MOLECULAR AND CELLULAR BIOLOGY, Sept. 1985, p. 2454–2465
0270-7306/85/092454-12$02.00/0

Vol. 5, No. 9

Apparent Gene Conversion between β-Tubulin Genes Yields Multiple Regulatory Pathways for a Single β-Tubulin Polypeptide Isotype

KEVIN F. SULLIVAN, JOSEPH T. Y. LAU, AND DON W. CLEVELAND*

Department of Biological Chemistry, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Received 4 April 1985/Accepted 3 June 1985

We have determined the complete nucleotide sequences of two chicken β-tubulin genes, β1 and β2. These genes display an unusual pattern of segmental homology which indicates that they originally arose by gene duplication and have subsequently coevolved by a process that included localized gene conversion or intergenic recombination. Since the β-tubulin polypeptides encoded by the two genes are virtually identical (99.5%), particularly in the major β-tubulin isotype defining regions, they almost certainly constitute a single isotypic class of β tubulin. However, the regulatory properties of the two genes are highly divergent as indicated by analysis of their patterns of expression in different chicken cell types. β1 is the major transcript detected in skeletal muscle myoblasts, whereas β2 is the major β-tubulin transcript in cultured sympathetic neurons. The existence of these two genes appears to derive from a regulatory requirement whereby the expression of a single tubulin isotype is mediated through different regulatory programs in development and differentiation. These results thus provide direct experimental support for the hypothesis that gene conversion and intergenic recombination play an important role in evolution by uncoupling the evolution of structural genes from the regulatory sequences which control them.

The principal structural subunit of eucaryotic microtubules is the protein tubulin, a heterodimer which is comprised of related but nonidentical α and β polypeptides (29). Such microtubules are prominent filamentous structures which are found in eucaryotic cells and comprise an integral element of the cytoskeleton as well as of the mitotic apparatus, elongated neuronal processes, and eucaryotic cilia and flagella. The cellular functions even within the class of ubiquitous cytoskeletal microtubules are diverse. For example, these microtubules are important in executing changes in cell shape during embryogenesis (e.g., references 6, 26), in mediating the transport of substances and organelles within the cytoplasm (e.g., references 22, 38), and possibly in transduction of signals originating at the cell membrane (34). Thus, two important questions regarding the function of microtubules are immediately evident. First, how does a single structural element function in such diverse roles, and second, how is the expression of tubulin and other microtubule-associated proteins integrated with different programs that determine cell structure and function?

Recently, with the isolation of tubulin DNA sequences by molecular cloning techniques (for review see references 10 and 13), the outlines of an answer to these questions have begun to emerge. Analysis of a number of higher eucaryotic genomes has demonstrated that genes for both the α- and the β-tubulin polypeptides are present as dispersed, multigene families in most, if not all, metazoan species (1, 12, 16, 25, 31, 32, 51). To the important and unresolved question of the functional significance of tubulin multigene families (a general question which in fact remains unanswered for many other eucaryotic multigene families), two hypotheses have been advanced. Multiple tubulin genes may fulfill a structural requirement by encoding divergent polypeptides with functional properties tailored to particular applications within the various differentiated cells of an organism (17). Alternatively, the multiple tubulin genes may encode functionally equivalent polypeptides but fulfill a regulatory role by allowing tubulin expression to be differentially regulated in different cell types or during development and differentiation (28, 48).

To date, determination of β-tubulin protein and DNA sequences has provided circumstantial evidence in support of the structural hypothesis (1, 16, 20, 30, 57, 58, 60). For all available metazoan β-tubulin sequences, amino acid substitutions are not found randomly throughout the primary structure but are localized into clusters, most notably at the carboxy terminus (20) but also in the regions around residues 33–57, 80–110, and 315–351 (57, 58). Furthermore, the evolutionary conservation of certain isotype specific sequences implies that the specific sequence variations found in these regions reflect positively selected functional requirements for different β-tubulin isotypes (13). Thus, these data support the hypothesis of structural and functional differentiation among tubulin isotypes, but resolution of this question awaits direct functional analysis of individual tubulin gene products.

These results, however, do not exclude the possibility that multiple, orthologous polypeptide sequences are encoded within a single organism by genes which are acted upon by different regulatory signals. Previous work from this laboratory demonstrated that the chicken genome contains at least four functional β-tubulin genes (32) that possess strong homology to a cloned chicken brain β-tubulin cDNA. More recently, a fifth gene has been isolated, and the existence of two additional genes has been inferred on the basis of indirect observations (57, 58). In an effort to resolve the questions introduced above and to establish the structural and functional properties of this small, multigene family, we have begun to analyze each member by direct DNA sequence analysis.

We have now determined the DNA sequences of two chicken β-tubulin genes (β1 and β2) and have found that

* Corresponding author.

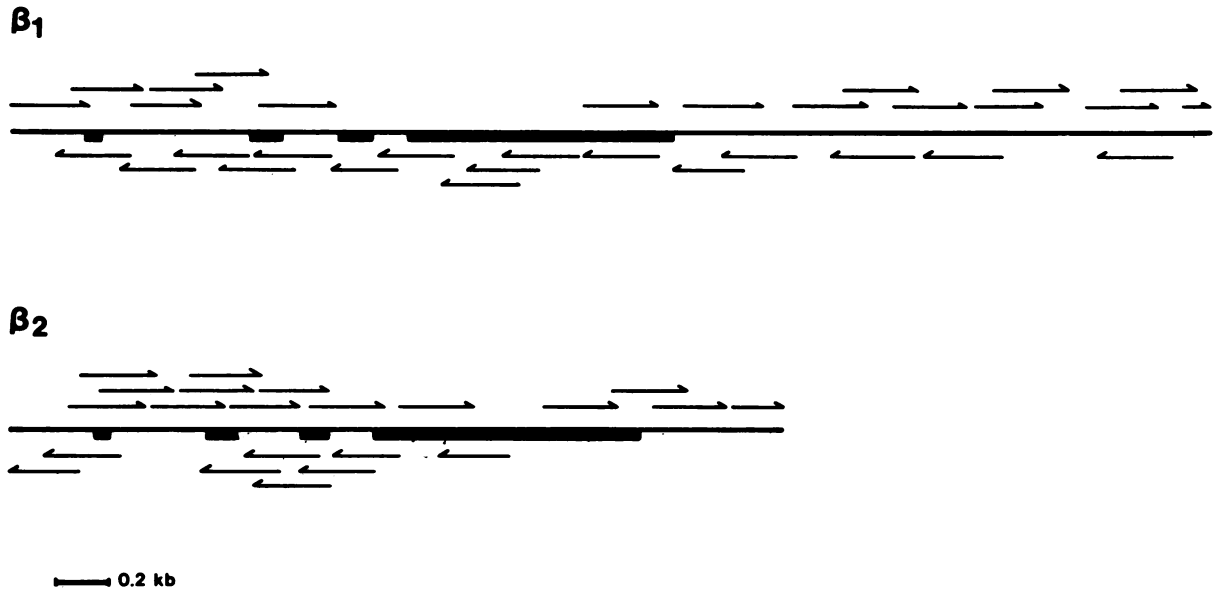


FIG. 1. Strategy for sequencing the β1 and β2 tubulin genes. DNA sequences were determined by the dideoxy technique using the progressive BAL 31 deletion procedure (44). Shown are the regions of pβG1 and pβG2 which were subjected to sequence analysis. Individual clones used to assemble the sequence are indicated by arrows. The intron/exon structure of the genes is indicated (noncoding sequences, thin lines; coding sequences, bold lines) and corresponds to that previously reported for other vertebrate β-tubulin genes (31, 58). The direction of transcription is from left to right.

