Alternative 5' exons either provide or deny an initiator methionine codon to the same α -tubulin coding region

Paul R. Dobner, Edward Kislauskis, Bruce M. Wentworth⁺ and Lydia Villa-Komaroff⁺

University of Massachusetts Medical Center, Department of Molecular Genetics and Microbiology, 55 Lake Avenue North, Worcester, MA 01605, USA

Received July 9, 1986; Revised and Accepted December 4, 1986

ABSTRACT

The primary structures of two overlapping novel α -tubulin cDNA clones isolated from a *Macaca fascicularis* testis cDNA library and the corresponding human gene are presented. Although the general structure of the human gene conforms to that of previously described mammalian α -tubulin genes, there is a surprising difference: the ATG initiator codon is conspicuously absent. The macaque testis cDNA similarly lacks the initiator methionine, but otherwise encodes a variant α -tubulin isotype precisely conserved in the human gene. RNA blot analysis in the macaque, using a 3' untranslated region probe, revealed the existence of two additional related transcripts expressed in every tissue examined except the adult testis. Sequence comparisons indicate that the 2.0 kb testis transcript and one of the additional transcripts result from differential transcription of the same gene. The two transcripts differ only at the 5' end as a result of the recruitment of different 5' exons. Curiously, the 5' exon utilized outside the testis encodes an initiator methionine in the expected location.

INTRODUCTION

Microtubules are central to a diverse set of eukaryotic cell functions including cell division, axoplasmic transport, and flagellar movement (1). The major soluble protein components of microtubules, α - and β -tubulin, are each encoded by multiple, highly conserved genes in most organisms (for review see ref. 2). Functional diversity may reside in the tubulin subunits themselves, requiring multiple tubulin genes to encode functionally distinct tubulin subunits (3). The examination of the α - and β -tubulin gene families in a variety of organisms has revealed that, although the genes are highly conserved, genetic diversity does exist, and several evolutionarily conserved α - and β -tubulin isotypes have been defined (4, 5, 6, 7). The functional significance of these isotypes remains to be established; however, the greatest variation between isotypes occurs at the carboxy terminus of both α - and β -tubulins, a region required for interactions with microtubule associated proteins (MAPs) (8). Tubulin isotypes could contribute to the diversity of microtubule function by differentially binding MAPs.

The existence of multiple tubulin genes may also be the result of separate requirements for gene expression in particular pathways of terminal differentiation (9). The best evidence supporting this theory is the ability of recessive mutations in a *Drosophila* testis-specific β -tubulin gene to disrupt multiple microtubule functions (10). The existence of two human α -tubulin genes, which encode nearly identical α -tubulin isotypes, but exhibit vastly different patterns of expression would also seem to support this notion (11). Thus, the assembly of particular cytoskeletal structures may require not only differences in the constituent proteins but also differences in the regulated expression of the genes encoding these proteins. Tissue-specificity, timing, and stoichiometry could be of particular importance.

A number of microtubule based structures are elaborated during spermatogenesis in a highly ordered fashion perhaps requiring tubulin isotypes and modes of gene expression peculiar to the testis. By screening a *Macaca fascicularis* (crab eating monkey) testis cDNA library with a chicken α -tubulin coding region probe, we have identified a novel α -tubulin cDNA. Using a probe derived from a large unique region located at the 5' end of this cDNA, we have isolated the corresponding human gene. The initiator methionine codon, which in other mammalian α -tubulin genes (12, 13) constitutes the only coding capacity of exon one, is conspicuously absent in both the macaque cDNA and human gene sequences. However, expression of the gene in tissues other than the testis appears to involve the recruitment of a separate 5' exon which encodes an initiator methionine in the expected location. Thus, a choice made between alternative 5' exons has the unusual consequence that a methionine codon is either provided or denied to the same α -tubulin coding region.

MATERIALS AND METHODS

Materials.

The following items were obtained from the indicated sources: ³²P-labeled nucleotides, Amersham; nucleotides, PL Biochemical; restriction and most other enzymes, Boehringer Mannheim; reverse transcriptase, Life Sciences.

cDNA library construction, sequencing, and blotting procedures.

Poly(A)⁺ RNA, isolated from *Macaca fascicularis* testes, was used to prepare double-stranded cDNA as described (11). DNA sequences were determined by either the dideoxy (14) or chemical degradation (15) method. RNA and DNA blotting were performed as described (11)

Construction of subclones.

MT5'. The cDNA clone MTG3 α was linearized with Sst II, the ends were repaired with Klenow fragment and the DNA was digested with Hind III. The products of these reactions were separated on a 1% low melting temperature agarose gel and the desired Hind III/Sst II band was excised. DNA was recovered from the gel slice as described (16) and ligated into pUC9 (17) which had been digested with Hind III and Sma I.

3'UT. A 92 bp Alu I fragment of H α 44 (nucleotides 5304-5395, Fig. 5) was ligated into the Sma I site of M13mp9 (17). A phage containing the insert in the desired orientation was identified by DNA sequence analysis of two complementing phage.

Single-stranded probe templates.

322 bp Hae III. MTG3 α was digested with Pst I and the larger of the two insert bands was isolated from a low melting temperature agarose gel. This material was digested with Hae III, a 322 bp band was isolated and ligated into M13mp8 (17) which had been digested with Sma I.

