Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 8477–8481, November 1986
Biochemistry

Tissue-specific and constitutive α-tubulin genes of *Drosophila melanogaster* code for structurally distinct proteins

(gene family)

WILLIAM E. THEURKAUF, HOWARD BAUM, JIEYING BO, AND PIETER C. WENSINK*

Department of Biochemistry and Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02254

Communicated by David Botstein, July 21, 1986

ABSTRACT **We have determined the nucleotide sequences of all four *Drosophila* α-tubulin genes (α1, α2, α3, and α4). Two of the genes, α1 and α3, are constitutively expressed and code for proteins that are very similar to previously sequenced α-tubulins. They differ from each other by only two amino acid substitutions. These two genes also have blocks of homology between the noncoding leader regions of their transcription units. In contrast to these constitutive genes, the tissue-specific α2 and α4 genes code for tubulins with different structures. The α2 mRNA is male-specific in adults and codes for a tubulin that differs from α1 at 21 of the 450 residues. Six nonconservative substitutions are clustered within the 14 carboxyl-terminal amino acids, a region implicated in the regulation of microtubule assembly. The α4 mRNA is maternal and is found only in ovarian nurse cells, eggs, and early embryos. It codes for the most highly divergent α-tubulin yet reported and differs from α1 at 149 positions.**

Microtubules are the major structural elements of cilia, flagella, the cytoskeleton, and the mitotic and meiotic spindles. As might be expected from their presence in these structures, the disruption of normal microtubule assembly affects cell motility, division, and secretion, as well as intracellular transport (reviewed in ref. 1). This diversity in microtubular structure and function suggests that different microtubule subunits assemble into specialized microtubules (2). The basic subunit of all microtubules is a heterodimer of α- and β-tubulin polypeptides. This αβ subunit coassembles with species- and tissue-specific microtubule-associated proteins (MAPs) (3) to form microtubules *in vivo*. Thus, functional specialization should be reflected in the structures of the α- and β-tubulins, the MAPs, or both.

Studies of tubulin proteins and genes from a wide variety of species indicate that although tubulin structure is highly conserved among species, there is some minor variation between the tubulins found in a single organism. This variation is due to both posttranslational modification of the tubulin heterodimer (4, 5) and primary-structure differences encoded in multiple tubulin genes (reviewed in ref. 6). The degree to which primary-structure differences contribute to tubulin heterogeneity is only beginning to emerge, because complete sequence analysis of an entire tubulin gene family of a higher eukaryote has not been reported. The available sequence data indicate that multiple tubulin genes do code for structurally different tubulins. For example, among the known sequences of chicken β-tubulins the greatest difference is 8.7%. This clearly represents a minimum estimate of the range of differences in a single organism because only four of the nine or more β-tubulin genes from chicken have been analyzed.

The sequences of all the α- or β-tubulin genes from a higher eukaryote would indicate the entire range of tubulin primary structures utilized by the organism and facilitate molecular approaches to the study of tubulin specialization. In this paper we describe the sequence of the entire α-tubulin gene family of the higher eukaryote *Drosophila melanogaster*.

There are four α-tubulin genes in *Drosophila*, referred to as α1, α2, α3, and α4,† at chromosomal locations 84B3-6, 85E6-10, 84D4-8, and 67C4-6, respectively (8). The α1 and α3 genes appear to be constitutively expressed, whereas the other two genes have highly specialized developmental patterns of expression (9–11). In adults, α2 transcripts are found only in males, where they may be testes-specific (9). The α4 transcripts accumulate only in very early embryos and in adult female ovaries (9). *In situ* hybridization to ovarian tissue indicates that the α4 message is maternal and is synthesized in the nurse cells (M. Harris and P.C.W., unpublished data). This suggests that α1 and α3 may have functions, such as providing cytoskeletal structure, that are common to most cells, whereas α2 and α4 may provide tubulins with specialized functions.

In this paper we report the complete nucleotide sequences and the transcript maps of all four *Drosophila* α-tubulin genes. The constitutively and coordinately expressed α1 and α3 genes code for nearly identical proteins that are very similar to the known sequences of α-tubulins from other species. The two sex-specific α-tubulin genes, α2 and α4, code for quite different polypeptides that may be functionally specialized.

MATERIALS AND METHODS

Sequencing. DNA fragments containing *Drosophila* α-tubulin genes were excised from plasmids pDmTα1–4 (8, 12) and subcloned into plasmids pUC8 and pUC9 for sequence analysis. A series of deletions in the pUC subclones were prepared using the DNase I method of Hong (13), modified for use with double-stranded plasmid DNA. Plasmid DNA was prepared by the alkaline-lysis method (14). Deletions were sized by gel electrophoresis rather than by the sequencing protocol of Hong.

Single-stranded templates for dideoxy sequencing (15) were prepared by digesting plasmid DNA with a restriction endonuclease that cut within 2.0 kilobases downstream (relative to the direction of primer extension) of the primer site. The DNA was then precipitated with ethanol, washed in 70% (vol/vol) ethanol, dried, and resuspended in 50 μl of 70 mM Tris·HCl, pH 8.0/1 mM $MgCl_2$/10 mM dithiothreitol. This DNA was digested (16) with 25 units of exonuclease III at 37°C for 3 hr or at 25°C overnight. Digestions were terminated and protein was removed by three extractions

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.

Abbreviation: MAPs, microtubule-associated proteins.

*To whom reprint requests should be addressed.