these genes share a very high degree of nucleotide sequence homology and encode essentially identical polypeptides. The patterns of expression of the two genes are, however, very different. Whereas β1 is the predominant β-tubulin gene expressed in cultured skeletal muscle myoblasts, β2 is the predominant β-tubulin gene expressed in neural cells. Thus, this pair of β-tubulin genes, which encode apparently equivalent protein products but which are regulated by different programs during development and differentiation, represent a demonstration of the "regulatory" hypothesis within a tubulin multigene family. Furthermore, nucleotide sequence comparison of the two genes reveals a distinctly polar pattern of segmental homology, consistent with the involvement of gene duplication and subsequent gene conversion in the generation and evolution of the two genes. Taken together, these findings directly suggest a role for gene conversion in uncoupling the structural evolution of gene products from the evolution of the control signals which regulate them.

MATERIALS AND METHODS

DNA sequencing. DNA sequencing was performed using the dideoxy chain terminator method, as described by Sanger et al. (53). DNA fragments for sequencing were generated by digesting plasmids with a restriction endonuclease followed by BAL 31 exonuclease digestion (Bethesda Research Laboratories) essentially as described by Poncz et al. (44). After digestion with a second restriction enzyme, fragments were ligated to the appropriately cut M13 vector (35) and transformed into *Escherichia coli* JM101. Positive plaques were identified by lifting and probing with an appropriate ^{32}P-labeled DNA fragment isolated from the original plasmid. DNA was prepared from individual plaques essentially as described by Sanger et al. (52).

Nucleotide sequence analysis was performed with a VAX/VMS computer using the algorithms developed by Wilbur and Lipman (62) and Goad and Kanehisa (18). For sequence alignments, data were analyzed both for global and local homology, and alignments were compiled by hand using the results of the computer analysis as a guide. In judging alignment, homologous regions of six bases or more were located and fixed, and the remaining sequence was aligned by maximizing the number of base matches while minimizing the number of deletions required to fit the two sequences. Homology within coding regions was calculated by the procedure described by Chan et al. (8) as modified from Perler et al. (43). Homology in noncoding regions was determined by dividing the number of nucleotide matches by the number of bases in the shared region without attempting to correct for multiple events.

RNA analysis. RNA was isolated from secondary chicken fibroblasts, primary embryonic skeletal muscle myoblasts, and primary sympathetic neuron cultures by the method of Chirgwin et al. (9). Poly(A)$^+$ mRNA was prepared by chromatography on oligo(dT)-cellulose. RNA was quantitated by spectrophotometry.

An S1 analysis (4) was performed essentially as described previously (57). 3′ Probes were generated by digestion of pβG1 and pβG2 with *Bam*HI, which cuts at a site within codon 344 in the fourth exon. Probes were labeled using [α-^{32}P]ddATP (63). For analysis of RNA initiation sites, a 350-base-pair (bp) *Eco*RI-*Nar*I fragment was isolated from pβG1 and labeled at the 5′ ends using polynucleotide kinase (PL Biochemicals) and [γ-^{32}P]ATP (Amersham) after treatment with calf intestinal alkaline phosphatase (Boehringer-Mannheim). The strand complementary to β1 mRNA is thus labeled at the *Nar*I site at position +52 in the β1 sequence (with the A of the methionine translation initiation codon as nucleotide +1). A 5′ probe was prepared for β2 using a synthetic 17-mer (5′CGGTGTCTGCCGGTGCC3′) which is complimentary to nucleotides −6 to −23 of the 5′ untranslated region of β2. The oligonucleotide was labeled at the 5′ end using polynucleotide kinase (PL Biochemicals) and used