386 bp Sst II. A pUC8 subclone containing a 3.0 kb Eco RI fragment derived from H α 44 was digested with Sst II and the ends were repaired with Klenow fragment. The products were separated by electrophoresis on a low melting temperature agarose gel, and DNA isolated from a 386 bp band which was ligated with M13mp9 cut with Sma I.

336 bp Taq I. Prepared as above except that the initial digestion was with Taq I and a 336 bp fragment was isolated for ligation into M13mp9.

Preparation of single-stranded probe and S1 protection assay.

Homogeneously labeled single-stranded probes were prepared by extending either a 15 or 17 base sequencing primer (New England Biolabs) using appropriate M13 subclones as templates (18). Single-stranded probe was isolated on 5% acrylamide, 8 M urea gels.

S1 protection assays were performed as described (19). The GC content of the probes and the incubation temperatures used for the hybridization reactions were as follows: 322 bp Hae III, 54% GC, 56°C; 386 bp Sst II, 71% GC, 63°C; 336 bp Taq I, 59% GC, 58°C. Primer extension was performed as described (20). The uniformly labeled primer used was composed of the non-coding strand between positions 1126 and 1273 (Fig. 5).

RESULTS

Isolation of a novel α -tubulin cDNA clone.

A testis cDNA library was constructed using poly(A)⁺ RNA isolated from *Macaca fascicularis* testis and screened with a ³²P-labeled chicken α -tubulin

Fig. 1. RNA blot analysis of α -tubulin gene expression in the testis. Monkey testis RNA was separated into poly(A)⁺ and poly(A)⁻ fractions by oligo dT-cellulose chromatography. 10 μ g from each fraction was size-fractionated on a 1.1% formaldehyde-agarose gel, transferred to Zeta-bind paper (AMF) and the same filter was hybridized either with a ³²P-labeled chicken α -tubulin cDNA insert (lane 1, poly(A)⁻; lane 2, poly(A)⁺) or the MT5' subclone (lane 3, poly(A)⁻; lane 4, poly(A)⁺). Hybridizing bands were visualized by autoradiography for 20 hrs. with intensifying screens. The two black bars denote the positions of 18S and 28S ribosomal RNAs run as size markers.

DNA insert (21). In addition to cDNAs representing the previously characterized human α -tubulin isotype, α 1 (11), two novel cDNAs, designated MTG3 α and MT6 α , were isolated. Preliminary sequence analysis revealed a region at the extreme 5' end of MTG3 α with no homology to any previously reported α -tubulin and which apparently preceded the coding region. A subclone, designated MT5', containing the 5' terminal 195 bp of MTG3 α , was constructed for use as a specific probe.

A distinct transcript is detected by the MT5' probe.

The existence of a large, unique region at the 5' end of MTG3 α suggested that the mRNA from which this cDNA was reverse transcribed might be larger than previously described α -tubulin mRNAs. Furthermore, the absence of any previously identified α -tubulin sequence, coding or non-coding, within the subclone MT5' predicted that it could be used to specifically detect the mRNA from which it was derived. To investigate this possibility, poly(A)⁺ and poly(A)⁻ RNAs, isolated from *M. fascicularis* testes, were analyzed by RNA blotting procedures (Fig. 1). A chicken α -tubulin probe, containing the highly conserved coding region, identified two RNA species, 2.0 and 1.8 kb in size, which were found predominantly in the poly(A)⁺ fraction (lanes 1 and 2). In contrast, the MT5' subclone detects only a 2.0 kb RNA, also found

Fig. 2. Tissue specific expression: a testis-specific α -tubulin gene. A ^{32}P -labeled single stranded probe was synthesized from a M13 template containing an inserted Hae III fragment from MTG3 α (see diagram). The portion of the insert (wide bar) which is highly homologous to other α -tubulins is stippled; the open portion represents sequences which are unique to the MTG3 α cDNA. The position of amino acid two in previously characterized α -tubulins and the Eco RI site used to create the 3' end of the probe are indicated. The bracketed regions represent the size of the probe (394 bases), and the expected protected fragments (see text). S1 nuclease assays were performed as described (see Methods) using 5 μg of macaque poly(A) $^{+}$ RNA except where indicated. A. Lane 1, undigested probe; lane 2, liver RNA; lane 3, kidney RNA; lane 4, testis RNA; lane 5, heart RNA; lane 6, lung RNA; lane 7, intestine RNA; lane 8, spleen RNA; lane 9, skeletal muscle RNA; lane 10, total poly(A) $^{+}$ RNA from *D. melanogaster*; lane 11, no RNA; lanes M, pBR322 DNA digested with Hinf I, ^{32}P end-labeled, and used as a size marker. B. Lane 1, undigested probe; lane 2, testis RNA; lane 3, kidney RNA; lane 4, human testis RNA; lane 5, no RNA; lane M, size marker as in A.

mainly in the poly(A)⁺ fraction (lanes 3 and 4). These results indicate that the insert contained in MTG3α is a reverse transcript of a 2.0 kb RNA expressed in the testis and that the MT5' subclone can be used to specifically assess the expression of this RNA.

The 2.0 kb mRNA is testis-specific.

