Structure of two human α -tubulin genes

(microtubules/heteroduplex/inverted repeat)

C. DEBORAH WILDE*, LOUISE T. CHOW[†], FRANKLIN C. WEFALD^{*}, AND NICHOLAS J. COWAN^{*}

*Department of Biochemical Sciences, Princeton University, Princeton, New Jersey 08544; and [†]Cold Spring Harbor Laboratory, P.O. Box 100, Cold Spring Harbor, New York 11724

Communicated by Walter Kauzmann, September 22, 1981

ABSTRACT The ability of a chicken α -tubulin cDNA probe to cross-hybridize with human DNA under stringent conditions has been exploited to screen two independently constructed human genomic libraries. Nine clones were isolated, accounting for 60% of the bands observed in a whole genomic Southern blot of human DNA. Two clones were selected for further analysis by restriction mapping, orientation experiments using 3'- or 5'specific probes, and electron microscopy of heteroduplexes. One clone, 2α , contains an α -tubulin-specific region of 5.0 kilobases that includes three intervening sequences. The second clone, 19α , contains an α -tubulin-specific region of 5.4 kilobases and has somewhat diverged 5' and 3' ends. Clone 19α has only two intervening sequences that correspond to the first two in clone 2α . However, these intervening sequences differ in size between clones 2α and 19α and show no detectable sequence homology. The sum of the lengths of sequences in either clone that hybridize to the cDNA probe accounts for essentially the entire length of the cDNA molecule.

A remarkable feature of microtubules is the diversity of eukaryotic cellular functions with which they are associated. Included in such functions are mitosis, cell motility, intracellular transport, and secretion. The principal components of microtubules are two soluble proteins named α - and β -tubulin, each with a molecular mass of $\approx 55,000$ daltons and each encoded by a separate gene or genes (1). Several observations suggest that the sequence of tubulin proteins is highly evolutionarily conserved. Hybrid microtubule structures may be formed by *in vitro* copolymerization of α - and β -tubulins from different species (2), NH₂-terminal sequence data indicate considerable homology between sea urchin and chicken tubulins (3), and cDNA probes constructed from chicken α - and β -tubulin mRNAs are able to cross-hybridize with genomic DNAs from a variety of eukaryotic species under stringent conditions (4).

How many genes encode α - and β -tubulins, and what factors govern their expression? One approach to this question involves a direct examination of the genes themselves. The existence of multiple tubulins is implied by several reports of electrophoretic microheterogeneity in α - and β -tubulin proteins (5–8) and has been recently confirmed by protein sequence data that show multiple closely related forms of porcine α -tubulin (9). To isolate genomic sequences containing human α -tubulin genes, we have exploited the ability of a chicken α -tubulin cDNA probe to cross-hybridize with corresponding sequences from other eukaryotic species (4). We chose to study human tubulin gene organization because of the possibility that defective gene structure or expression might underlie certain disease conditions (10). We previously reported the isolation and characterization of several human β -tubulin genes (11). This paper reports the structure of two human α -tubulin genes.

MATERIALS AND METHODS

Screening of Libraries, Restriction Mapping, and Southern Blot Analysis. The partial EcoRI (12) and partial Hae III/Alu I (13) libraries were screened (14) as described (15, 16) in accordance with National Institutes of Health guidelines under P2/EK2 containment conditions. Restriction maps were generated by digestion of purified bacteriophage DNA with restriction endonucleases (New England BioLabs) under conditions suggested by the manufacturer. After electrophoresis on 0.7% agarose gels, DNA was transferred to nitrocellulose sheets (17) and the blot was hybridized with ³²P-labeled (18) α - or β tubulin cDNA probes to identify bands containing tubulin-specific sequences.

Electron Microscopic Heteroduplex Analysis. Desired restriction fragments of 2α and 19α were recovered from lowmelting-temperature agarose gels by extraction with phenol. pT1 DNA was linearized with *Eco*RI and extracted with phenol. The 2α DNA restriction fragment and pT1 DNA were denatured with alkali, neutralized, and then renatured in 70% formamide/TE buffer (0.1 M Tris·HCl/0.01 M EDTA, pH 8). Hybridization was carried out at 30–45°C for 15–20 min at DNA concentrations of 5 and 2.5 μ g/ml each. To prepare $19\alpha/pT1$ heteroduplexes, renaturation was carried out in 50–70% formamide/TE buffer at 22–30°C at similar DNA concentrations for 15–30 min. The preparation of electron microscopic grids and data analysis were carried out as described (11).