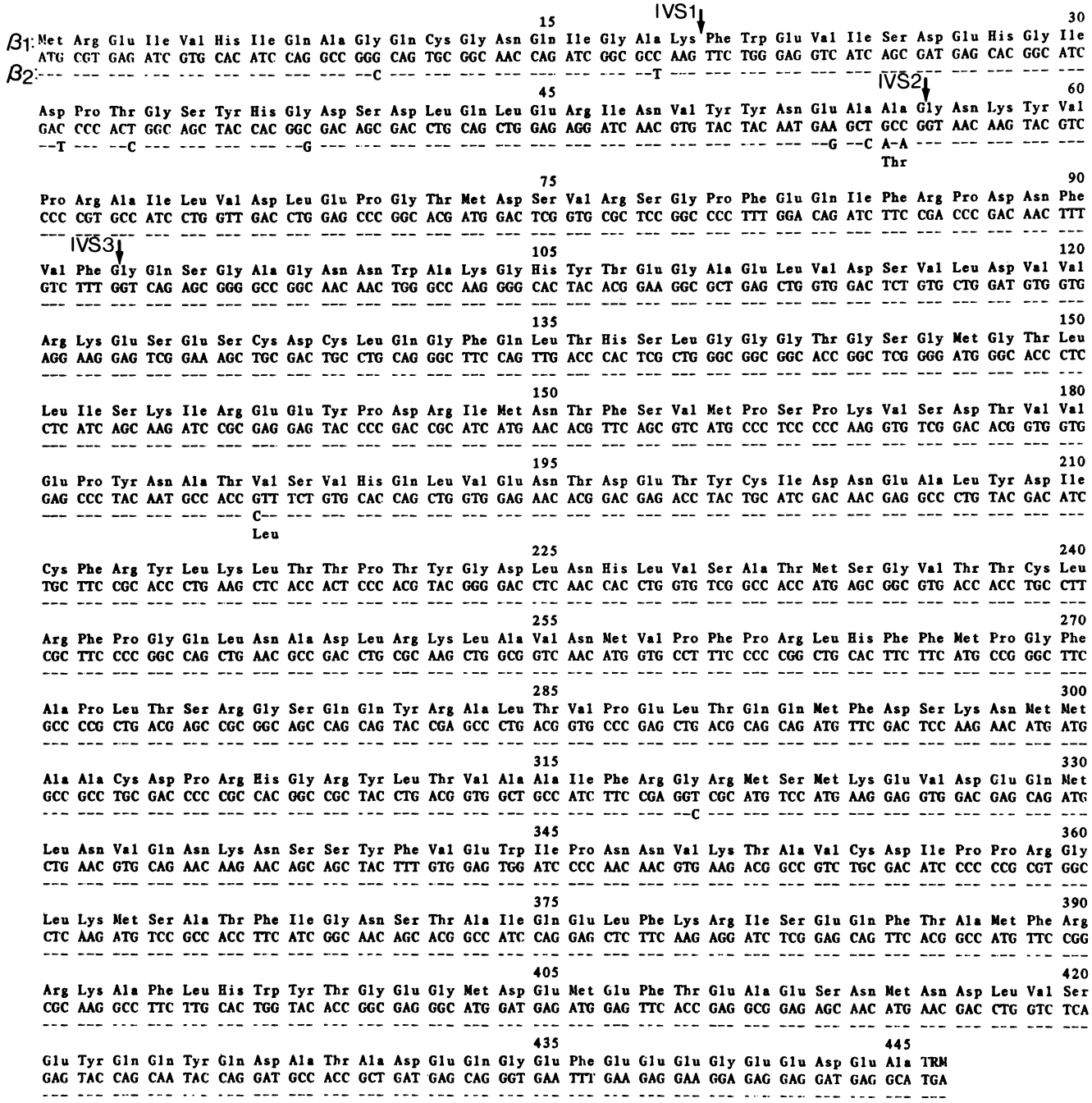


FIG. 2. Comparison of the coding sequences of β1 and β2. The coding sequence of β1 is shown with the predicted translation to protein sequence above it. The β2 sequence is shown beneath; identical nucleotides are denoted by dashes, and the predicted amino acid sequence is not shown except where it differs from that of β1 (e.g., codons 55 and 187). The sequence in the coding region of β2 is identical to that determined from a cDNA clone, pT2, derived from the β2 gene (60). Positions of the intervening sequences are denoted by arrows.

as a primer for probe synthesis from a single-stranded M13 subclone containing the 5′ region of β2. Routinely, 100 ng of poly(A)$^+$ mRNA was coprecipitated with 3×10^{-14} to 10×10^{-14} mol of probe and hybridized for 10 to 16 h at 60°C in the presence of 80% formamide. After digestion with S1 nuclease (Bethesda Research Laboratories), protected fragments were detected by electrophoresis and autoradiography.

RESULTS

The genomic DNA fragments containing two nonallelic chicken β-tubulin genes, β1 and β2, have been previously isolated by molecular cloning and subcloned into the plasmid pBR322 (32). Both genes are found on the same chromosome (11) but are not closely linked (>10 kilobases; 32). The complete sequences of the β-tubulin regions of the resultant plasmids, pβG1 and pβG2, respectively, have now been determined using the dideoxy chain terminator method (53). As detailed in Fig. 1, a region spanning 4,778 bp was determined for β1, and a 3,099-bp region was determined for β2. Both genes possess a structure typical of vertebrate β-tubulin genes, consisting of four protein coding exon sequences interrupted by three intervening sequences (IVSs)

(31), and both genes are flanked by appropriate consensus transcriptional control sequences as discussed in detail below.

Sequence comparison. (i) Coding sequences. An alignment of the coding regions of β1 and β2 is shown in Fig. 2 along with the corresponding predicted polypeptide sequences. Immediately apparent is a striking degree of homology between the two coding regions, which differ at only 11 of 1,338 nucleotide positions for an overall homology of 99.2%. This conservation is also reflected at the level of amino acid sequences, which differ by only two conservative substitutions (β1 Ala55-β2 Thr55 and β1 Val187-β2 Leu187). One of these substitutions (β1 Ala55-β2 Thr55) occurs within a region previously identified as a variable region among chicken β tubulins (57), whereas the other results in a change at Leu187 which is otherwise conserved in all other β tubulins so far examined (13). However, since the two polypeptides are identical in the other variable region clusters, particularly the highly variable carboxy terminus (16, 20, 57, 58), it seems appropriate to consider the two polypeptides encoded by β1 and β2 to constitute a single isotypic class of β tubulin.

Unexpectedly, close inspection of Fig. 2 reveals that nucleotide substitutions are not uniformly distributed throughout the coding region but are found primarily at the 5′ end of the genes. Of the 11 substitutions, 9 are found in the 165-bp region which constitutes exons 1 and 2, while only 2 are found in the remaining 1,173 bp of exons 3 and 4. Calculating the degree of divergence of these two domains independently (Table 1) demonstrates the nonuniform or segmental nature of the homology between these two genes. Using the estimated value for the rate of fixation of neutral mutations (0.7% per 10^6 years) that has been derived from previously studied vertebrate gene families (15, 43), we can estimate the time of divergence of the 5′ domain (exons 1 and 2) at 25×10^6 to 30×10^6 years ago, while in the 3′ domain the calculation predicts that the two genes shared sequence identity as recently as 0.5×10^6 years ago. While these estimates may not be accurate since they rest on the assumption that silent codon positions are selectively neutral (see Discussion), the calculations serve to underscore the unusual pattern of segmental homology between β1 and β2 and indicate that the mechanism(s) which achieved sequence concordance between the two genes has acted most recently in the 3′ domain alone.

(ii) Noncoding regions. The noncoding regions (promoter, 5′ untranslated region, introns, and 3′ untranslated region) also reflect the bipartite conservation between the two domains of β1 and β2 (see Fig. 4). Although the exon sequences 5′ to IVS1 show strong homology, IVS1 and the 5′ portion of IVS2 show very little homology, whereas the 3′ portion of IVS2 and all of IVS3 show the very high degree of homology characteristic of the noncoding sequences within recently diverged genes (e.g., references 15, 24). Remarkably, the 3′ untranslated and 3′ flanking regions show virtually complete divergence, except for the position and sequence of the unusual polyadenylation signal, ACATAAA.

TABLE 1. Divergence between β1 and β2 coding regions

Region	% Divergence[a]		
	Total	Replacement	Silent
Complete coding region	0.83	0.2	2.7
5′ Domain (exons 1 and 2)	5.45	0.8	19.4
3′ Domain (exons 3 and 4)	0.17	0.11	0.34

[a] Percent divergence was calculated as described by Chan et al. (8).