To examine the pattern of expression of the 2.0 kb mRNA, an S1 protection experiment was performed using poly(A)⁺ RNA isolated from various macaque tissues (Fig. 2). The ³²P-labeled single stranded probe was synthesized from a M13mp9 subclone containing a 322 nucleotide Hae III fragment derived from the MTG3α cDNA clone and straddles the unique 5' and common coding region. Hybridization with macaque testis poly(A)⁺ RNA (Fig. 2A, lane 4) resulted in the protection of a 322 base fragment, the size expected if all of the non-vector sequence is protected. Hybridization with poly(A)⁺ RNA from a wide variety of other monkey tissues resulted in the protection of a 239 base fragment (Fig. 2A, lanes 2-3, 5-9), the size expected if only the sequence shared with other α-tubulins is protected. These results indicate that the expression of the 2.0 kb mRNA is testis-specific. The origin of the small amount of protected material observed with skeletal muscle RNA (Fig. 2A, lane 9) is uncertain since the 2.0 kb mRNA has not been detected in subsequent RNA blotting experiments using the same skeletal muscle RNA preparation (Fig. 8).

To verify that this RNA is expressed in humans, we performed a S1

Fig. 3. Dissection of the macaque and human α-tubulin multigene families. Aliquots (5 μg) of either *M. fascicularis* lung or human fetal brain DNA were digested with Eco RI, Bam HI, or Sst I, the resulting fragments were separated by electrophoresis on a 0.8% agarose gel, blotted to Zeta-bind paper and hybridized to ³²P-labeled DNA probes. Lanes 1-3, monkey DNA digested with Eco RI, Bam HI, or Sst I, respectively; lanes 4-6, human DNA digested with these same enzymes. A. Hybridization with a ³²P-labeled chicken α-tubulin coding region cDNA probe. B. Hybridization with a ³²P-labeled MT5' insert probe. Lane M λ DNA digested with Hind III.

protection experiment in which human testis poly(A)⁺ RNA was hybridized with the macaque probe (Fig. 2B, lane 4). A full-length protected product, which co-migrates with the largest protected product obtained using macaque testis poly(A)⁺ RNA (Fig. 2B, lane 2), as well as several smaller products, are visible. Thus, a highly homologous RNA is expressed in the human testis. The smaller products are likely to be due to mismatched bases, between the macaque probe and the human RNA (Fig. 5), which are only partially cleaved by S1 nuclease. The lower intensity of the protected products obtained with human RNA may be due to an actual lower level of the corresponding RNA or an increased sensitivity of the hybrid to S1 nuclease degradation because of sequence mismatches.

A single gene detected by the MT5' probe.

To determine the complexity of sequences related to those contained in the MT5' probe, macaque or human high molecular weight genomic DNA was digested with Eco RI, Bam HI, or Sst I, and examined by Southern blotting procedures. As expected, when a chicken α -tubulin coding region probe was used, a complex pattern of 15-20 bands was observed for both macaque and human DNA regardless

Fig. 4. Sequencing strategies. Sequences were determined using the strategies indicated by the arrows at the top and bottom and represent the minimum coverage of both strands, many more templates were actually sequenced (for details see Fig. 5). The stippled regions are highly homologous to previously described α -tubulins. The broken bars represent the 5' region of MTG3 α unique to this tubulin gene which comprises part of exon one in H α 44. The small 3'-untranslated region at the end of MTG6 α and the corresponding sequence in H α 44 are represented by solid bars. Arrowheads at the left hand side of H α 44 denote several repeat units of the human Alu family of repeated sequences. Thin lines represent human genomic DNA for which no homologous sequence is contained in the monkey cDNA clones. The thick arrow marks the location of the CAP site. The positions of Taq I, Sst II, and Hae III restriction fragments used to construct M13 clones for the synthesis of single-stranded probes (see Figs. 2 and 6) are indicated. Restriction enzyme sites are designated: B, Bam HI; E, Eco RI; H, Hind III; P, Pst I; S, Sst I.

Fig. 5. Sequence of H α 44, MTG3 α , and MT6 α . Sequences were determined using the strategy depicted in Fig. 4 using either the dideoxy or chemical method. The numbering starts at the first nucleotide inside the Eco RI site created by the cloning process. The positions of the three intervening sequences are indicated by arrows and the conserved consensus splice junctions are shown with the GT donor and AG acceptor dinucleotides underlined. Also underlined are the CAP site (1088), the AATAAA polyadenylation signal, and the nucleotide

to which the poly(A) tail is attached (5366). An asterisk has been placed above the first nucleotide (1274) of MTG3 α and differences in the monkey sequence are indicated. The bottom line indicates amino acid differences between the predicted amino acid sequence of H α 44 and that of k α 1, a sequence derived from a human keratinocyte cDNA clone (11).

of the restriction enzyme used (Fig 3A, lanes 1-6). However, the MT5' probe identified only a single band in each digest of both human and macaque DNA (Fig. 3B, lanes 1-6) indicating that this α -tubulin gene is single-copy in both genomes.

Isolation of the human gene.