†Recommended symbols (7) for these genes are *αTub84B*, *αTub85E*, *αTub84D*, and *αTub67C*.

with phenol/chloroform (1:1). This DNA was precipitated, washed, and dried as above and then resuspended in 15 μl of 10 mM Tris·HCl, pH 8.0/1 mM EDTA. The resuspended templates were sequenced by the dideoxy procedure (15), but using ^{32}P-end-labeled primer (universal sequencing primer, New England Biolabs). End-labeled primer was essential for generating unambiguous sequence data from these templates. Products of the reactions were electrophoresed in 6% polyacrylamide buffer-gradient sequencing gels (17).

Other Methods. DNA restriction fragments were radiolabeled using either bacteriophage T4 polynucleotide kinase or avian myeloblastosis virus reverse transcriptase as described (18). RNA isolation, plasmid purification, agarose gel electrophoresis, ligations, and bacterial transformations were done as described (14, 19). Restriction endonucleases and other DNA-modifying enzymes were purchased from Boehringer Mannheim and New England Biolabs. Radiolabeled nucleotides were obtained from ICN.

RESULTS AND DISCUSSION

Nucleotide Sequence and Transcript Maps of the *Drosophila* α-Tubulin Genes. The strategy of Hong (13), adapted for use with double-stranded plasmid vectors (see *Materials and Methods*), was used to generate deletions in subcloned fragments of the four α-tubulin genes. DNA neighboring the deletion breakpoints was then sequenced by the dideoxy method (15), resulting in the sequences shown in Figs. 1 and 2.

The intron/exon structures of the four genes are shown in Figs. 1 and 2. The structures of α1, α2, and α4 are similar. The first exon of each encodes only the first amino acid of the protein. In each case the codon for the first amino acid is located within a consensus splice-donor sequence and the

α1

```
GTCGACAGCTTGCCGTCTCTAGCTCCGGTGCCTATATAAAGCAGCCCGCTTTCCACATTTCATATTCGTTTTACGTTTGTCAAGCCTCATAGCCGGCAGTTCGAAC +45
GTATACGCTCTCTGAGTCAGACCTCGAAATCGTAGCTCTACACAATTCTGTGAATTTTCCTTGTCGCGTGTGAAACACTTCCAATAAAAACTCAATATGGTGAGTA +150
CTTTAAAAAAAAATCTAGTGAAATAATGCTGAAAAGAAATTTGTGTGGGCAAAATTCAATGGGCAAAAACGCGATGCGGCTTTTTCTCAAAATGGCGGCCGGCCTG +255
CGTTTTTTCCTCAAAAGTGATGACGTCATGCCTGTTTTTTTTTTTTTTGTTCGCAATGAGGAATGGCTCTTAAAATCTAGATAAAAAAAATATTCATTATTTCTAT +360
GCTGCTGGAACGCTTCATTAATCTTAAAAATTCTAAATTCGGTTACCATGATACTTCGACGCATAACTGTAGATTTTGGATAGAATTAAAGAGAAAATGGCGAGAG +465
AGTAAAATTCCGGCGTCGGCAAAGTAGAGCAAAAAAATCAGTATACCATTTAGCTACCTCTCTCACTCGCACGCAGTGCCGGCTCAAGTTGGGCGCGGCTCTGCAA +570
TTATCGATTTTCTTGGGGTGTGTAACTAATCATCCGTTTTCCCTTCCTCCTCATCCACAGCGTGAATGTATCTCTATCCATGTTGGTCAGGCTGGTGTCCAGATTG +675
GAAACGCCTGCTGGGAGCTCTACTGCTTGGAGCACGGCATCCAGCCCGATGGCCAGATGCCGTCTGACAAGACCGTGGGCGGAGGTGATGACTCGTTCAACACCTT +780
CTTCAGCGAGACTGGAGCTGGCAAGCACGTGCCCCGCGCCGTGTTTGTGGATCTGGAACCCACTGTGGTCGATGAGGTCCGTACCGGAACCTACCGTCAGCTGTTC +885
CACCCTGAGCAGCTGATCACTGGTAAGGAGGATGCGGCCAACAACTACGCCCGTGGCCACTACACCATCGGCAAGGAGATCGTCGATCTGGTTCTGGACAGGATCC +990
GCAAGCTGGCCGATCAGTGCACCGGTCTGCAGGGCTTCCTCATCTTCCACTCGTTCGGTGGAGGTACCGGCTCCGGCTTCACCTCGCTGCTGATGGAGCGTCTCTC +1095
CGTGGACTACGGCAAGAAGTCCAAGCTGGAGTTCGCCATCTACCCAGCCCCCCAGGTGTCCACTGCCGTGGTCGAGCCCTACAACTCCATCCTGACCACCCACACC +1200
ACCCTGGAGCATTCCGACTGCGCCTTCATGGTCGACAACGAGGCTATCTACGACATCTGCCGCCGCAACCTGGACATTGAGCGCCCCACGTACACCAACCTGAACC +1305
GTCTGATTGGCCAGATCGTGTCCTCGATTACCGCCTCTCTGCGATTCGATGGTGCCCTTAACGTGGATCTGACTGAGTTCCAGACCAACTTGGTGCCCTACCCACG +1410
TATTCACTTCCCTCTGGTGACCTACGCCCCCGTTATCTCCGCCGAGAAGGCCTACCACGAGCAGCTGTCGGTGGCTGAGATCACCAACGCCTGCTTCGAGCCGGCC +1515
AACCAGATGGTCAAGTGCGATCCCCGTCACGGCAAGTACATGGCCTGCTGCATGCTGTACCGCGGTGATGTTGTGCCCAAGGACGTCAACGCCGCTATTGCCACCA +1620
TCAAGACCAAGCGCACCATTCAATTCGTCGACTGGTGCCCCACTGGCTTCAAGGTTGGCATCAACTACCAGCCACCCACCGTGGTGCCTGGAGGTGATTTGGCCAA +1725
GGTGCAGCGTGCCGTGTGCATGTTGTCCAACACCACGGCCATCGCCGAGGCCTGGGCCCGTCTGGACCACAAGTTCGATCTGATGTACGCCAAGCGTGCCTTCGTC +1830
CACTGGTACGTTGGTGAGGGTATGGAGGAGGGAGAGTTCTCCGAGGCCCGTGAGGATTTGGCTGCCCTCGAGAAGGACTACGAGGAGGTCGGCATGGACTCCGGTG +1935
ACGGCGAGGGTGAGGGCGCTGAGGAGTACTAAGCGTCGCGCCACTTCAACGCTCGATGGGAGCGTCATTGGTGGGCGGGGTAACCGTCGAAATCAGTGTTTACGCT +2040
TCCAATCGCAACAAAAAATTCACTGCAACACTGAAAAGCATACGAAAACGATGAAGATTGTACGAGAAACCATAAAGTATTTTATCCACAAAGACACGTATAGCAG +2145
AAAAGCCAAGTTAACTCGGCGATAAGTTGTGTACACAAGAATAAAATCGGCCAGATTCAGTGTTGTCAGAAATAAGAAAACCCCACTATGTTTTTCTTTGCCTTTT +2250
```