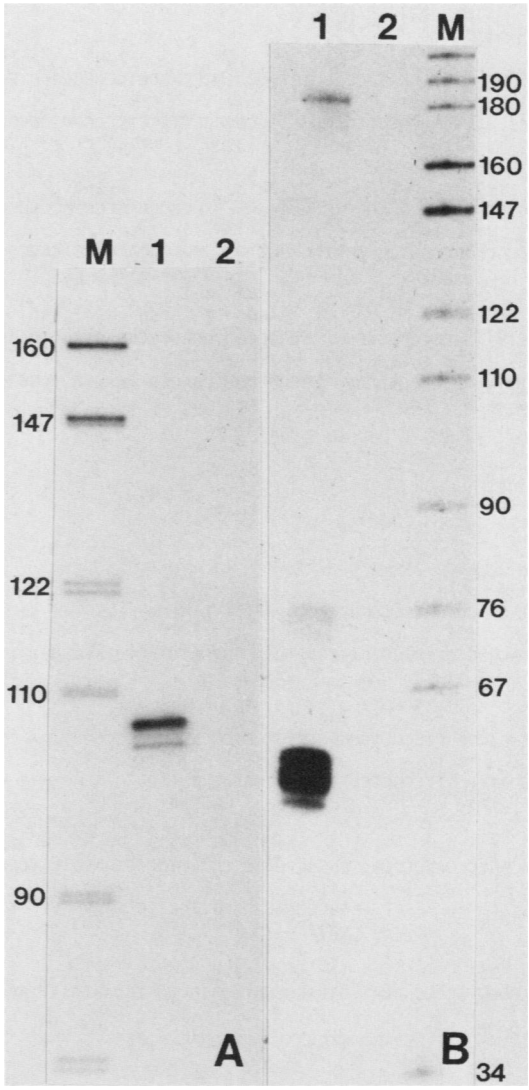
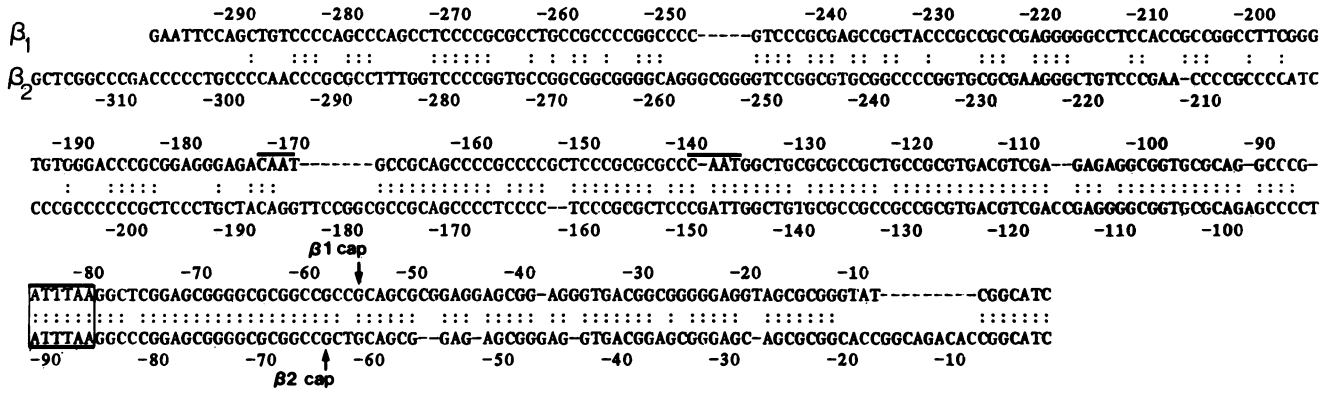
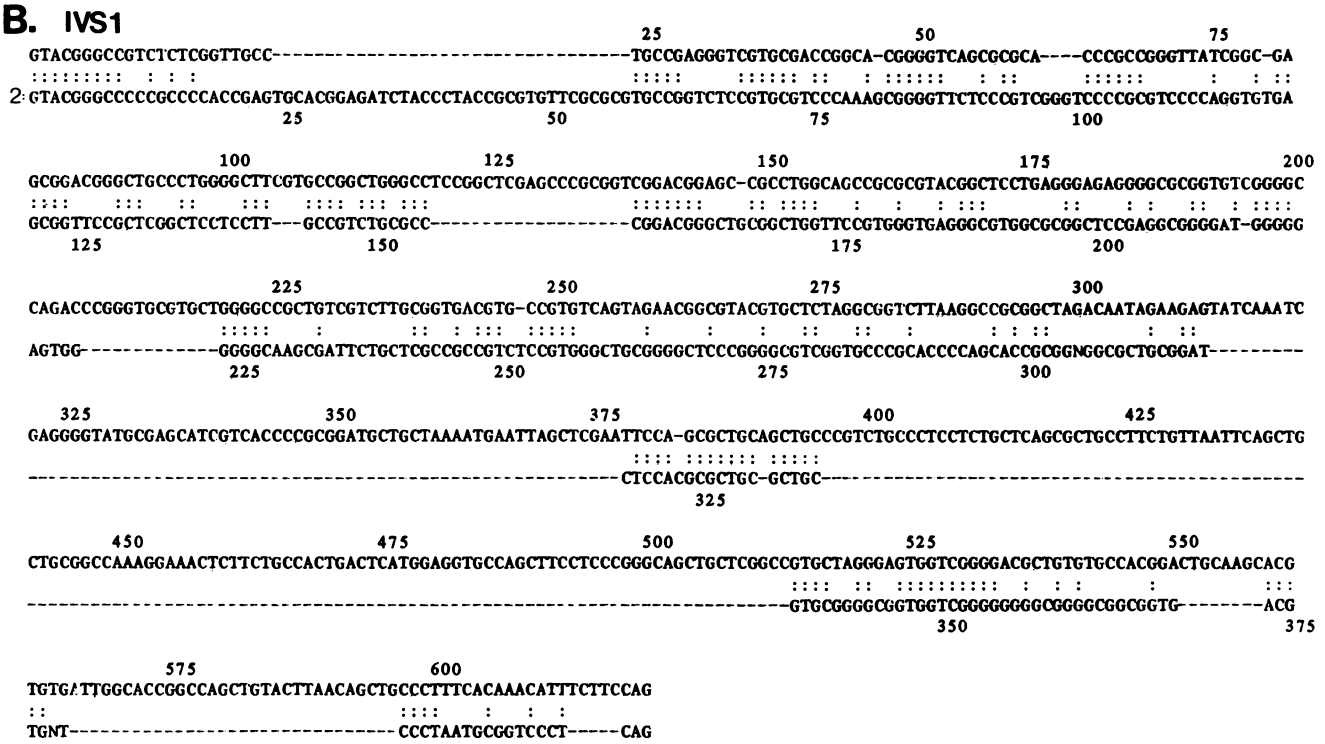
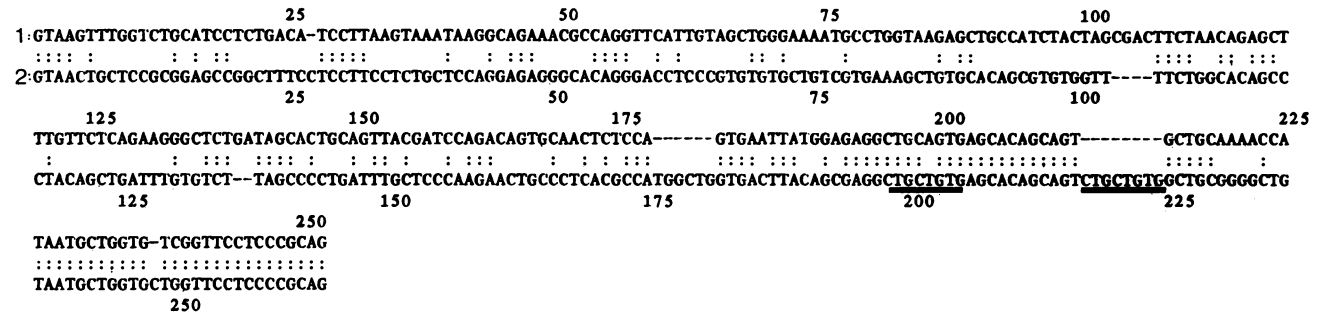


