SUKLMEURKGKAMKAYKAAEKSNSISKEMST SM	1901
GACTGGCTGGATGACAGGCCGCGGGGGATTGGCTAAACTTGGGGGGCGTAAATTGGAGGGCTAAGTGGTGGACTAGTGGGGCGATAATTGGAGGCAATCGGC	6120
T G S M T R S K S C G D S L N L S A R K *	1921
GGTAGCCACCAGCCAGCGAGCGACCTGCCACAGATTGGGTGCAACCAGCGGCGCTGCAAGCAGTGCGAAGAAGTGCTATGGCAAGAAGTGCTATGGCAAGAAGTGCTTCTAAAG	6240
uuuuluuuanaatiskuaatistuskaasuskastiskasessa kaisessa kaisessa kaisessa kaisessa kaisessa kaisessa kaisessa kai	6360
Toacocoatigoggaaatigaaatigaatigagtaggaggaggaggaggaggaggagtitaaggaggtitaaggaggagtictacgggagaataatitatigaggaatiga	6480
Togatactaatigaaatigaggaggaggaggaggaggaggaggaggaggaggaggag	6600
$\label{eq:construction} TGATTCCATTGTTGTGTGTGTAATTGTCGTTTTTGCTTTTTGCTTTTGCAATGAAGCGTTTTACTGATATATTTTCTAATGGTTCTTTGCCTTTTAGTTAAGTTTTTGTTTTTTTT$	6720 6840
AGATTITEGTGTCTTGAACTACTCCGCATTGAAGCATCGCACGCCGCCGCGCGCG	7080 7107

FIG. 2. Nucleotide sequence and the deduced amino acid sequence of the composite kinesin-73 cDNA. The number of nucleotide residues and the number of amino acid residues are shown on the right. The motor domain (amino acid residues 1–359) is indicated with a dashed line. The phosphate-binding loop (P-loop) of the ATP binding site is shown using boldface amino acid residues. Underlined amino acid residues (1819–1874) correspond to the cytoskeleton-associated protein Gly-rich domain (CAP-Gly domain). Double underlining indicates a fibronectin cell attachment amino acid sequence (RGDS). The polyadenylylation signal (AATAAA) of kinesin-73 cDNA also is underlined.



FIG. 3. The probability of forming stable coiled–coils predicted by the algorithm of Lupas *et al.* (31) plotted against the amino acid residues of kinesin-73. Possible regions with stable coiled–coils are amino acid residues 594–622, 668–695, and 734–761.

in the syncytial embryo (not shown). As shown in Fig. 64, kinesin-73 mRNA is expressed in the germ band and part of the procephalic region in a stage 9 embryo. Kinesin-73 mRNA is prominently expressed in the central nervous system of a Stage 13 embryo (Fig. 6B). Kinesin-73 mRNA is present in both the central and peripheral nervous systems of stage 17 *Drosophila* embryos (Fig. 6 *C* and *D*). These results show that the kinesin-73 gene is widely expressed in early *Drosophila* embryos; however, later in embryonic development expression of the gene is restricted mostly to the central and peripheral nervous systems.

DISCUSSION

Kinesin-73 cDNA was cloned and sequenced and shown to encode a novel kinesin heavy chain protein. Kinesin-73 protein consists of 1921 amino acid residues and contains a motor domain with a putative ATP binding site and a microtubule binding site in the N-terminal portion of the protein. The amino acid sequence of the motor domain and the central region of kinesin-73 show that kinesin-73 is a member of the unc-104 subfamily of kinesin heavy chain proteins, which include *C. elegans* unc-104 (6), mouse KIF1A, (7) and KIF1B (8), and human H-ATSV (32). The stalk domains of most kinesin heavy chain proteins are thought to form coiled–coil structures that enable kinesin heavy chain molecules to form dimers. Kinesin-73, KIF1A, and KIF1B (7) apparently have only a few short coiled–coils



FIG. 4. Dot matrix amino acid sequence comparisons of kinesin-73 with unc-104 (A), KIF1A (B), and KIF1B (C). Amino acid sequence comparisons were performed using the Genetics Computer Group program COMPARE (window = 30, stringency = 16), and then plotted using DOTPLOT.

in their stalk domains, and, together with unc-104, lack extensive coiled–coils in the central regions of these proteins.