α2

```
CGCGTCCACATTTCACATAACTTCCACATATAGCATGTAGCATATAGCATATGCTACATATGTCTGTCTTGGGGTCTTATAAAAGAACGCGTATCGTCAAGATTCG -4
GTTAGTTGACTCCCAGCTGTGTGTCGCATCACAGACCAGTGCTAGGATTTTCGTGATTGAATCTGCTTTTTTAAACTTCAGCTTAAACTTTTGATCATCATGGTAA +100
GGATAATAGGCACTTTCAATGATTTAGCTGGACTTTTGAAATTTCGTCGGCTGCACAGAACTATGGATCCTTAAAAGTCCTTATATGCGATTTCATGCAAAACTAG +205
ACCATATGGTGCAATATTTATTCATTTTTGCAAGTGATTAATTAATGGTTTTAAAAAATATAACTGCTTGTAAACTATTATCTTAAATGAATACCCAAGAACTTAC +310
AGCGATCAAAATTGTTTGTAGCATATGGTCCAAATCACAGATTTGCAACCGGAAGTCCAAAGTTAAATTTTAGAATTACCCAAATTTACAGCGACATGAGTCACAT +415
TATTATAAACTCTGTTCCAGAGGGAATGCATTTCGGTTCACATTGGCCAAGCTGGTGTCCAGATCGGCAATGCCTGCTGGGAACTTTACTGCCTGGAGCACGGCAT +520
CCAGCCCGATGGCCATATGCCCTCCGATAAAACCGTGGGCGGAGGCGATGACTCCTTCAGCACCTTCTTCAGCGAGACTGGAGCTGGTAAGCATGTGCCGCGTGCC +625
GTGTTCGTGGATCTGGAGCCCACTGTGGTGGATGAGGTGCGGACAGGAACCTACCGCCAGCTGTTCCATCCGGAGCAACTGATCACCGGTAAGGAGGATGCGGCCA +730
ACAACTATGCCCGCGGCCACTATACCATCGGTAAGGAGATCGTCGACGTGGTCCTGGACAGGATTCGCAAGTTGGCGGATCAGTGCACGGGTCTGCAGGGATTCCT +835
GGTCTTCCACTCCTTCGGCGGCGGCACTGGCTCTGGATTTACGTCGCTGCTGATGGAGCGCCTGTCCGTGGACTATGGAAAGAAGTCAAAGCTGGAGTTCTCCATT +940
TACCCAGCACCACAAGTATGTCATTTGCATATGGATATTATATCTATTGCTTGAGCATTTATAGAAGACCAATTAACATAATTAGTAAAAGACATGGGAAATTAGT +1045
CTTCGAAAGGCAATCTAATAATCGAACTTCCTGTTAGGTGTCCACAGCTGTGGTCGAGCCATACAACTCGATCTTAACCACCCACACAACGCTGGAGCACTCGGAC +1150
TGTGCGTTTATGGTTGACAACGAGGCAATCTACGACATCTGCCGGCGCAACTTGGATATCGAGCGACCCACCTACATGAATCTCAATCGCCTAATTGGCCAGATCG +1255
TTTCGTCCATTACGGCCTCGTTGCGCTTCGATGGCGCCCTGAATGTGGATCTGACCGAGTTCCAGACCAATTTGGTGCCCTACCCACGCATTCATTTCCCACTGGC +1360
GACATACGCTCCCGTCATTTCCGTGGAGAAGGCCTACCATGAGCAACTGACCGTGGCCGAGATCACCAATGCCTGCTTCGAGCCGGCCAACCAGATGGTCAAGTGT +1465
GATCCGCGTCGTGGCAAGTACATGGCCTGCTGCATGCTCTACCGCGGTGATGTGGTGCCCAAGGATGTGAACGCAGCCATTGCCACCATCAAGACCAAGCGCTCTA +1570
TCCAGTTCGTCGACTGGTGTCCCACGGGCTTCAAGGTGGGCATCAACTATCAGCCACCCACCGTTGTTCCTGGCGGAGATCTGGCCAAGGTGCAGCGCGCCGTCTG +1675
CATGCTGTCCAATACCACTGCCATTGCCGAGGCCTGGGCCCGTCTGGATCACAAGTTCGATCTGATGTACGCCAAGAGGGCCTTCGTCCACTGGTACGTGGGTGAG +1780
GGCATGGAGGAGGGCGAGTTCGCCGAGGCTCGCGAGGATCTCGCTGCCCTCGAGAAGGACTACGAGGAGGTAGGCATCGACTCCACCACCGAACTGGGGGAAGATG +1885
AGGAATACTAGAGAAAACGTTGTCAACAAATGATACAGCCATTACAACTTTATAAACATCACTTTCTTTCTACCAAATACCAATGCATAAAACAGTTTCGCAATAA +1990
AACAGTTTAACACAATTGCTTAAAATGGCTTATTTAAATGGGAAATATTCGCTTTAGAT
```