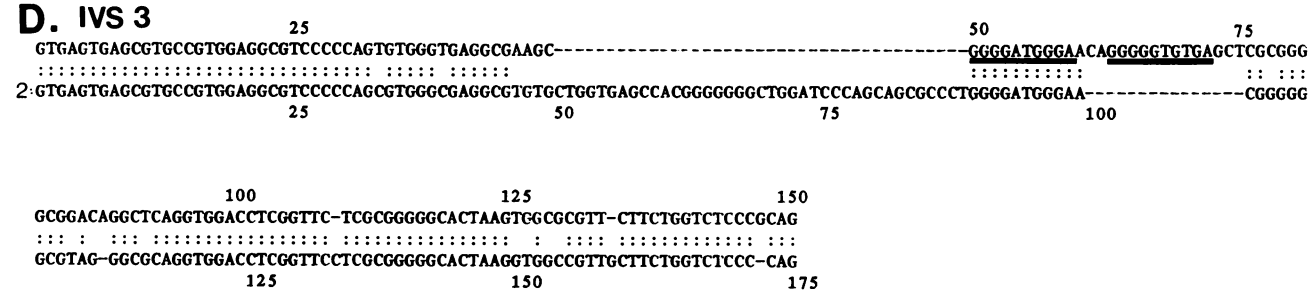
FIG. 3. Transcriptional start sites for β1 and β2. Start sites for transcription were determined using an S1 nuclease assay as described in Materials and Methods. (A) Identification of the putative transcription initiation site for β1. Lane 1, Fragments of a β1 probe (5′ end labeled at position +52) which were protected from S1 digestion by hybridization to chick fibroblast poly(A)$^+$ mRNA; lane 2, protected probe fragments from a hybridization reaction exactly like that in lane 1, except that no RNA was added. (B) Identification of transcription initiation sites of β2. Lane 1, S1-resistant fragments derived from hybridization of chick brain mRNA to a β2-specific probe 5′ end labeled at position −6; lane 2, fragments protected in a control hybridization reaction identical to that in lane 1, except that no RNA was added. Lanes containing molecular weight markers (in bases) are marked with M.

(iii) 5′ Flanking and untranslated sequences. To distinguish unambiguously between promoter and 5′ untranslated regions of each gene, the transcription initiation sites for β1 and β2 were determined by using S1 nuclease analysis of gene-specific DNA-RNA hybrids as described in Materials and Methods. For β1, a DNA fragment that begins at nucleotide +52 with respect to the A of the ATG translation initiation codon was labeled at the 5′ ends and was used to


A. 5' FLANK
β1 cap
β2 cap


B. IVS1


C. IVS 2



E. 3' FLANK

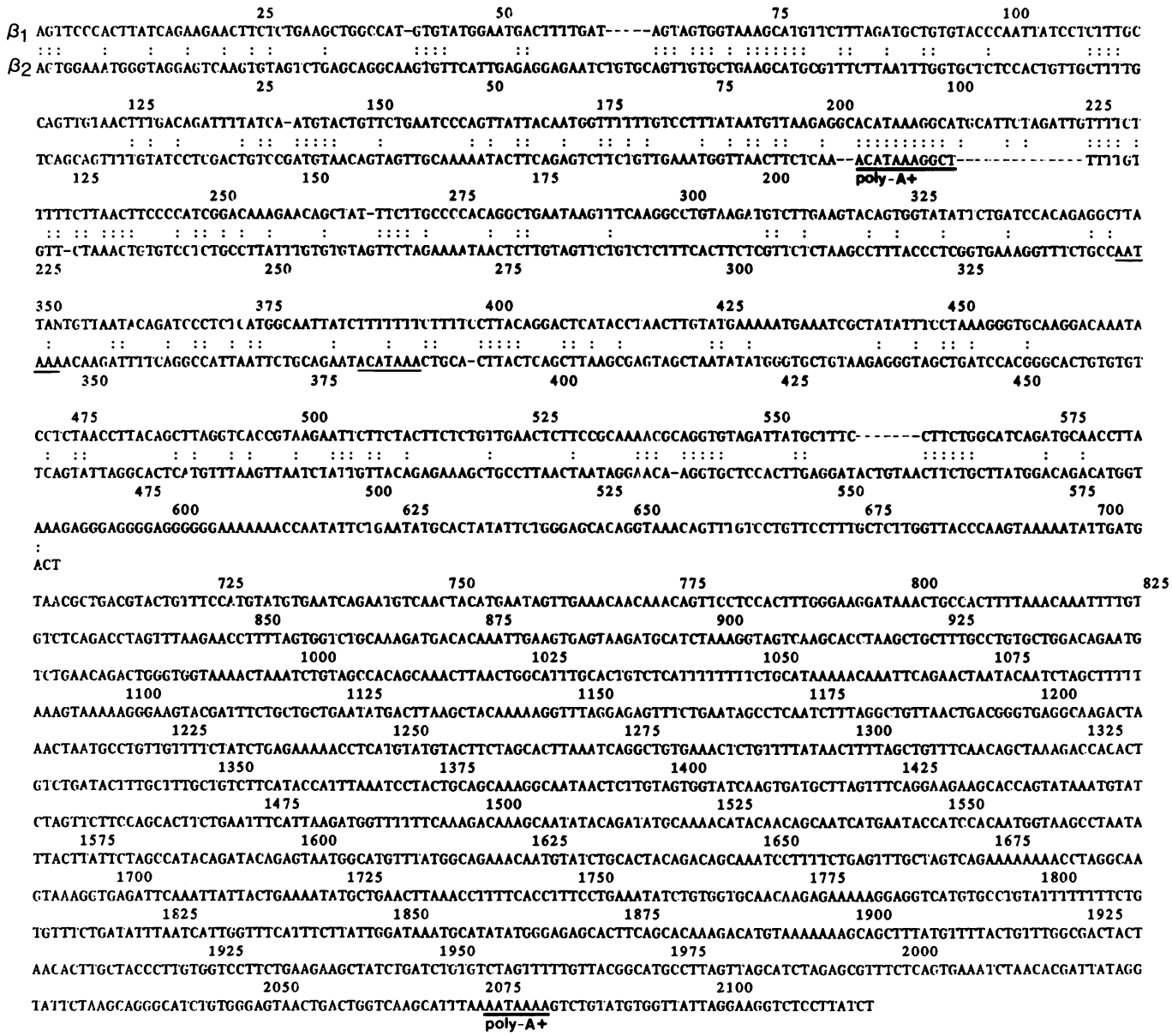


FIG. 4. Comparison of the noncoding sequences of β1 and β2. Sequences were aligned as described in Materials and Methods. The β1 sequence is shown in the top lines with the β2 sequence shown beneath. (A) 5′ Flanking sequences. The positions of putative regulatory signals, CAAT (bold lines) and the TATA box (boxed), are indicated, as are the apparent transcriptional start sites (arrows). (B) IVS1. (C) IVS2. (D) IVS3. The intervening sequences are numbered, with the G of the 5′ GT dinucleotide of each intervening sequence as nucleotide 1. Note the sequence at the boundary of homology within IVS2, which is detailed in Fig. 7. In part C, a nontandem 9-base direct repeat is underlined in the β2 sequence. In part D, an imperfect direct repeat in the β1 sequence is underlined. (E) 3′ Untranslated and flanking sequences of β1 and β2. Data are presented as in Fig. 2. Polyadenylation sites which are utilized in the synthesis of mRNA are denoted. Unused consensus polyadenylation signals in β2 are underlined with a thin line.