RESULTS

Isolation of Clones Containing α -Tubulin Sequences. Previous experiments (4) have shown the ability of cloned chicken tubulin cDNAs to hybridize under stringent conditions to human genomic DNA. To isolate human α -tubulin sequences, two independently generated human genomic libraries (12, 13) were screened using the chicken α -tubulin cDNA probe (clone pT1) labeled with ³²P by nick translation. Screening of $\approx 10^6$ bacteriophage from each library yielded 11 different clones that hybridized strongly with the chicken α -tubulin cDNA probe. Bacteriophage yielding positive hybridization signals were purified, and the isolated DNAs were examined by agarose gel electrophoresis after digestion with *Eco*RI. The results are shown in Fig. 1A. *Eco*RI fragments containing strongly hybridizing sequences, identified after transfer to nitrocellulose (17), are shown in Fig. 1B.

To determine the proportion of genomic α -tubulin sequences isolated by our screening experiments and to ascertain that the cloned fragments did not undergo rearrangement, an *Eco*RI digest of total DNA from human placenta was prepared, resolved by agarose gel electrophoresis, the fragments transferred to nitrocellulose, and the blot analyzed for tubulin-like sequences using the ³²P-labeled large *Pst* I fragment from clone

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: kb, kilobase pair(s).



FIG. 1. (A) EcoRI digests of DNA purified from positively hybridizing bacteriophage. DNA isolated from plaque-purified bacteriophage was digested with EcoRI. The fragments were resolved on a 0.9% agarose gel and visualized by staining with ethidium bromide. M, molecular size markers; group I, clones isolated from the partial EcoRI library (12); group II, clones isolated from the partial Hae III/Alu I library (13). (B) Southern blot prepared from the gel shown in A. After transfer, the nitrocellulose was hybridized with 5×10^6 cpm of clone pT1 (4) labeled by nick translation (18). (C) Southern blot of an EcoRI digest of human genomic DNA. A sample (10 μ g) of human placental DNA was digested with EcoRI. Fragments were resolved on a 0.8% agarose gel and transferred to nitrocellulose, and the blot was hybridized with 4×10^7 cpm of the large Pst I fragment (derived from clone pT1) labeled by nick translation. Unlabeled DNA markers were run in an adjacent lane.

pT1 (19) (see Fig. 3). The result (Fig. 1C) shows the presence of \approx 15 bands, 9 of which correspond in size to *Eco*RI fragments contained in recombinant bacteriophage from one (or both) of the human genomic libraries screened. The isolated clones thus account for \approx 60% of the corresponding *Eco*RI fragments in the human genome.

Structural Analysis of Two α -Tubulin Sequences. Sequences contained in four clones $(2\alpha, 41\alpha, 7\alpha, \text{ and } 19\alpha; \text{ Fig. } 1)$ were selected for study by restriction mapping and electron microscopic heteroduplex analysis. The human DNA in each recombinant bacteriophage was analyzed by endonuclease digestion, and the fragments were tested for their ability to hybridize to the ³²P-labeled clone pT1 probe after transfer to nitrocellulose. The results are shown in Fig. 2. In no case did these clones hybridize with a labeled cloned chicken β -tubulin cDNA probe. Two pairs of clones (2α and 41α , 7α and 19α) contain overlapping regions of DNA, each with a common hybridizing sequence. The hybridizing region defined by such Southern blots spans ≈ 4.8 kilobases (kb) in clones 2α and 41α and ≈ 4.1 kb in clones 7α and 19α .