FIG. 1. Nucleotide sequences and transcript maps of the α1- and α2-tubulin genes of *Drosophila*. Exons are underlined. The "TATA boxes," the first and last codons, and the polyadenylylation consensus sequences are overlined. Position +1 is the transcription start site. The intron/exon structures of the four genes were predicted by localizing open reading frames coding for proteins homologous to porcine brain α-tubulin (20) and by identifying regions that match consensus sequences found at initiation, termination, and processing sites (21). In ambiguous cases, the exon boundaries were determined by nuclease protection, primer-extension, or RNA blot hybridization analysis and are described in detail elsewhere (11, 22). Blot hybridization analysis of electrophoretically fractionated RNA localizes the first exon of α2 between −325 and +415. In the figure, this exon is positioned according to the best fit to consensus sequences for the TATA box and the donor splice junction within this region.

α3

```
CTTGAAATAGGCATTATTAATTGTGTTAATATGCAGTTCACGTTATACAAGGCATTGTTATTATTGAACTGTTTGGTATATTTAAATTTCCCACCGGCGGTCACA +3
CAAAATCTACTGAACAAGCGTGCTTGTGCTAAATAATTTTACTGATCTTTAAATCTCGTAAATACCGCCATTTAGCCGCATTGCTGAACACTCTTGCCGGTGGTA +108
ATCAATTATTGGCCAAGAGCTAGCAATATTGTGTAAAAATCAGCGTGAAACTAAAGCCGAATTCAAACAAACAAACCAAAAACCGGCCTGACAGGGTTTTTCGTG +211
CCACCAGAGAAAATCGATTGGTAAAAGAAACGCGTTTCGTATAATATAGATAGTTATCTAATATCTAATCTAATATCTGAGTTGTTGTGAAGCCCTTTTTCGTAA +316
AACAGACGTACTGGCGATTTACGTCTGTTCGAAATCCATTCCCGGGCATTCCAACGTATACGCCCGCAGAGTCAGACCCCAAAATCCGTAGGCTTACACTTTTTC +421
CGTTGATTTTCCGTAGTCGCGTGTGAGAAACCGAAACTCGAAACAAGAAACCCGGAATCGGAAAACCCCATTAAAAATCAATATGCGCGAATGTATCTCTATCCAC +526
GTGGGCCAGGCCGGAGTCCAGATTGGAAATGCCTGCTGGGAGCTCTATTGCCTGGAGCACGGCATCCAGCCCGATGGCCAGATGCCCTCCGACAAGACCGTAGGC +631
GGAGGTGATGACTCGTTTAACACCTTCTTCAGCGAGACTGGAGCTGGCAAGCACGTGCCTCGCGCCGTGTTCGTGGATCTGGAGCCCACTGTGGTCGACGAGGTC +736
CGTACTGGAACCTACCGTCAGCTGTTCCACCCCGAGCAACTGATCACCGGTAAGGAGGATGCGGCCAACAACTACGCCCGTGGCCACTACACCATTGGCAAGGAG +841
ATCGTCGACCTGGTGCTGGACAGGATCCGCAAGCTGGCCGATCAATGCACCGGTCTGCAGGGCTTCCTCATCTTCCACTCGTTCGGTGGAGGTACCGGCTCCGGC +946
TTCACCTCGCTGCTGATGGAGCGTCTCTCCGTGGACTACGGCAAGAAGTCCAAGCTGGAGTTCGCCGTGTACCCAGCTCCGCAGGTGTCTACCGCCGTGGTGGAG +1051
CCCTACAACTCCATCCTAACCACGCACACCACCCTCGAGCACTCCGACTGCGCCTTTATGGTGGACAACGAGGCCATCTACGACATCTGTCGCCGAAACCTGGAC +1156
ATCGAACGACCCACTTACACTAACCTGAATCGTTTAATCGGCCAAATAGTGTCCTCGATCACCGCCTCTCTGCGATTCGATGGAGCCCTAAATGTGGACCTCACC +1261
GAGTTCCAGACCAACCTGGTTCCTTACCCCCGCATCCACTTCCCGCTGGTGACGTACGCTCCCGTCATCTCCGCGGAGAAGGCCTACCACGAGCAGCTGTCGGTG +1366
GCCGAGATCACCAACGCCTGCTTCGAGCCAGCCAACCAGATGGTCAAGGTGGATCCCCGTCACGGCAAGTACATGGCCTGCTGCATGCTGTATCGCGGCGATGTG +1471
GTGCCCAAGGACGTGAACGCCGCCATCGCCACCATCAAGACAAAGCGTACCATTCAGTTCGTGGACTGGTGCCCCACCGGCTTCAAGGTGGGCATCAACTACCAG +1576
CCACCCACTGTCGTCCCAGGCGGGGATCTGGCCAAGGTGCAGCGTGCCGTGTGCATGTTGTCCAATACCACGGCTATTGCCGAGGCCTGGGCCCGTCTGGACCAC +1681
AAGTTCGATCTGATGTACGCCAAGAGGGCCTTCGTCCACTGGTATGTTGGTGAGGGCATGGAGGAGGGCGAGTTCTCCGAGGCCCGTGAGGATCTGGCTGCCCTC +1786
GAGAAGGACTACGAGGAGGTCGGCATGGATTCCGGTGACGGTGAGGGCGAGGGTGCTGAGGAGTACTAGAGCGGATAGGTGTGCATCCCTCAAAATCTTGCTTAT +1891
TTTCATACGCTCCCACTCGCATGTATAAATTTTCTGCGATAATGTATAGCGAACAAAGAAGAAAATAAAACCCGAGAAAACACCAATTTATTAACCACGTTTAG +1996
```