A pre-existing panel of recombinant bacteriophage λ , containing human DNA inserts which hybridize with a chicken α -tubulin cDNA probe (22), was screened with 32 P-labeled MT5' DNA; one positively hybridizing phage, designated H α 44, was identified and further analyzed. Preliminary restriction mapping of H α 44 revealed an insert of 17 kb which, upon digestion with Eco RI, yielded fragments of 13, 2.9, and 1.4 kb. A restriction enzyme map of H α 44 is shown in Fig. 4 along with maps of the macaque cDNA clones which were essential for its isolation and characterization.

Sequence analysis reveals a missing initiator methionine codon.

The DNA sequence of the region of H α 44 depicted in Fig. 4 was determined, and the relevant areas are presented in Fig. 5 along with the corresponding composite sequence of the cDNA clones MTG3 α and MT6 α . Comparison of these sequences reveals that the H α 44 gene contains three intervening sequences in the conserved locations characteristic of mammalian α -tubulin genes (12, 13). The only coding capacity of the 5' exon in a previously characterized human α -tubulin gene, b α 1 (13), and in the homologous rat gene (12), is the initiator methionine codon, ATG. To our surprise, the 5' exon of H α 44 ends with the sequence CAG, and thus cannot provide an ATG codon in the homologous position. This is not likely to be a cloning artifact since this feature is shared by the MTG3 α cDNA clone and the H α 44 genomic clone.

The amino acid sequence predicted from exons 2, 3 and 4 of H α 44 (Fig. 4), when compared with that of a previously described human α -tubulin, k α 1, differs in 17 positions. In addition, the carboxy terminus of the H α 44 α -tubulin is truncated by three amino acids, including the carboxy-terminal tyrosine residue which is encoded by all previously characterized α -tubulin cDNAs (2). This variant α -tubulin coding region is precisely conserved between macaque and human.

The absence of an initiator methionine codon in the typical location

Fig. 6. S1 protection mapping of the 5' end of H α 44. ³²P-labeled single stranded probes derived from partially overlapping M13 subclones were hybridized to poly(A)⁺ RNA, digested with S1 nuclease, and the reaction products were analyzed on a 5% acrylamide, 7M urea sequencing gel. H α 44 (stippled bar and solid lines) as well as vector sequences (dashed lines) contained in the probes are indicated in the diagram below. Lane 1, undigested Taq I probe; lanes 2-5, Taq I probe and 5 μ g poly(A)⁺ RNA from macaque testis, lane 2; macaque kidney, lane 3; human testis, lane 4; or 5 μ g of yeast tRNA, lane 5. Lane 6, undigested Sst II probe; lanes 7-10, Sst II probe and RNA as in lanes 2-5, respectively.

coupled with the presence of an intact, highly conserved, α -tubulin coding region suggested that translation might begin elsewhere, perhaps within the 5' exon. Extension of the α -tubulin reading frame into the 5' exon revealed that it remains open throughout the entire length of the MTG3 α cDNA insert and the corresponding region of the H α 44 gene. Although the MTG3 α sequence contains an in phase ATG codon (76 codons upstream of the codon for amino acid 2 in

Fig. 7. Primer extension mapping of the 5' end of the Ha44 transcription unit. A 169 base ^{32}P -labeled primer was synthesized using a M13 subclone which contains a Sma I fragment derived from Ha44 (see text). A schematic of the primer in which the open bar and thin straight line represent Ha44 5' exon and vector sequences, respectively, is included. Primer (50,000 cpm) was annealed with either 1 μg (lanes 1, 3, and 5) or 5 μg (lanes 2, 4, and 6) of poly(A)⁺ RNA isolated from monkey testis (lanes 1 and 2), monkey kidney (lanes 3 and 4), or human testis (lanes 5 and 6) or 10 μg of yeast tRNA (lane 7), extended with reverse transcriptase and the reaction products were analyzed on a 6% acrylamide, 7M urea sequencing gel (below lanes; M, monkey; H, human). As a marker, a set of dideoxy sequencing reactions, synthesized using the same M13 subclone as was used for primer production, were run in adjacent lanes (G, A, T, C). The lane marked P represents unreacted primer. The position of the specific extension products, 39 nucleotides larger than the primer, is indicated by an arrow.

other characterized α -tubulins), the corresponding human sequence does not; therefore, it became necessary to define the full extent of the 5' exon to define its coding capacity.

Mapping the 5' end of Ha44.

Two partially overlapping subclones were constructed and used to synthesize probes for use in S1 protection experiments to more precisely map the 5' end of the gene (Fig. 6). A 386 base probe region, extending between two Sst II sites, and overlapping the region homologous with the 5' end of MTG3 α , was fully protected when hybridized to poly(A)⁺ RNA from either monkey or human testis (lanes 7 and 9) but not protected when hybridized to poly(A)⁺ RNA isolated from monkey kidney (lane 8). A Taq I fragment, which overlaps the Sst II fragment but extends 292 bases further in the 5'

direction, was not protected by hybridization with any of these RNA preparations (lanes 2-4). A 39 base protected fragment is predicted (see below), however, this short hybrid appears to be unstable under the hybridization conditions used. The 5' end of the gene must, therefore, lie very close to the 3' end of the Taq I fragment.