The C-terminal region of kinesin-73 protein contains a CAP-Gly domain that is similar to the CAP-Gly domains of some cytoskeleton associated proteins and dynein-associated proteins such as restin (24) [CLIP-170 (25)], which links endocytic vesicles to microtubules (24); *Drosophila* Glued



FIG. 5. Amino acid sequence alignment of the CAP-Gly domain of kinesin-73 protein with CAP-Gly domains of other proteins. Species abbreviations are as follows: d, *Drosophila*; h, human; r, rat; ce, *C. elegans*; and y, the yeast *S. cerevisiae*. A dash shown in white on a black background corresponds to the same amino acid residue as shown for kinesin-73. The column of numbers on the right corresponds to the percent of amino acid residues that are identical to those shown for kinesin-73. Amino acid residues shown in black on a shaded background correspond to conservative amino acid replacements, defined as follows: S, T, G, A, P/L, I, M, V/E, D, Q, N/K, R, H/F, V, W/, and C (34). The symbol "." corresponds to the absence of an amino acid residue. The CAP-Gly domain of the proteins correspond to the following amino acid residues: kinesin-73, 1826–1868; restin CAP-Gly domain-1, 78–120; restin CAP-Gly domain-2, 232–264; Glued, 56–98; dynactin, 47–89; MO1A8.4, 39–81; BIK1, 26–69. References are as follows: restin (24), *Glued* (26), dynactin (27), MO1A8.4 (28), and BIK1 (29).



FIG. 6. Distribution of kinesin-73 in RNA in *Drosophila* embryos detected by *in situ* hybridization with a digoxigenin-labeled (-) RNA probe. (A) Stage 9 embryo. (B) Stage 13 embryo. (C and D) Stage 17 embryos, side view (C) and ventral view (D). Kinesin-73 mRNA is present in the central and peripheral nervous systems.

(26) and rat dynactin (27), which are involved in cytoplasmic dynein-mediated vesicle transport; and *C. elegans* M01A8.4 protein (28) and yeast BIK1 protein (29). CLIP-170 contains two CAP-Gly domains that have been shown to bind to microtubules (25). Only the CAP-Gly domain of kinesin-73 protein exhibits homology with these proteins. Kinesin-73 protein is the only kinesin heavy chain protein that has been reported thus far that contains a CAP-Gly domain. The presence of a CAP-Gly putative microtubule binding site in the tail region of kinesin-73 protein suggests that kinesin-73 may be a motor protein for anterograde axonal transport of tubulin oligomers and/or microtubules along microtubule tracks.

Microtubules elongate by addition of tubulin to the plus ends of the microtubules. Axonal microtubules are oriented so that the elongating plus ends point away from the cell body toward the axon tip. In addition to microtubule assembly by addition of tubulin subunits to the plus ends of microtubules in axons, evidence for tubulin or microtubule transport in axons from neuronal soma toward axon tips has been reported (for a review see ref. 35). Further work is needed to determine whether the function of kinesin-73 involves microtubule transport.

We thank Lily Jan and Yuh Nurgston for *Drosophila* lines, Norma Heaton for help with maintenance of *Drosophila* lines, Vicky Guo for oligonucleotide synthesis, David Landsman for computer analysis assessing the probability of coiled–coils in kinesin-73, and Alan Peterkofsky for comments on the manuscript.

- 1. Bloom, G. S. & Endow, S. A. (1994) Protein Profile 1, 1059-1116.
- Walker, R. A. & Sheetz, M. P. (1993) Annu. Rev. Biochem. 62, 429–451.
- Barton, N. R. & Goldstein, L. S. B. (1996) Proc. Natl. Acad. Sci. USA 93, 1735–1742.
- 4. Scholey, J. M. (1996) J. Cell Biol. 133, 1-4.
- Brady, S. T. & Sperry, A. O. (1995) Curr. Opin. Neurobiol. 5, 551–558.
- Otsuka, A. J., Jeyaprakash, A., García-Añoveros, J., Tang, L. Z., Fisk, G., Hartshorne, T., Franco, R. & Born, T. (1991) *Neuron* 6, 113–122.
- Okada, Y., Yamazaki, H., Sekine-Aizawa, Y. & Hirokawa, N. (1995) Cell 81, 769–780.