α4

```
TCAATATGATATTATATAGTTTTTAATAATAATTTAAGTAAAGGTAAAGCAAATCCTGAGATACATACATACTATTTCAAATTAGTATCGATAAGCCGCCAGAAAT -83
GGGTTGGAAAAACCTTGGAATGATTATTGGTTAGCAGCCCTGTCTCCCAATTTGACATTCCTCTTTGGCGCTCTGTAACGGCGCTTTTCCAGTGACTCCCCAATCA +23
GCGAACCAACTAACCTGAGGCTGCACTTTTCGCGTGGCGACCAAAAAAATTATAAAAAAAAGGTCCCGCATTAGATTTTAATTTGTATACACCCAATTTTAGTCCA +129
GCAAAAAAAAACTAATAAAATGGGTGAGAAGAAGAAAAACTAATTAGTTTAAGAATTGCTGAAAGCGGGAAGTGAAATTTAAAATCAGAAAAGATGTGTAAGAAGAT +235
GTGAAATGAAAAGGTGTGGGAAGTGAACATGAATAAATGTTCATTGAAAATGAACAGCTAATTGCGAATTTTTTGAAACTTAACTGTTAACTGTTACTGCTCACAA +341
ATCCTTGAAAAATATTAAGATCTTAGAATTGTTTATTTAAAATGAATTCCTCGTATGTATTATAACAGACAATCAAGTTTATTAAAGAATATTTAAACTTGTATTT +447
TTTTTTGAGTTTGATAAAATGTCCCAAATATAAAAGAGTATTACCTTTTCGTCATATAATACTGTTCTTAATATTAGTTTTAGTTAGTAATGTGTTTATAAAGTAA +553
ATTACAAAGGGGAGGACCTAACCTAACCTTAAACAATGGCGCCTTCTCCTTGCCCAGCTCATCACTTTTATTGTTTCACCCGCCAGCGCGAAGTAGTGTCCATCCA +659
GATTGGCCAGTGCGGCATCCAGATCGGCAATGCCTGCTGGGAGCTGTACCTGCTGGAGCACGGCATCAACCTGGACGGCAGCCTGAAGACCAAGGAGGAGCTAACG +765
GCCAGCGGGAGCAGTGCCAGTGTGGGTCACGACACCTCGGCCAACGATGCTAGGACCTTCTTCACGGAAACCGGCAATGGAAAACAAGTGCCACGCTCGATTTTCG +871
TCGATCTGGAACCGACGGTCATCGATGATGTGCGGAATGGCTGCATGAGAGAGCTCTATCATCCGGAGCAACTGATTTCTGGAAAGGAAGATGCGGCCAATAACTA +977
TGCCCGAGGTCGTTACTCCATCGGCAAGGAGGTGATCGATAGGGTGACCTCACGGCTGCAGAAGATCGCAGAGCAGTGCGACAGCTTGCAGGGATTCCTCATCTTC +1083
CACTCGCTGGGCGGTGGTACTGGTTCGGGATTCACCTCCCTGTTGGTGGAGCGCTTGTCCACCGATTACAGCAAGAAGTGCAAGCTGGACTTTGCCGTCTATCCAT +1189
CGCCTAAGGTCTCCACCGCCGTGGTGGAGCCATACAACGCGCTGCTCACCACCCACTCCACCATGGACCACTCGGACTGTGTGTTCATGGTGGATAACGAGGCCAT +1295
CTATGATATATGTAACAATAGCTTGGGTGTGGACAGGCCCGCCTATAGGAATCTGAATCGTTTGATCGCCCAAATAGTGAGCTCCACAACGGCTTCTCTGCGTTTC +1401
AGTGGCTCCATGAATGTGGATCTGAATGAGTTCCAGACGAATCTAGTGCCCTTCCCCAGAATCCACTTTCCCCTGGTTGCCTATGCCCCACTGATGTCCGCCGAGA +1507
GATCGGCCCATGAACAGCATGCGATTACCACATTGACCAATGCCTGTTTCGAATCCTCCAACATGATGGTAAAGTGTGATCCTCGTGCGGGCAAATTCATGGCCTG +1613
TTGCATGCTTTACAGGGGTGATGTGGTGCCCAAGGATGTAAATGCCGCTGTCTCGGCCATCAAGTCCAAGCGGCACATTCAATTCGTGGACTGGTGTCCCACTGGT +1719
TTCAAGATTGGCATCAACTACGAGAAGCCAGCCTTTGTGCCGGATGGAGATTTGGCCAAGACCTCGAGGGCCTGCTGCATGCTGTCCAACACCACCGCCATCTCGG +1825
TGGCCTTCTCCAATCTCTCCTACAAGTTCGATCTGATGTTCAAGAAGCGGGCCTTCGTCCATTGGTACGTGGGCGAGGGCATGGAGGAGGGCGAGTTCACGGAGGC +1931
GCGCGAGAATATCGCCGTACTGGAACGGGACTTCGAGGAGGTCGGTCTGGACAATGCCGAGGAGGGCGGCGACGAGGACTTCGATGAGTTCTAAGCCCACACCGCA +2037
ACAATCCCAGCCAGACTTCAGTCATTACAATATCCAATAATTATGTGTAAATAAATTAACCAAGAGTTTGAATCATTATGAGTTTCATTGAACAAGGGATGAGGAT +2043
```