Alternatively, the Taq I fragment might lie within an intron and an additional exon might be present upstream. To decide between these alternatives, we synthesized a single-stranded primer using a M13 subclone containing a Sma I fragment derived from Ha44. The 3'-OH terminus of the primer was formed by digestion with Taq I and is thus coincident with the 5' terminus of the Taq I probe used in the S1 protection experiment described above (see Fig. 6, downward arrow). The primer was hybridized with two different concentrations (1 μ g or 5 μ g) of poly(A)⁺ RNA isolated from either macaque testis, or kidney, or human testis, extended with reverse transcriptase and the products were analyzed by gel electrophoresis (Fig. 7). Prominent extension products, 39 bases longer than the primer, were observed with macaque (lanes 1 and 2) or human (lanes 5 and 6) testis RNA, but not with macaque kidney (lanes 3 and 4) or yeast tRNA (lane 7). Two closely spaced bands are observed in the extension products of macaque testis RNA (lanes 1 and 2) indicating that some microheterogeneity in the 5' end of the transcript may exist. Comparison of the positions of the extension products with the dideoxy sequencing ladder indicates that the CAP site occurs at position 1088 $\pm$ 1 bp; thus, exon one is 470 bp in length. Inspection of the human 5' exon reveals that it does not contain an in phase methionine codon.

Multiple transcripts detected with a 3' untranslated region probe.

Because the structure of the 2.0 kb testis α -tubulin mRNA is unusual, we sought to verify the results we obtained with 5' exon probes using a subcloned probe derived from the 3' untranslated region of Ha44 (3'UT). Analysis of poly(A)⁺ RNA, isolated from various macaque tissues, with both the 3'UT (Fig. 8A) and a 5' exon (Fig. 8B) probe has instead revealed the existence of two additional transcripts, 1.8 and 2.2 kb in size (Fig. 8A). These additional transcripts apparently share 3' untranslated sequences with the 2.0 kb testis-specific transcript indicating that they are transcribed from either the same or closely related genes. As will be discussed below, the 1.8 kb transcript is almost certainly transcribed from the same gene as the 2.0 kb testis transcript. The 1.8 kb RNA is detected in every tissue examined except the adult testis (Fig. 8A). The 2.2 kb RNA is clearly visible in lung (lane 4), skeletal muscle (lane 7), juvenile testis (lane 9), and brain

Fig. 8. Multiple transcripts detected with the Ha44 3'UT probe. Poly(A)⁺ RNA (10 μ g), isolated from various monkey tissues, was analyzed as described in Fig. 1. The same filter was either hybridized to a ³²P-labeled single-stranded probe derived from the 3' untranslated region (A), or 5' exon (B) of the Ha44 gene. Lane 1, liver; lane 2, kidney; lane 3, heart; lane 4, lung; lane 5, intestine; lane 6, spleen; lane 7, skeletal muscle; lane 8, adult testis; lane 9, juvenile testis; lane 10, brain.

(lane 10), and is expressed at very low levels in the other tissues examined (visible upon overexposure of the autoradiograph). Interestingly, there is a clear shift from the expression of the 1.8 and 2.2 kb mRNAs in the juvenile testis to the 2.0 kb RNA in the adult testis (Fig. 8A, compare lanes 8 and 9). Thus, the 2.0 kb transcript is not observed until after the onset of spermatogenesis indicating that it is specific to germ cells.

DISCUSSION

The missing initiator methionine codon.

Two previously described mammalian α -tubulin genes are composed of four exons divided by three introns located in conserved locations (12, 13); the first intron occurs between the initiator methionine codon and the codon specifying amino acid two. The number and placement of the introns within the Ha44 gene are precisely conserved, however, the initiator methionine which normally precedes the first intron is conspicuously absent (Fig. 5). Similarly, the otherwise intact α -tubulin coding region predicted from two overlapping macaque testis cDNA clones (MTG3 α , MT6 α) is not preceded by an initiator methionine codon. A lesion in the initiator methionine codon has

Fig. 9. Two 5' exons of the Ha44 gene. A. The region of the Ha44 gene homologous with the 5' end of the monkey MTG3 α cDNA is depicted together with the predicted amino acid sequence. The human gene sequence begins at the testis cap site and ends at the inferred position of intron one. Differences in the macaque sequence, which begins at the asterisk, are indicated. B. The region of the Ha44 gene homologous with the 5' end of the mouse Ma4 cDNA (32) is depicted. The following features are indicated: Splice donor dinucleotides, doubly underlined; "TATA" homology, boxed; potential SP1 binding sites, underlined. The Ma4 sequence starts at the asterisk and differences are indicated.

been proposed to have been the initial silencing event in the formation of the primate β -globin non-processed pseudogene (23). The absence of an initiator methionine codon in both the human Ha44 gene and macaque MTG3 α cDNA indicated that the Ha44 gene may have been similarly silenced prior to the divergence of the macaque (Old World monkey) and human lineages 30 million years ago (24). However, analysis of the primate β -globin and other pseudogenes indicates that after the initial silencing event the entire gene accumulates nucleotide substitutions at the neutral rate (23, 25). Thus, even at the decelerated rate of neutral evolution in primates (23, 26) additional differences between the macaque and human α -tubulin coding regions should have accumulated. The α -tubulin coding region predicted from the human gene and macaque cDNAs are identical and contain no additional lesions. This strict evolutionary conservation provided the first indication that the Ha44 gene is not simply a pseudo-gene.

Tissue-specific expression.