To determine the direction of transcription within these genes, restriction digestions were performed on purified pT1 plasmid DNA to generate probes containing only the 3'-terminal or 5'-terminal portion of the cloned cDNA insert. These probes were then used in Southern blot experiments in which the restriction enzymes chosen to digest the recombinant bacteriophage DNA were selected to cleave at least once within the gene-containing region. Thus, digestion of clone 2α DNA with EcoRI yielded two fragments, one of 16 kb that hybridizes mainly with the 5'p probe and a second of 2.1 kb that hybridizes mainly with the 3's probe. Similar digestions with Bgl II (with or without EcoRI) and HindIII also yielded pairs of hybridizing fragments: within each pair, each fragment showed predominant hybridization to either the 5'p or the 3's probe (Fig. 3). The direction of transcription within clones 2α and 41α was thus defined as shown in Fig. 2. Analogous experiments using clone 19 α DNA allowed the determination of the direction of transcription in this gene. In this case, no hybridization was observed with any restriction fragment using the 5'p probe. However, the 5' end of 19α was identified by using the longer (5'h) probe (Fig. 3).

To determine the structures of the genes contained in clones 2α and 19α and to further investigate the failure of Southern blot experiments to identify a 5'-hybridizing sequence in clone 19α , experiments were performed using (as in original) electron microscopic heteroduplex analysis. A purified Sac I fragment (see Fig. 2) of clone 2α containing the α -tubulin sequence was denatured and reannealed to pT1 DNA that had been digested with *Eco*RI restriction endonuclease. The latter enzyme was chosen because it does not cut the cDNA but linearizes pT1 to allow the orientation of the gene in the heteroduplexes (11). The



FIG. 2. Restriction maps of some cloned fragments containing α tubulin sequences. Restriction digests were resolved by agarose gel electrophoresis, transferred to nitrocellulose, and tested for their ability to hybridize with clone pT1 (4) labeled with ³²P by nick translation. Bands giving a positive hybridization signal served to define the homologous regions within each cloned fragment. (A) Overlapping clones 19α (upper) and 7α (lower). (B) Overlapping clones 2α (upper) and 41α (lower). Internal vertical bars denote common *E*coRI sites. Ba, *BamHI*; Bg, *Bgl* II; H, *Hin*dIII; K, *Kpn* I; P, *Pst* I; Sa, *Sac* I; Sm, *Sma* I; S, *Xba* I. Additional cleavage sites for these enzymes exist within some of the cloned fragments. With the exception of clone 41α , the left arm of the vector is on the left.



FIG. 3. Southern blots of digests of clone 2α DNA hybridized with 5'- (*Left*) and 3'- (*Right*) specific α -tubulin probes. Aliquots of DNA purified from clone 2α were digested with a restriction enzyme that cleaves within the α -tubulin gene-containing regions (see Fig. 2). The digests were resolved in a 0.7% agarose gel, and the fragments were transferred to nitrocellulose. The blots were then hybridized with one of a number of 5'-specific or 3'-specific probes. The probes were generated by digestion of clone pT1 with one of the restriction enzymes as shown in the line drawing and labeled with ³²P by nick translation. *Pvu* II and *Hind*II cut once and twice within the pBR322 vector, respectively. In experiments with *Sma* I, further digestion with *Eco*RI was used to cleave the vector. Bg, *Bgl* II; H, *Hind*III; H2, *Hind*II; R, *Eco*RI; Ps, *Pst* I; Pv, *Pvu* II; S, *Sma* I.

Sac I fragment contained several pairs of short inverted duplications involving several hundred nucleotides each and formed stem-loop structures in single-stranded DNA. At least some of these sequences have alternative pairing partners, giving rise to many different stem-loop structures in individual single strands. Because the chicken cDNA is inserted into pBR322 via oligo(dG) and oligo(dC) tailing, linear pT1 also formed a stem loop in single strands. To form heteroduplexes between such DNAs, fairly stringent conditions were used to melt out the secondary structures in one of the hybridizing partners. Nonetheless, heteroduplexes were found at reasonable frequency, indicative of a high degree of sequence homology. Interpreta-