FIG. 2. Nucleotide sequences and transcript maps of the α3- and α4-tubulin genes of *Drosophila*. Features of the sequence are marked as described in the legend to Fig. 1.

codon for the second amino acid is within a consensus splice-acceptor sequence. The remainder of the protein is encoded by either one (α1 and α4) or two exons (α2). Both rat (23) and human (24) α-tubulin genes have been found that have introns located between the first and second amino acid codons. The α3 gene, unlike the other three, has a single exon coding for the complete protein.

Nucleotide Sequence Homologies. As expected, the four genes are homologous in their protein-coding regions. However, the degree of homology varies. Genes α1 and α3 are the most closely related members of the gene family, having 92% nucleotide sequence homology within the protein-coding regions. The coding sequences of the α2 and α4 genes are, respectively, 87% and 68% homologous to α1.

The coordinately expressed α1 and α3 genes are the only members of the family that have extended regions of homology outside the coding sequences. The regions of homology are in the untranslated leader portions of the α1 and α2 transcripts (Fig. 3). The homologous sequences begin immediately upstream of the first codon and extend upstream, lying between nucleotides +27 and +147 of α1 and nucleotides +354 and +503 of α3.

```
  +27                                                    +87
5'CATAGCCGGCAGTTCGAACGTATACGCTCTCTGAGTCAGACCTCGAAAT_CGTA_GCTCTACA 3'   α1
   |  | |||| ||| ||||||||||| | | |||||||||| | |||| |||| ||| ||||
  ATTCCCGGGCA_TTCCAACGTATACGCCCGCAGAGTCAGACCCCAAAATCCGTAGGCT_TACA       α3
 +354                                                    +414

                                                    +143
  C_AATTCTGTGAATTTTCC_TTGTCGCGTGTGAAACACTTCCAATAAAAACTCAATATG 3'        α1
  |   ||| ||  ||||||| | ||||||||||| |      || || |||| ||||||||
  CTTTTTCCGTTGATTTTCCGTAGTCGCGTGTGAGA....CCCCATTAAAAATCAATATG           α3
                                +447 +484               +503
```

FIG. 3. Homology between the untranslated leader regions of α1 and α3 messages. Homology between α1 (upper line) and α3 is indicated by vertical lines. Gaps have been introduced to maximize homology. Sequences are numbered as described for Fig. 1. The initiation codon (ATG) is underlined.

The conservation of sequences in the 5′ leaders of the two coordinately expressed genes suggests a regulatory function for the leader region. Tubulin mRNA levels appear to be subject to feedback regulation in a number of cultured cell lines (25, 26). Drugs that promote depolymerization of microtubules and increase the tubulin subunit pool size also cause a decrease in the steady-state levels of α- and β-tubulin mRNA. This regulatory mechanism appears to act posttranscriptionally (27, 28). We speculate that the homologies in the 5′ leader sequences of α1 and α3 are involved in a posttranscriptional regulatory circuit, perhaps interacting with cytoplasmic factors to reduce stability of the message. It is also possible that these sequences are involved in transcriptional regulation of the two genes.

Amino Acid Sequence Homologies and Heterogeneities. The sequences of the four genes were used to predict the primary structure of the *Drosophila* α-tubulins (Fig. 4). The α1- and α3-tubulins are nearly identical and are very similar to the known sequences of α-tubulins from other species. However, the two tissue- and sex-specific α-tubulin genes, α2 and α4, code for clearly different polypeptides.

The α1 and α3 proteins differ by two substitutions. Isoleucine-170 and cysteine-305 of α1 are replaced by valine residues in α3 (Fig. 4). These substitutions do not cause differences between the predicted secondary structures (Fig. 5), indicating that the products of these two genes are likely to have identical functions. They resemble previously characterized α-tubulins. As an example, α1 differs at 16 of 450 residues relative to a porcine α-tubulin, and of those 16 differences, only 6 are nonconservative (Fig. 4).

The sequence of the α2 protein differs from α1 at 21 of 450 residues (Fig. 4). While most of these are conservative differences, a few do change the predicted secondary structures (Fig. 5). The most striking cluster of nonconservative substitutions is at the carboxyl terminus of the protein, between residues 439 and 450 (Fig. 4).