A testis-specific 2.0 kb poly(A)⁺ RNA is detected in the macaque using

5' exon probes derived from either the macaque MTG3 α cDNA or human H α 44 gene (Figs. 1 and 9). Two additional poly(A)⁺ RNAs of 1.8 and 2.2 kb are detected in every tissue examined except the adult testis, using a probe (3'UT) derived from the 3' untranslated region of H α 44 (Fig. 9A). These results suggest that the three RNA species share common 3' ends, but vary at their 5' termini. Transcription from different promoters (27), alternative splicing of the same primary transcript (28), or a combination of the two (29, 30), can result in different mRNAs which, however, share common exons. Alternatively, related transcripts could be the products of separate but related genes (31). Southern blots of macaque and human DNA have revealed the existence of sequences other than the H α 44 gene which hybridize at high stringency with the 3'UT probe, thus either or both of the alternatives is possible. However, sequence comparisons between the H α 44 gene and a mouse cDNA clone (see below) indicate that at least one of the additional transcripts is generated by transcription of the same gene from a different promoter.

Provision or denial of an initiator methionine codon.

Two cDNAs, recently isolated from either a mouse testis (Ma4_L) or bone marrow (Ma4) cDNA library, encode an α -tubulin isotype identical to that encoded by the primate sequences we have isolated (32). The 5' sequence of the Ma4_L cDNA is homologous to the 5' exon of H α 44 and MTG3 α , and like the primate sequences does not encode an initiator methionine codon in the expected location (32). The Ma4 cDNA contains a different 5' end which maps to the first intervening sequence of the H α 44 gene (Fig. 9B). The existence of a consensus splice donor sequence immediately following the region of homology indicates that it constitutes a distinct 5' exon. Interestingly, the human mRNA which would result from the utilization of this 5' exon and the mouse mRNA represented by the Ma4 cDNA both encode an initiator methionine in the expected location.

We have mapped the cap site of the testis specific 2.0 kb mRNA to position 1088 $\pm$ 1bp of the H α 44 gene. The presence of a "TATA" homology (33) and several apparent SP1 transcription factor binding sites (34) upstream of the region of homology with the 5' end of the Ma4 cDNA indicates that recruitment of the two different 5' exons is mediated by separate promoters not by alternative RNA splicing of a single primary transcript originating from the testis cap site. The position of the "TATA" homology suggests that the resultant mRNA corresponds to the 1.8 kb mRNA species detected with a 3'UT probe in the macaque (Fig. 8) and mouse (32).

A pseudo-transcript?

The ability to produce a 1.8 kb mRNA encoding an initiator methionine in the expected location could solely account for the evolutionary conservation of the H α 44 α -tubulin coding region. The 2.0 kb testis-specific transcript may be adventitious and non-functional. This notion is hard to reconcile with the sustained production of this mRNA at least since the mammalian radiation. Perhaps equally likely is the possibility that translation of the 2.0 kb mRNA begins at an atypical location or that the unusual structure signifies an unusual requirement for translation initiation or control in the testis.

The existence of substantial open reading frames in phase with the α -tubulin coding region in both the human H α 44 5' exon and the corresponding region of the monkey MTG3 α cDNA opens the possibility that the testis 2.0 kb mRNA encodes an amino-terminally extended α -tubulin. The 5' exon choice would then be similar to that made in the production of two different myosin light chain genes (29, 30) in that the initiator methionine codon and, thus, the amino terminus of the coding region is affected. Alternative 5' exons also encode different amino-termini in four mouse c-abl mRNAs (35). However, an in phase methionine codon contained in the monkey cDNA (underlined in Fig. 9A), which if utilized would result in a 75 amino acid extension of α -tubulin, is not conserved in the human 5' exon sequence. Since translation initiation at codons other than AUG is extremely rare in eukaryotes (36, 37), the absence of an in phase methionine codon in the human 5' exon suggests that it is translationally inactive. In addition, the rate of nucleotide substitution in the 5' exon between macaque and human (1.1×10^{-9} per nucleotide per year) is very close to the primate neutral rate (23, 26). A similar comparison with the corresponding region of the mouse M α 4 α cDNA (32) confirms that the 5' exon is evolving at the neutral rate; furthermore, the 5' exon open reading frame predicted from the human and macaque sequences is not conserved in the mouse. Thus, the 5' exon appears to be evolving without constraint indicating that it has no protein coding function. In contrast, the alternative 5' exons of two myosin light chain mRNAs, which encode either the initiator methionine alone or a 38-42 amino acid amino-terminal extension, are highly conserved between chicken (29) and mouse (30).

The existence of long open reading frames in a region of the mRNA which presumably does not encode protein, although unusual, is not unprecedented. The mRNA encoding rat multiplication stimulating activity (rat IGF II) encodes an open reading frame of 392 amino acid residues, 201 codons of which are likely to be 5' untranslated sequences and the structure of the human IGF II

mRNA is similar (38). The high G + C content of the human H α 44 5' exon (68%) and corresponding region of the macaque MTG3 α cDNA (64%) could be responsible for the paucity of in frame termination codons as has been suggested for the rat and human IGF II mRNAs (38).