tion and data analysis were possible despite the multiple intrastrand interactions of 2α DNA. The α -tubulin gene had four conserved segments, designated C1-C4, separated by three intervening sequences, S1-S3 (Fig. 4 A and B). Measurements of 14 or more heteroduplexes for each of the segments gave the following results. Starting near the NH2 terminus, the order was as follows: the first conserved segment, C1 (0.15 \pm 0.03 kb), the first intervening sequence, S1 (1.0 \pm 0.05 kb), C2 (0.17 \pm 0.03 kb), S2 (0.7 \pm 0.04 kb), C3 (0.72 \pm 0.04 kb), S3 (1.9 \pm 0.09 kb), C4 (0.37 \pm 0.04 kb). The hybridizing sequences totalled 1.41 kb, spanning a region of 5.0 kb. As shown in Fig. 4B, two pairs of short inverted duplications involved sequences in S2, S3, and sequences downstream from the gene. In other molecules, C4 measured only 210-290 nucleotides, suggesting a certain degree of mismatch in the nontranslated region of the gene beyond the COOH terminus (11).

Analogous experiments were performed using a purified restriction fragment from clone 19α for heteroduplex formation with EcoRI-linearized pT1 DNA. In this gene, there were two intervening sequences separating three conserved sequences (Fig. 4C). The first conserved segment, C1, measured 0.16 \pm 0.04 kb and was followed by S1 (0.3 \pm 0.04 kb), C2 (0.17 \pm 0.03 kb), S2 (3.7 ± 0.12 kb), and C3 (1.0 ± 0.04 kb) (with 12 or more measurements for each segment). The α -tubulin-specific sequence of 1.36 kb spanned a region of 5.35 kb. In many heteroduplexes, the first two conserved segments failed to hybridize to the cDNA even though there was no secondary structure in the single-stranded DNA within the gene. Decreasing the stringency during renaturation had no effect, indicative of substantial sequence mismatch in these two segments. With the exception of the 3'-untranslated regions and the 5' end of 19α , no extended mismatched regions were observed in the hybridizing segments of either genomic clone under the nonstringent conditions for the preparation of electron microscope grids. The α -tubulin-specific regions in clones 2α and 19α are schematically represented in Fig. 4D. In each case, the orientation of the genes is the same as that determined by hybridization with 5'- and 3'-specific probes.

The above experiments used a heterologous (i.e., chicken) probe to identify and characterize human genomic sequences. To ensure that the sequences isolated using the chicken cDNA probe were indeed homologous to human α -tubulin sequences, the ability of cloned genomic fragments to hybridize to labeled human α -tubulin cDNA was tested under increasingly stringent conditions. Human α -tubulin cDNA was prepared using human α -tubulin mRNA selected from total HeLa poly(A)⁺mRNA by hybridization with linearized clone pT1 bound to a nitrocellulose filter (4, 11). The purity of the mRNA thus selected was determined by translation in the rabbit reticulocyte cell-free system. In this experiment, the only detectable translation product was α -tubulin (11). Selected human α -tubulin mRNA was then used as a template for cDNA synthesis in the presence of [³²P]dCTP, and this probe was tested for its ability to hybridize under increasingly stringent conditions to EcoRI fragments derived from all the clones isolated from the human genomic libraries. Under the most stringent conditions used (15 mM NaCl/1.5 mM Na citrate, 68°C), the labeling pattern of individual EcoRI fragments was identical to that obtained using the cloned chicken α -tubulin cDNA probe. We conclude, therefore, that there is considerable homology between chicken and human α -tubulin genes. This conclusion is consistent with recently obtained DNA sequence analysis data that shows 80% homology between a portion of the coding region of a human β -tubulin gene and the corresponding chicken cDNA sequence (unpublished results). Approximately half of these sequence



FIG. 4. Structures of α -tubulin-specific sequences in clones 2α and 19α as determined by electron microscopic analysis of heteroduplexes. (A) Heteroduplex between a Sac I fragment of 2α (Fig. 2) and EcoRI-linearized pT1 DNA. The inverted duplications in 2α did not pair and gave a well laid out heteroduplex. (B) As A except that the two pairs of inverted duplications in 2α had annealed. (C) Heteroduplex between EcoRI-linearized pT1 DNA and a 19α fragment generated by digestion with EcoRI/Sma I (Fig. 2). Interpretative tracings of regions of interest accompany each micrograph. —, 2α or 19α restriction fragment; -----, pT1 DNA. \rightarrow , Inverted duplications. C1-C4 are regions of hybridization; S1-S3 are intervening sequences. (D) Schematic representations of α -tubulin-specific sequences. \blacksquare , Sequences specific for α -tubulin; equivalent regions of 2α and 19α are indicated by dotted lines; results are in kb. Additional heteroduplex analyses show the Xba I (x) site near the 3' end of the gene (Fig. 2A) is <100 nucleotides from the end of the fourth conserved sequence. Another Bgl II site is in the second conserved sequence.

differences are the result of "silent" base changes that do not affect the amino acid sequence.