The α4 protein is clearly different from the other *Drosophila* α-tubulins and from the sequenced α-tubulins of other species (Fig. 4). The α1 and α4 proteins differ at 149 residues. Such high divergence has not been observed between any two previously described α-tubulins. Substitutions are found throughout the polypeptide, with two major clusters near the amino and carboxyl termini, respectively. The amino-terminal cluster lies between residues 34 and 61, and includes an 11 amino acid insertion in α4 relative to other α-tubulins. The α1 tubulin from *S. pombe* has a 4 amino acid insertion at this same site (29). The predicted secondary structure of the protein is quite different from those of the

```
D1 MRECISIHVGQAGVQIGNACWELYCLEHGIQPDGQMPSDKTVGGGDDSFN----------TFFSETGAGKHVPRAVFVDLEPTVVDEVRTGTYRQLFHP 89
D2       V I                        H              S
D3
D4    VV  QI  C I          L     NL  SLKTKEELTASGS ASVGHDTSANDAR   T   N  Q   SI        I D  N CM E Y
Y1    V  V                        G   FPTENSE HKNNSYL DGFG              Q  F   SIY     N I Q    P KD
Y2    I         T                   N Y NP TASQNS GG S                   Q  Y   SIY     N I Q   DP  D
P                                        I                                                 I

D1 EQLITGKEDAANNYARGHYTIGKEIVDLVLDRIRKLADQCTGLQGFLIFHSFGGGTGSGFTSLLMERLSVDYGKKSKLEFAIYPAPQVSTAVVEPYNSIL 189
D2                              V                  V                                  S
D3                                                                                    V
D4     S            R S    VI R TS LQ I E  DS         L            V    T  S  C  D  V  S K          AL
Y1   MV       S         V   MI S  E   RM  N S      V           LGT  L   NME     N Q SV        S       V
Y2           S         V   L  E T K  RIG N S      V            GA  L   AME T     Q SV        S       V
P                             I                   SV                                  S

D1 TTHTTLEHSDCAFMVDNEAIYDICRRNLDIERPTYTNLNRLIGQIVSSITASLRFDGALNVDLTEFQTNLVPYPRIHFPLVTYAPVISAEKAYHEQLSVA 289
D2                                      M                                                A       V        T
D3
D4    S MD    V            NNS GVD  A R      A     T      S SM    N        F        A   LM   RSA   HAIT
Y1    A  DN   T       C              E      A V          A S     N                   S IV  A  F  SN  Q
Y2    A  DLA V   L       L KKS   D  S A     VA VI  L T    D A    IN              N       IV  A  F  SN  Q
P                                                                                   A

D1 EITNACFEPANQMVKCDPRHGKYMACCMLYRGDVVPKDVNAAIATIKTKRTIQFVDWCPTGFKVGINYQPPTVVPGGDLAKVQRAVCMLSNTTAIAEAWA 389
D2                    R                                 S
D3                V
D4 TL      SS M       A  F                    VSA  S  H            I    EK AF  D     TS  C          SV FS
Y1 Q   Q    Y         T R   T L      I R  Q  VTS  SR               I  C E  QH   SGI   N          S     S
Y2 Q   Q    Y         A R   T L      I R  Q  VT   A                I  CDR  QHIE SEI   D          S     S
P                              L                                        E

D1 RLDHKFDLMYAKRAFVHWYVGEGMEEGEFSEAREDLAALEKDYEEVGMDSGDGEGEGA-EEY 450
D2                                A                 I  TTEL  D-
D3
D4 N SY     FK                    T    NI V  R F    L NAEEG DEDFD F
Y1           S                               R      Q  M N MYE D
Y2           S                               R      Q  MEVDYME-
P                                      M            V  VE    EEG
```

FIG. 4. Comparison of amino acid sequences of α-tubulins. The species abbreviations are D1, D2, D3, and D4 for *D. melanogaster* α1, α2, α3, and α4 proteins, respectively; Y1 and Y2 for *Schizosaccharomyces pombe* (NDA2)1 and (NDA2)2, respectively (29); and P for porcine brain α-tubulin (20). The amino acids are written from left to right and top to bottom in an amino-terminal to carboxyl-terminal direction. Standard one-letter abbreviations are used. Only differences from the α1 sequence are shown. A dash represents the position of an insertion relative to α1.

FIG. 5. Secondary structure predictions for *Drosophila* α-tubulins. Chou and Fasman (30) predictions are diagrammed. α-Helical regions are represented by solid boxes; β-sheets, by cross-hatching; β-turns, by hatching; and unpredicted structures, by open boxes. The dashed lines in α1, α2, and α3 are the positions of deletions relative to α4.

other *Drosophila* α-tubulins (Fig. 5). There are, however, regions of predicted secondary structure that are shared by all four α-tubulins. Many of the α-helices and β-turns predicted for the α1, α2, and α3 proteins have corresponding structures in α4-tubulin (Fig. 5). This analysis indicates that the α4 protein, although clearly different, is similar in length, shares most of the sequence, and may fold to form a protein with many of the structural features of other α-tubulins.

The highly divergent α4 polypeptide ends with a carboxyl-terminal phenylalanine residue instead of the tyrosine found at the termini of all other α-tubulins sequenced to date. The terminal tyrosine of α-tubulins is involved in a novel posttranslational modification. This residue can be removed from and religated to α-tubulin by a specific carboxypeptidase and tyrosine ligase (4). This modification may be involved in regulation of microtubule function (31). The carboxypeptidase and ligase that catalyze this cyclic reaction have been found to use phenyalanine as an alternate substrate, indicating that α4-tubulin may undergo a similar modification.

One cluster of differences between the constitutive and developmentally regulated genes lies at the carboxyl terminus of the proteins. This region has been implicated in regulation of microtubule assembly and interactions with MAPs (32, 33). It is possible that this region interacts with specific MAPs to form specialized microtubular structures. This hypothesis is particularly attractive for α2-tubulin. In adult flies the α2 gene is expressed only in males, where it may be testes-specific (9). The testes are the only tissue in the fly that are known to assemble flagella. This correlation suggests that the α2 polypeptide may represent a testes-specific tubulin that is selected for assembly into the flagellar axoneme.