One intriguing structural feature of the H α 44 α -tubulin coding region is the substitution of a methionine codon for an isoleucine at position 16 (Fig. 5). Since in eucaryotes there appears to be some flexibility in the location of the initiation codon (39), translation could conceivably start at this conserved methionine codon resulting in an amino-terminally truncated α -tubulin. The nucleotides neighboring the initiator methionine codon have been demonstrated to be important for translation initiation and a consensus sequence has been compiled (40). The nucleotides surrounding the methionine codon at position sixteen poorly match the consensus sequence except for the G immediately following the ATG. However, the nucleotides preceeding the ATG closely match those upstream of the initiator methionine codon of another human α -tubulin, κ 1 (11). Translation from this site would necessitate a violation of the scanning hypothesis (41), but several mRNAs have been described where such violations occur. One of these is the yeast GCN4 mRNA which like the 2.0 kb testis mRNA contains an exceptionally long leader sequence (42). The GCN4 mRNA is translationally regulated and upstream AUG codons within the leader sequence are directly involved (43). Thus, the H α 44 5' exon may be required for control of translation from the methionine codon at position 16.

A divergent α -tubulin isotype.

The α -tubulin isotype described here differs at 20 positions with the previously described human κ 1 isotype (Fig. 5). These differences are likely to be significant since they are precisely conserved between the primate sequences reported here and mouse (32). As was first observed for two β -tubulin isotypes (4), the region of highest diversification is located at the carboxy terminus. Evidence exists indicating that this region of both α - and β -tubulin subunits is responsible for microtubule associated protein (MAP) interactions (8). Thus, different MAP interactions may be specified by the carboxy-terminal sequences of different tubulin isotypes (4, 5, 6, 7). An additional peculiar feature of the isotype reported here is the absence of a carboxy terminal tyrosine residue. However, the α -tubulin subunit could be tyrosylated post-translationally by an enzyme which is particularly abundant in brain and muscle (44).

The divergent primary sequence of the H α 44 isotype suggests that it

fulfills some special functional requirement. Preliminary screening of a skeletal muscle cDNA library suggests that this isotype (i.e. as encoded by the mRNA which encodes an initiator methionine in the typical location) may be the only one expressed in adult skeletal muscle (data not shown). Since adult muscle is composed primarily of non-dividing cells, a cytoskeletal or intracellular transport role is indicated.

Summary

The evidence discussed above supports a model in which the human H α 44 gene is transcribed in most tissues to yield a 1.8 kb mRNA encoding a divergent α -tubulin isotype, but is transcribed from a distinct promoter in the adult testis to yield a 2.0 kb testis-specific mRNA. The two mRNAs differ only at their 5' ends, the result of a choice between two alternative 5' exons apparently dictated by promoter utilization. Typical of two previously described mammalian α -tubulin genes, the H α 44 5' exon recruited in the formation of the 1.8 kb mRNA encodes only the initiator methionine, however, this codon is conspicuously absent in the testis-specific 2.0 kb mRNA. The rapid evolution of the testis 5' exon at the neutral rate would seem to eliminate the possibility that it encodes an amino-terminally extended α -tubulin. Instead, the occurrence of a novel methionine codon 15 amino acids downstream of the usual location suggests that the testis 2.0 kb mRNA encodes an amino-terminally truncated α -tubulin. An additional possibility is that the testis 2.0 kb transcript is translationally inactive and non-functional, although this notion is hard to reconcile with its continued expression at least since the mammalian radiation. Transcription of the H α 44 gene in the testis could play some as yet undefined role in a process other than translation. Such a non-translation role has been suggested for C μ transcripts which result from initiation within the IgH enhancer region and contain a 5'-nontranslatable exon (45).

ACKNOWLEDGEMENTS

We thank the University of Washington Regional Primate Center tissue banking and distribution program (supported by NIH grant RR 00166), and the Harvard University New England Regional Primate Center for providing Macaca fascicularis tissues; Dr. Mani Menon of the University of Massachusetts Medical Center for providing human testes samples; and Linda Niedzwiec and Michael Burley for expert technical assistance. We are indebted to Drs. Walter Gilbert and Allan Jacobson for helpful comments on the manuscript. We extend special thanks to Dr. Nicholas J. Cowan for providing the H α 44 genomic

clone, helpful discussions, and critical review of the manuscript.

This work was supported by National Institutes of Health grants HL33307 and GM26028 to P.R.D. and L.V.K., respectively, and a grant from Biogen, N.V. to L.V.K.

⁺Present address: Department of Neuroscience, The Children's Hospital, 300 Longwood Avenue, Boston, MA 02115, USA