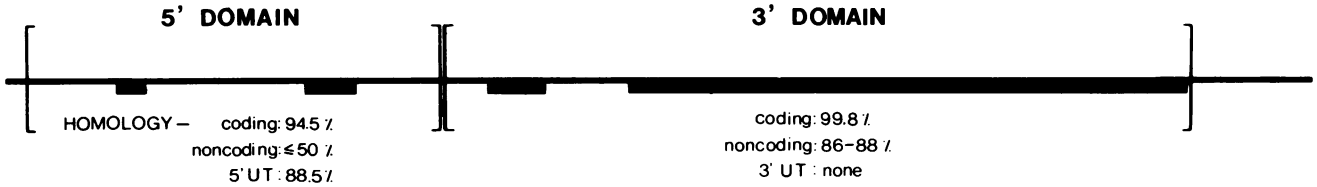


FIG. 5. Summary of the segmental homology observed between β1 and β2. The intron/exon structure of the β1 and β2 tubulin genes is shown. Brackets delineate the two regions of homology between β1 and β2. Shown beneath are the percent homologies which are found in the different portions of the two genes. Note that the boundary of the 5′ domain extends into noncoding flanking sequences, presumably reflecting conservation of transcriptional initiation signals, while the 3′ boundary of the 3′ domain is virtually coincident with the translation termination codon.

probe poly(A)$^+$ mRNA from primary chicken embryo fibroblasts (Fig. 3A, lane 1). The major protected fragment of 108 bases corresponds to a transcriptional start at nucleotide −55 (25 bases 3′ to an ATTTAA sequence).

Similarly, a probe for β2 was synthesized using a 5′-labeled oligonucleotide primer which starts at nucleotide −6 in β2, and this probe was used to analyze embryonic chicken brain poly(A)$^+$ mRNA. A prominent series of protected fragments approximately 57 bases in length was observed (Fig. 3B, lane 1), indicating an apparent transcriptional start at about −63 (22 bases 3′ to a TATTTAA sequence). Also observable are two less abundant series of protected fragments of ~75 and ~182 bases in length, which represent apparent transcription initiations at about −81 (which is within the putative TATA box) and at −188. (That a fraction of β2 RNAs do indeed derive from an initiation site 5′ to the major initiation at −63 is also clearly demonstrated in the sequence of pT2, a cDNA clone of a β2 mRNA [32]. Inspection of the known sequence [60] reveals that the cDNA begins precisely at position −88.)

Having established the border separating the promoters and 5′ untranslated sequences, comparison of the DNA sequences reveals extensive homology between β1 and β2 extending about 170 bp 5′ to the translation initiation codon (nucleotides −168 in β1 and −177 in β2; Fig. 4A). Within this region are found the "TATA" promoter sequences at −84 and −90 for β1 and β2, respectively, the transcription start sites, and the 5′ untranslated mRNA leader sequences. A total of 9 base substitutions, 10 single or double base gaps, and a large insertion/deletion adjacent to nucleotide −7 distinguish the two genes in this region, resulting in an overall homology of 88.5% (with deletions or insertions scored as single mutation events). No additional sequence homology can be discerned in the adjacent chromosomal sequences. In particular, the tetranucleotide sequence CAAT is found in β1, immediately flanking the 5′ border of homology. Although this sequence has been implicated in control of transcription in certain eucaryotic genes (5), a corresponding sequence is not present in β2. Overall, the data clearly demonstrate that the promoters for β1 and β2 are functional and predominantly direct transcription of mRNAs which initiate at almost identical positions. The two primary transcripts are therefore strikingly homologous in their 5′ untranslated leader sequences, differing mainly in the region adjacent to position −7 (see Fig. 4A).

(iv) Intervening sequences. Alignments of the intervening sequences of β1 and β2 are also shown in Fig. 4B through D. Very little homology is seen between the two genes in IVS1 and in the 5′ portion of IVS2. With the exception of 7 bp immediately flanking the 5′ splice site, only a few sparse clusters of homology are seen in IVS1 (Fig. 4B). Similarly, the 5′ portion of IVS2 (Fig. 4C) shows only vestiges of homology between the two genes.

In striking contrast to the sequences immediately adjacent, the 3′ portions of IVS2 (Fig. 4C) show a significant homology between the two genes, as does IVS3 (Fig. 4D). In IVS2, the region beyond position 175 differs by a single block transition involving 4 bp ($[A:T]_4$ to $[G:C]_4$), 7 base substitutions, and one deletion of 8 bp which is associated with a nontandem direct repeat (nucleotides 198–205 and 217–224 of β2 in Fig. 4C). Such repeat units have been previously shown to flank deletion points in other multigene families (15, 24). Counting both deletions and the block transition as single events yields a divergence of only 13.3% in the 3′ portion of IVS2.

IVS3 (Fig. 4D) shows a level of homology between the two genes similar to that of the 3′ region of IVS2. In this instance, the two introns differ primarily by two large insertion/deletion mutations, yielding three regions of homology separated by two stretches of nonhomologous sequence. The segmental mutation on the 3′ side of IVS3 is consistent with the deletion of an imperfect nontandem repeat (residues 50–59 and 63–72 of β1, Fig. 4D) which resulted in the loss of one repeat and some flanking sequence in β2. Overall, the IVS3s in the two genes differ by 11 single base substitutions, 4 single base gaps, and 2 large deletion/insertions. Counting the segmental mutations as single events yields a divergence of 12% between the two introns.

(v) 3′ Flanking sequences. Alignment of the 3′ untranslated and flanking sequences of β1 and β2 is shown in Fig. 4E. Clearly, the extreme homology which characterizes the 3′ domain of the two genes (Fig. 2) terminates abruptly 3 bp beyond the translation stop codon. Except for the positions (+202 and +206 for β1 and β2, respectively, where nucleotide 1 is taken as the first base 3′ to the translation stop codon) and unusual sequence (ACATAAA) of the proximal polyadenylation sites in the two genes, the 3′ untranslated regions do not share detectable homology. Curiously, β2 has two additional polyadenylation signals at +339 and +377, although only the more proximal is actually utilized (see below). For β1, in addition to the ACATAAA sequence at +202, a second, distal poly(A) signal sequence is located at +2079.