The tissue-specific distribution of the α4 transcript suggests that the α4 polypeptide may also have specialized functions. The α4 polypeptide is the product of a maternal transcript that is detected only in ovarian nurse cells and 0- to 3-hr embryos (ref. 9; M. Harris and P.C.W., unpublished data). During the first 2.5 hr of embryonic development, nuclei are undergoing 13 synchronous divisions without the formation of cell membranes (34, 35). These divisions are extremely rapid, averaging 9 min each. Concurrent with some of these nuclear divisions are the movement of nuclei to the periphery of the embryo, contraction of the yolk sac, and nuclear elongation. All of these events require microtubules (35) and appear to be unique in the life cycle of *Drosophila*. Some or all of these microtubule-dependent events may require the divergent α4 polypeptide.

The two tissue- and sex-specific α-tubulins differ from each other and from the constitutively expressed α-tubulins. The divergence of the α4 gene product is particularly striking. It is unclear whether the degree of heterogeneity observed within the *Drosophila* α-tubulin gene family reflects the unique requirements of this species or the incomplete nature of the structural data on tubulin gene families of other eukaryotes.

We speculate that the tissue-specific α-tubulin genes code for functionally specialized proteins. While this is an attractive hypothesis, it clearly need not be the case. Genetic analysis of the testes-specific β-tubulin of *Drosophila* clearly demonstrates that this tissue-specific protein participates in assembly of several different microtubule-based structures (36). The nucleotide sequences of all four α-tubulins of *Drosophila* should facilitate direct analysis of the regulation and function of the members of this gene family.

We thank Chan Fulton, Michael Wormington, and Paul Mitsis for their critical comments on the manuscript. We are grateful for the financial support of the National Institutes of Health (Grants GM21626 and GM31234). W.E.T. is the recipient of a Gillette predoctoral fellowship.

1. Dustin, P. (1984) *Microtubules* (Springer, New York).
2. Fulton, C. & Simpson, P. A. (1976) in *Cell Motility*, eds. Goldman, R., Pollard, T. & Rosenbaum, J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 987–1005.
3. Vallee, R. B., Bloom, G. S. & Theurkauf, W. E. (1984) *J. Cell Biol.* **99**, 38s–44s.
4. Flavin, M. & Murofushi, H. (1981) *Methods Enzymol.* **106**, 223–237.
5. L'Hernault, S. W. & Rosenbaum, J. L. (1985) *Biochemistry* **24**, 473–478.
6. Cleveland, D. W. & Sullivan, K. F. (1985) *Annu. Rev. Biochem.* **54**, 331–365.
7. Treat-Clemons, L. G. & Doane, W. W. (1984) in *Genetic Maps 1984*, ed. O'Brien, S. J. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), Vol. 3, pp. 311–323.
8. Kalfayan, L. & Wensink, P. C. (1981) *Cell* **24**, 97–106.
9. Kalfayan, L. & Wensink, P. C. (1982) *Cell* **29**, 91–98.
10. Mischke, D. & Pardue, M. L. (1982) *J. Mol. Biol.* **156**, 449–466.
11. Natzle, J. E. & McCarthy, B. J. (1984) *Dev. Biol.* **104**, 187–198.
12. Baum, H. J., Livneh, Y. & Wensink, P. C. (1983) *Nucleic Acids Res.* **11**, 5569–5587.
13. Hong, G. F. (1982) *J. Mol. Biol.* **158**, 539–549.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Smith, A. J. H. (1979) *Nucleic Acids Res.* **6**, 831–848.
17. Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3963–3965.
18. Hung, M. C., Barnett, T., Woolford, C. & Wensink, P. C. (1982) *J. Mol. Biol.* **154**, 581–602.
19. Schleif, R. F. & Wensink, P. C. (1981) *Practical Methods in Molecular Biology* (Springer, New York).
20. Ponstingl, H., Krauhs, E., Little, M. & Kempf, T. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2757–2761.
21. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
22. Baum, H. J. (1985) Dissertation (Brandeis Univ., Waltham, MA).
23. Lemischka, I. & Sharp, P. A. (1982) *Nature (London)* **300**, 330–335.
24. Hall, J. L. & Cowan, N. J. (1985) *Nucleic Acids Res.* **13**, 207–223.
25. Ben Ze'ev, A., Farmer, S. R. & Penman, S. (1978) *Cell* **17**, 319–325.
26. Cleveland, D. W., Lopata, M. A., Sherline, P. & Kirschner, M. W. (1981) *Cell* **25**, 537–546.
27. Cleveland, D. W. & Havercroft, J. C. (1983) *J. Cell Biol.* **97**, 919–924.
28. Caron, J. M., Jones, A. L., Rall, L. B. & Kirschner, M. W. (1985) *Nature (London)* **317**, 648–651.
29. Toda, T., Adachi, Y., Hiraoka, Y. & Yanagida, M. (1984) *Cell* **37**, 233–242.
30. Chou, P. Y. & Fasman, G. D. (1978) in *Advances in Enzymology*, ed. Meister, A. (Wiley, New York), Vol. 47, pp. 45–148.
31. Gundersen, G. G., Kalnoski, M. H. & Bulinski, J. C. (1984) *Cell* **38**, 779–789.
32. Serrano, L., Avila, J. & Maccioni, R. B. (1984) *Biochemistry* **23**, 4675–4681.
33. Serrano, L., De La Torre, J., Maccioni, R. B. & Avila, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 5989–5993.
34. Zalokar, M., & Erk, I. (1976) *J. Microsc. Biol. Cell* **25**, 97–106.
35. Foe, V. E. & Alberts, B. M. (1983) *J. Cell Sci.* **61**, 31–70.
36. Kemphues, K. J., Kaufman, T. C., Raff, R. A. & Raff, E. C. (1982) *Cell* **31**, 655–670.