REFERENCES

1. Kirshner, M.W. (1978) *Int. Rev. Cyto.* 54, 1-71.
2. Cleveland, D.W., and Sullivan, K.F. (1985) *Ann. Rev. Biochem.* 54, 331-365.
3. Fulton, C. and Simpson, P.A. (1976) In Goldman, R., Pollard, T., and Rosenbaum, J.L. (eds), *Cell Motility*, Cold Spring Harbor Press, New York, pp. 987-1005.
4. Hall, J.L., Dudley, L., Dobner, P.R., Lewis, S.A., and Cowan, N.J. (1983) *Mol. Cell. Biol.* 3, 854-862.
5. Sullivan, K.F. and Cleveland, D.W. (1984) *J. Cell. Biol.* 99, 1754-1760.
6. Lewis, S.A., Gilmartin, M.E., Hall, J.L. and Cowan, N.J. (1985) *J. Mol. Biol.* 182, 11-20.
7. Elliott, E.M., Henderson, G., Sarangi, F., and Ling, V. (1986) *Mol. Cell. Biol.* 6, 906-913.
8. Serrano, L., de la Torre, J., Maccioni, R.B., and Avila, J. (1984) *Proc. Natl. Acad. Sci. USA* 81, 5989-5993.
9. Raff, E.C. (1984) *J. Cell Biol.* 99, 1-10.
10. Kemphues, K.J., Kaufman, T.C., Raff, R.A., and Raff, E.C. (1982) *Cell* 31, 655-670.
11. Cowan, N.J., Dobner, P.R., Fuchs, E.V. and Cleveland, D.W. (1983) *Mol. Cell. Biol.* 3, 1738-1745.
12. Lemischka, I. and Sharp, P.A. (1982) *Nature* 300, 330-335.
13. Hall, J.L. and Cowan, N.J. (1985) *Nucl. Acids Res.* 13, 207-222.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.
15. Maxam, A. and Gilbert, W. (1980) *Meth. Enzym.* 65, 499-560.
16. Langridge, J., Langridge, P. and Bergquist, P.L. (1980) *Analy. Biochem.* 103, 264-271.
17. Messing, J. (1983) *Meth. Enzym.* 101, 20-78.
18. Ley, T.J., Anagnou, N.P., Pepe, G. and Nienhuis, A.W. (1982) *Proc. Natl. Acad. Sci. USA* 79, 4775-4779.
19. Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T., Potter, H., Lenoir, G., and Leder, P. (1983) *Cell* 34, 779-787.
20. Muskavitch, M.A.T., and Hogness, D.S. (1982) *Cell* 29, 1041-1051.
21. Cleveland, D.W., Lopata, M.A., McDonald, R.J., Cowan, N.J., Rutter, W.J. and Kirschner, M.W. (1980) *Cell* 20, 95-105.
22. Wilde, C.D., Crowther, C.E. and Cowan, N.J. (1982) *J. Mol. Biol.* 155, 533-538.
23. Goodman, M., Koop, B.F., Czelusniak, J., Weiss, M.L. and Slightom, J.L. (1984) *J. Mol. Biol.* 180, 803-823.
24. Sibley, C.G. and Ahlquist, J.E. (1984) *J. Mol. Evol.* 20, 2-15.
25. Li, W.-H., Gojobori, T. and Nei, M. (1981) *Nature* 292, 237-239.
26. Britten, R.J. (1986) *Science* 231, 1393-1398.
27. Young, R.A., Hagenbuchle, O., and Schibler, U. (1981) *Cell* 23, 451-458.
28. Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S., and Evans, R.M. (1982) *Nature* 298, 240-244.

29. Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M., and Ogata, K. (1984) *Nature* 308, 333-338.
30. Robert, B., Daubas, P., Akimenko, M.-A., Cohen, A., Garner, I., Guenet, J.-L., and Buckingham, M. (1984) *Cell* 39, 129-140.
31. Shani, M., Nudel, U., Zevin-Sonkin, D., Zakut, R., Givol, D., Katcoff, D., Carmon, Y., Reiter, J., Frischauf, A.M., and Yaffe, D. (1981) *Nucl. Acids Res.* 9, 579-589.
32. Villasante, A., Wang, D., Dobner, P.R., Dolph, P., Lewis, S.A., and Cowan, N.J. (1986) *Mol. Cell. Biol.* 6, 2409-2419.
33. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
34. Gidoni, D., Dynan, W.S., and Tjian, R. (1984) *Nature* 312, 409-413.
35. Ben-Neriah, Y., Bernards, A., Paskind, M., Daley, G.Q., and Baltimore, D. (1986) *Cell* 44, 577-586.
36. Becerra, S., Rose, J.A., Hardy, M., Baroudy, B.M., and Anderson, C.W. (1985) *Proc. Natl. Acad. Sci. USA* 82, 7919-7923.
37. Zitomer, R.S., Walthall, D.A., Rymond, B.C., and Hollenberg, C.P. (1984) *Mol. Cell. Biol.* 4, 1191-1197.
38. Dull, T.J., Gray, A., Hayflick, J.S., and Ullrich, A. (1984) *Nature* 310, 777-781.
39. Sherman, F. and Stewart, J.W. (1983) In Strathern, J.N., Jones, E.W., and Broach, J.R. (eds), *The Molecular Biology of the Yeast *S. cerevisiae*: Metabolism and Gene Expression*, Cold Spring Harbor Press, New York, pp. 301-334.
40. Kozak, M. (1984) *Nucl. Acids Res.* 12, 857-872.
41. Kozak, M. (1983) *Microbiol. Rev.* 47, 1-45.
42. Thireos, G., Penn, M.D., and Greer, H. (1984) *Proc. Natl. Acad. Sci. USA* 81, 5096-5100.
43. Mueller, P.P. and Hinnebusch, A.G. (1986) *Cell* 45, 201-207.
44. Raybin, D. and Flavin, M. (1975) *Biochem. Biophys. Res. Comm.* 65, 1088-1095.
45. Lennon, G.G. and Perry, R.P. (1985) *Nature* 318, 475-478.