Gene duplication and concerted evolution: boundaries of homology between β1 and β2. As detailed above and summarized in Fig. 5, the overall level of homology between β1 and β2 indicates that they are products of gene duplication. Furthermore, in addition to maintenance of sequence homology through selective pressure, the discretely segmental nature of the homology between the two genes strongly suggests that within the 3′ domain a process of localized intrachromosomal (11) sequence exchange, which may have been mediated by gene conversion (2, 47) or by nonreciprocal recombination events (36, 59), has acted to maintain the high degree of homology observed. For the 5′ domain, although the original duplication unit probably extended

beyond the present borders, homology is now detectable only in regions which are presumptively subject to a functional constraint (e.g., coding region, 5′ untranslated region, and 3′ portion of each promoter). Noncoding flanking sequences show essentially no homology. We can thus estimate the minimum time since duplication at not less than 80 $\times 10^6$ to 100 $\times 10^6$ years ago, the predicted time required for complete divergence of two nonselected sequences (43).

Dual regulation of a single β-tubulin isotype encoded by two genes. The above analysis indicates strongly that β1 and β2 satisfy the structural requirements expected for a gene duplication event whose evolutionary endpoint has been to allow alternative pathways of regulated expression for a single polypeptide class. To begin to test whether these two genes are expressed in alternative programs during differentiation, we exploited the finding that although the two genes are identical in their 3′ coding regions they diverge immediately 3′ to the translation stop codon. Using an S1 protocol with probes 3′ end labeled within the 3′ coding region of either gene, RNAs derived from both genes can be detected simultaneously. For the homologous RNA, a DNA segment will be protected through the polyadenylation site actually utilized, while for the heterologous transcript a smaller segment extending only to 3 bases 3′ of the translation stop codon will be protected. Appropriate probes starting at codon 344 and running far into the genomic 3′ flanking sequences were isolated from each gene. Fragments were 3′ end labeled using [^{32}P]ddATP as described in Materials and Methods and were used to probe poly(A)$^+$ mRNA from chicken fibroblasts, cultured striated muscle myoblasts, and cultured sympathetic neurons.

The results (Fig. 6) demonstrate a dramatic differential expression of β1 and β2 in different cell types. With the β1 probe (Fig. 6A), a 310-base protected fragment corresponds to β2 transcripts while a 530-base protected fragment corresponds to β1 transcripts which terminate at the proximal polyadenylation site. β1 also possesses a distal polyadenylation site 2,079 bases downstream of the stop codon which, due to the limited resolution of large fragments on the acrylamide/urea gel, is not distinguishable in this experiment from residual undigested probe. This problem is eliminated, however, in the complementary experiment performed with the β2 probe (Fig. 6B), in which all β1 transcripts are summed in the 310-base fragment while β2 transcripts are reported by the 530-base fragment.

The relative quantities of the two transcripts were determined by densitometry of the data shown in Fig. 6B. These data (Table 2) demonstrate that β1 and β2 are expressed in nearly equivalent quantities in fibroblasts (β_1/β_2 = 1.6). In skeletal muscle myoblasts, however, β_1 is the predominant transcript (β1/β2 = 7.3), while the opposite is true in neurons (β1/β2 = 0.10). Furthermore, this analysis gives some insight into the mechanisms which control the expression of the two genes in these cell types. For β1, it can be seen that the level of transcripts per microgram of poly(A)$^+$ RNA accumulated in fibroblasts and myoblasts is approximately equal but is less than half that level in neurons, indicating a two- to threefold change in β1 mRNA accumulation in the different cell types. In contrast, relative to the level in fibroblasts, β2 transcripts show an approximately 4-fold decrease in skeletal muscle and a 6- to 7-fold increase in neurons, indicating that the level of accumulation of transcripts from this gene varies over a 30-fold range in the three cell types tested. Other experiments using RNA blot analysis have indicated that β1 is a minor transcript in most cell types except skeletal muscle myotubes, in which it is the major β-tubulin tran-

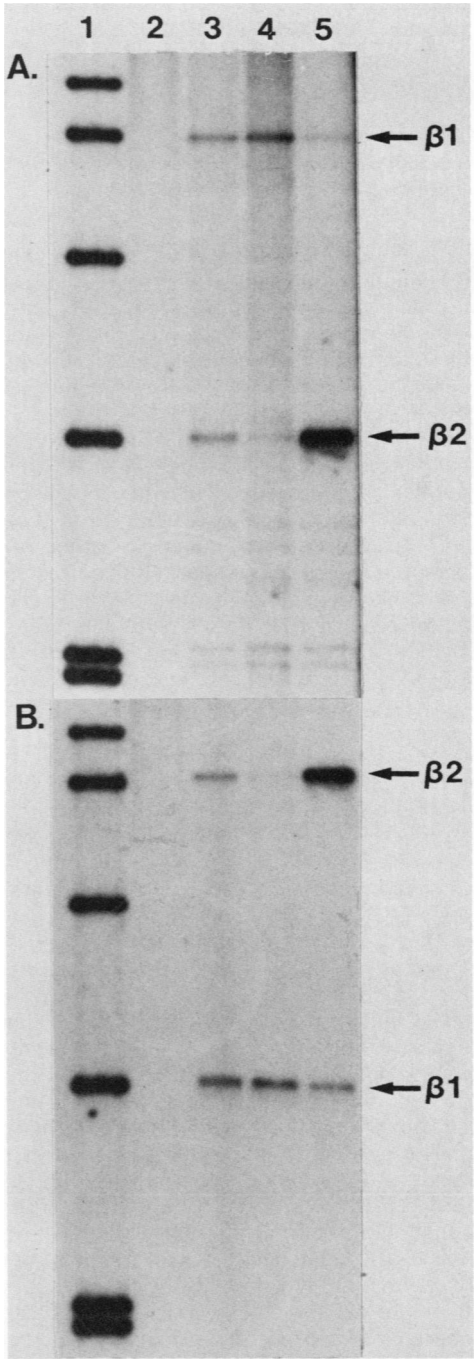


FIG. 6. Analysis of β1 and β2 transcript levels in different chicken cell types. S1 nuclease analysis was performed for a variety of chicken RNAs using gene-specific DNA probes labeled at 3′ ends at a site corresponding to codon 344 as described in Materials and Methods. For each probe, the protected fragment migrating at 310 bases corresponds to heterologous transcripts and the protected fragment at 530 bases corresponds to homologous transcripts as described in the text. Lane 1, Molecular weight standards: (from top to bottom) 622, 527, 404, 309, 242, and 238 bases, respectively. Lane 2, No RNA control. Lane 3, Secondary chicken fibroblast poly(A)$^+$ mRNA, 100 ng. Lane 4, Cultured embryonic chicken pectoralis myoblast poly(A)$^+$ mRNA, 100 ng. Lane 5, Cultured chicken embryonic sympathetic neuron poly(A)$^+$ mRNA, 100 ng. (A) β1 probe. (B) β2 probe.

TABLE 2. Expression of β1 and β2 transcripts

Cells	RNA abundance (arbitrary unit)		Ratio (β1/β2)
	β1	β2	
Fibroblasts	59.2	36.5	1.62
Myoblasts	61.0	8.3	7.3
Neurons	25.2	246.0	0.1

script detected (21). β2, on the other hand, is expressed in detectable quantities in many cell types but is particularly enriched in neuronal cells. Thus, although β1 and β2 share extensive homology in their promoter regions and encode essentially identical β-tubulin polypeptides, they display very different apparent modes of regulation during the differentiation of chicken tissue types.