- Nangaku, M., Sato-Yoshitake, R., Okada, Y., Noda, Y., Takemura, R., Yamazaki, H. & Hirokawa, N. (1994) *Cell* 79, 1209– 1220.
- Endow, S. A. & Hatsumi, M. (1991) Proc. Natl. Acad. Sci. USA 88, 4424–4427.
- Yang, J. T., Laymon, R. A. & Goldstein, L. S. B. (1989) Cell 56, 879–889.
- Zhang, P., Knowles, B. A., Goldstein, L. S. B. & Hawley, R. S. (1990) Cell 62, 1053–1062.
- Stewart, R. J., Pesavento, P. A., Woerpel, D. N. & Goldstein, L. S. B. (1991) Proc. Natl. Acad. Sci. USA 88, 8470–8474.
- Heck, M. M. S., Pereira, A., Pesavento, P., Yannoni, Y., Spradling, A. C. & Goldstein, L. S. B. (1993) *J. Cell Biol.* 123, 665–679.
- Pesavento, P. A., Stewart, R. J. & Goldstein, L. S. B. (1994) J. Cell Biol. 127, 1041–1048.
- Williams, B. C., Riedy, M. F., Williams, E. V., Gatti, M. & Goldberg, M. L. (1995) J. Cell Biol. 129, 709–723.
- 16. Hall, D. H. & Hedgecock, E. M. (1991) Cell 65, 837-847.
- O'Kane, C. J. & Gehring, W. J. (1987) Proc. Natl. Acad. Sci. USA 84, 9123–9127.
- Bier, E., Vaessin, H., Shepherd, S., Lee, K., McKall, K., Barbel, S., Ackerman, L., Carretto, R., Uemura, T., Grell, E., Jan, L. Y. & Jan, Y. N. (1989) *Genes Dev.* 3, 1273–1287.
- Pirrotta, V. (1986) in *Drosophila: A Practical Approach*, ed. Roberts, D. B. (IRL, Oxford), pp. 83–110.
- 20. Tautz, D. & Pfeifle, C. (1989) Chromosoma 98, 81-85.
- 21. Mellerick, D. M. & Nirenberg, M. (1995) Dev. Biol. 171, 306-316.
- 22. Cavener, D. R. & Ray, S. C. (1991) Nucleic Acids Res. 19, 3185–3192.
- Saraste, M., Sibbald, P. R. & Wittinghofer, A. (1990) Trends Biochem. Sci. 15, 430–434.
- Bilbe, G., Delabie, J., Bruggen, J., Richener, H., Asselbergs, F. A. M., Cerletti, N., Sorg, C., Odink, K., Tarcsay, L., Wiesendanger, W., DeWolf-Peeters, C. & Shipman, R. (1992) *EMBO J.* 11, 2103–2113.
- 25. Pierre, P., Scheel, J., Rickard, J. E. & Kreis, T. E. (1992) *Cell* **70**, 887–900.
- Swaroop, A., Swaroop, M. & Garen, A. (1987) Proc. Natl. Acad. Sci. USA 84, 6501–6505.
- Gill, S. R., Schroer, T. A., Szilak, I., Steuer, E. R., Sheetz, M. P. & Cleveland, D. W. (1991) J. Cell Biol. 115, 1639–1650.
- Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., et al. (1994) Nature (London) 368, 32–38.

Biochemistry: Li et al.

- 29. Trueheart, J., Boeke, J. D. & Fink, G. R. (1987) Mol. Cell. Biol. 7, 2316–2328.
- 30. Pierschbacher, M. D. & Ruoslahti, E. (1984) Nature (London) **309,** 30–33.
- 31. Lupas, A., Van Dyke, M. & Stock J. (1991) Science 252 1162-1164.
- 32. Furlong, R. A., Zhou, C. Y., Ferguson-Smith, M. A. & Affara,

N. A. (1996) Genomics 33, 421-429.

- 33. Riehemann, K. & Sorg, C. (1993) Trends Biochem. Sci 18, 82-83.
- Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1979) in Atlas of Protein Sequence and Structure, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), pp. 345–352. 35. Black, M. M. (1994) *Prog. Brain Res.* **102**, 61–77.