DISCUSSION

Our screening experiments resulted in the isolation of a number of different α -tubulin genes. Of the *Eco*RI fragments detected in a whole genomic Southern blot of human DNA, 60% are accounted for by the recombinant clones described. Analysis by Southern blotting of clones 2α , 41α , and 19α showed no evidence of additional α - or β -tubulin hybridizing regions within 12 kb upstream or downstream from the gene contained in clones 19α and 2α , respectively (Fig. 2). Restriction mapping of a number of other human α -tubulin gene-containing clones (Fig. 1) also shows no evidence of close linkage (unpublished data). Thus, unlike the genes encoding α - or β -globin (12), human α -tubulin genes seem likely to be members of a dispersed multigene family. This dispersed arrangement also appears to apply in the case of human β -tubulin genes (11) and in the tubulin genes of *Drosophila* (19) and chickens (20).

Electron microscopic analysis of heteroduplexes between genomic DNA contained in 2α and chicken α -tubulin cDNA indicated that the human α -tubulin-specific sequence is highly homologous, even under rather stringent conditions, to that of chicken, with possibly some divergence in the 3'-untranslated region (21). The sum of the hybridizing regions of 2α is virtually identical to that of the cDNA within the resolution of electron microscopy. The situation with the α -tubulin sequence in 19α is somewhat different. There is considerable mismatching between the first two conserved segments and the first 330 nucleotides of the chicken DNA. This is consistent with blotting experiments in which the 5' end of the gene cannot be located using a short 5' specific probe. The total length of hybridizing sequence in 19α was slightly less than the cDNA in pT1. This difference might be due to experimental error or to additional sequence divergence near the 3'-untranslated region. Indeed, such divergence seems likely since hybridization of cDNA to the 3' end of the gene beyond the Bgl II site (Fig. 4C) was not

detected in Southern blot experiments performed at higher stringency (Fig. 2A). Divergence of 3'-untranslated sequences has been previously described in chicken actin (4) and ovalbumin genes (22).

Because the cDNA probe used for heteroduplex analysis lacks ≈ 200 nucleotides corresponding to the 5' end of the mRNA, the existence of additional structural features at the 5' ends of the genes contained in 2α and 19α cannot be formally excluded. Attempts to form heteroduplexes using purified human α -tubulin mRNA (11) were unsuccessful, either because the amount of mRNA was too low for hybrids to form within the duration of the reaction or as a consequence of competing intramolecular reactions in the DNA due to the presence of several inverted repeats. Nonetheless, comparison of the structures of the α -tubulin sequences in 2α and 19α obtained using the cDNA probes shows some interesting features. The total lengths of sequences that hybridize to chicken cDNA are similar, and the first two conserved segments are identical in size. In addition, while the first two intervening sequences in the two clones are at very similar locations, they are of different lengths. Heteroduplexes between restriction fragments of 2α and 19α showed that there is no homology between these intervening sequences (data not shown). These heteroduplexes also confirmed that 19α lacked the third intervening sequence and that the 5' ends of the two genes have diverged, because they hybridized rather infrequently. Different intervening sequences at identical locations have been previously reported in the chicken ovalbumin and Xenopus vitellogenin gene families (22, 23). Rat preproinsulin genes are also related to one another by the presence or absence of an intervening sequence (24), as are the actin genes of *Drosophila* (25). Finally, the 2α clone contained several pairs of inverted duplications interspersed with the conserved segments of the gene whereas the 19α clone had no obvious secondary structures within the gene. Secondary structures have been previously observed in a human β -tubulin gene (11). The significance, if any, of such structures is not clear, although it has been suggested that members of the Alu I family, which have structures reminiscent of transposable elements (26), may function as sites for the origin of DNA replication (27). Some human β -tubulin genes consist of sequences that are incapable of yielding a functional β -tubulin mRNA (11). Whether one or both of the clones described here contain pseudogenes or functional genes will be revealed by DNA sequence analysis.

We thank Dr. T. Maniatis for the provision of the human genomic libraries. This work was supported by Grant GM 26456 from the National Institutes of Health.

- Cleveland, D. W., Kirschner, M. W. & Cowan, N. J. (1978) Cell 15, 1021–1031.
- Snyder, J. A. & McIntosh, J. R. (1976) Annu. Rev. Biochem. 45, 699-720.
- 3. Luduena, R. F. & Woodward, D. O. (1973) Proc. Natl. Acad. Sci. USA 70, 3594-3598.
- 4. Cleveland, D. W., Lopata, M. A., McDonald, R. J., Cowan, N. J., Rutter, W. J. & Kirschner, M. W. (1980) Cell 20, 965-1105.
- 5. Stephens, R. E. (1978) Biochemistry 17, 2882-2891.
- 6. Gozes, I. & Littauer, U. A. (1978) Nature (London) 276, 411-413.
- Marrotta, C. A., Harris, J. L. & Gilbert, J. M. (1978) J. Neurochem. 30, 1431-1440.
- 8. Sheir-Neiss, G., Lai, M. H. & Morris, N. R. (1978) Cell 15, 639-647.
- Postingl, H., Krauhs, E., Little, M. & Kempf, T. (1981) Proc. Natl. Acad. Sci. USA 78, 2757-2761.
- Shay, J. W., Feit, H. & Neudeck, U. (1980) in "Microtubules and Microtubule Inhibitors," eds. DeBrabander, M. & DeMey, J. (Elsevier/North Holland, Amsterdam), pp. 545-553.
- Cowan, N. J., Wilde, C. D., Chow, L. T. & Wefald, F. (1981) Proc. Natl. Acad. Sci. USA 78, 4877–4881.
- 12. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1980) Cell 19, 959-972.
- Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) Cell 15, 1157–1174.
- 14. Benton, W. D. & Davis, R. W. (1977) Science 196, 180-181.
- Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connel, C., Quon, D., Shen, G. K. & Efstratiadis, A. (1978) Cell 15, 687-701.
- Blattner, F. R., Blechl, A. E., Denniston-Thomson, K., Faber, H. E., Richards, J. E., Slighton, J. L., Tucker, P. W. & Smithies, O. (1978) Science 202, 1279-1284.
- 17. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
- Rigby, P. W., Dieckmann, M., Rhodes, C. & Berg, P. (1977) J. Mol. Biol. 113, 236-251.
- Sanchez, F., Natzle, J. E., Cleveland, D. W., Kirschner, M. W. & McCarthy, B. J. (1980) Cell 22, 845–854.
- Cleveland, D. W., Hughes, S. H., Stubblefield, E., Kirschner, M. W. & Varmus, H. (1981) J. Biol. Chem. 256, 3130-3134.
- Valenzuela, P., Quiroga, M., Zaldivar, J., Rutter, W. J., Kirschner, M. W. & Cleveland, D. W. (1981) Nature (London) 289, 650-655.
- Heilig, R., Perrin, F., Cannon, F., Mandel, J. L. & Chambon, P. (1980) Cell 20, 625–637.
- Wahli, W., Dawid, I. G., Wyler, T., Weber, R. & Ryffel, G. U. (1980) Cell 20, 107-117.
- 24. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. & Tizand, R. (1979) Cell 18, 545-558.
- Fyrberg, E. A., Bond, B. J., Huskey, N. D., Mixter, K. S. & Davidson, N. (1981) Cell 24, 107–116.
- Haynes, S. R., Toomey, T. P., Leinwand, L. & Jelinek, W. R. (1981) Mol. Cell. Biol. 1, 573–583.
- Jelinek, W. B., Toomey, T. P., Leinwand, L., Duncan, G. H., Biro, P. A., Choudary, P. V., Weissman, S. H., Rubin, C. M., Houck, C. M., Deininger, P. L. & Schmid, C. W. (1980) Proc. Natl. Acad. Sci. USA 77, 1398-1402.