In addition, the experiment shown in Fig. 6 also demonstrates that, at least in the cell types tested, β2 uses only the proximal polyadenylation site located 206 bases downstream of the stop codon. No signal was detected in the positions expected for transcripts terminated at either of the two downstream polyadenylation sites. In contrast, both the proximal and distal polyadenylation sites of β1 are utilized, generating mRNAs of 1,800 bases and 4,000 bases (21). As assayed by S1 mapping and RNA blotting experiments, both sites are used with approximately equal frequency in the different cell types tested (data not shown).

DISCUSSION

Comparison of the nucleotide sequences of two closely related β-tubulin genes from the chicken reveals that they are the products of a gene duplication event which took place at least 80×10^6 years ago. Since this duplication, the two genes have maintained virtual sequence identity in their coding regions by a mechanism involving segmental exchange of DNA sequences between the two loci. The result of this apparent coevolution has been the maintenance of two genes which encode essentially identical β-tubulin polypeptides. Since the predicted polypeptides share amino acid sequence identity in regions which have been shown to differ among β-tubulin isotypes (16, 20, 57, 58), they almost certainly constitute a single isotypic class of β tubulin. The two genes, however, show strikingly divergent regulatory behavior. The product of the duplication and evolution of these genes is thus a single structural product which has evolved different modes of cell type-specific regulation. A similar situation for two α tubulins can be observed by comparison of the sequence and expression of two human α-tubulin genes. In this instance, the two encoded polypeptides are again virtually indistinguishable (differing at only one amino acid position; compare the sequence of kα1 [14] with that of bα1 [19]), but the expression of one is restricted to neuronal and glial cells (19) whereas the other appears to be ubiquitously expressed (14).

In addition to localized selective pressures which derive from a functional constraint, a separate process mediated the maintenance of sequence homology between the two genes by acting discretely in the 3′ portions of the genes between a site within IVS2 and the 3′ end of the coding region. This conclusion is based on the different degrees of sequence divergence between the presumptively nonselected intervening sequences in the 5′ and 3′ homology domains (Fig. 4). Intervening sequences in the 5′ homology domain (IVS1 and the 5′ end of IVS2) appear to have diverged at the time of duplication, or at least 80×10^6 years ago. In contrast, the 3′ end of IVS2 and IVS3 shared sequence identity as recently as 20×10^6 to 25×10^6 years ago. This implies the action of a correction mechanism at some point(s) after duplication. Conceptually, such a correction process between nonallelic genes can be mediated by gene conversion, which is a nonreciprocal exchange of sequences between two genes that rectifies the sequence at the acceptor locus without altering the donor locus (e.g., references 39, 50). Gene conversion has been implicated in sequence exchange among different genes during the evolution of several multigene families such as the γ-globins (55), mouse histocompatibility genes (42, 61), immunoglobulins (3, 41), and chicken embryonic δ- and ε-globin genes (49). Alternatively, multiple nonequal recombination events between nonallelic loci can also generate a pattern of segmental sequence identity such as is observed between the β1 and β2 tubulin genes. The human adult α-globin genes are an example of the products of this mechanism (36).

Although sequence data alone are not sufficient to unambiguously define the molecular mechanisms of "concerted evolution" in vertebrate genomes, we favor the involvement of gene conversion in the concerted evolution of these two β-tubulin genes. The alternative process of nonequal crossing over would generate chromosomes with one or three genes instead of the original two, as has been documented for certain human α thalassemias (e.g., reference 36). We have not found any evidence of restriction site polymorphism which would be expected from such chromosomes in analysis of over 100 individual chicken genomes (D. Cleveland, unpublished data). However, we hasten to add that it is not clear that domestic chicken populations are sufficiently outbred to allow unambiguous interpretation of this result.

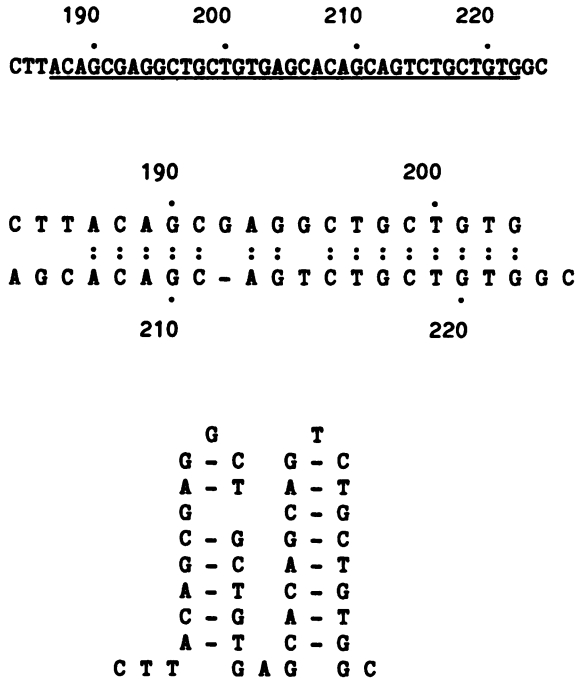


FIG. 7. Detail of the sequence found at the border of the 5′ and 3′ homology domains. The region from β2 is shown in detail. (Top) The linear sequence is shown with the region of direct and inverted symmetry underlined. (Middle) The direct repeat structure of this sequence is shown. (Bottom) The palindromic nature of this region of sequence is detailed to show potential hairpin loop structures.

Irrespective of the precise mechanism which has allowed these genes to evolve in concert, it is clear that it operated in a polar or segmental fashion. The junction between the two homology domains can be clearly located within IVS2, some 80 bp upstream of the 3′ splice junction (Fig. 4C and 5). The sequence adjacent to this boundary is not a "simple sequence" such as the $(GT)_n$ sequences found at the borders of human γ-globin gene conversion units (55), nor does it resemble the sequence found at the borders of mammalian α-globin conversion units (36, 54). Rather, as illustrated in detail in Fig. 7 for the region in β2, it has an unusual structure consisting of a tandemly repeated sequence of 17 bp, each of which forms a palindrome. Such regions of sequence symmetry have been proposed as sites for the initiation of gene conversion-associated strand transfer (56), and structurally similar, albeit more extensive, sequences have been previously observed at the termini of eucaryotic transposable elements (e.g., reference 7), retroviruses (23), and *Drosophila* foldback elements (46). In addition, short, nontandem direct and inverted repeats are associated with rearrangements within immunoglobulin loci (27, 33, 40). It seems likely, therefore, that this unusual sequence at the boundary of the converted region has served as a site for initiation of the recombination events associated with gene conversion. Further, the unique conversion of sequences 3′ to this initiation site suggests that it possesses an intrinsic polarity with respect to the directionality of the conversion-associated recombination process.

The coincidence of the 3′ boundary of homology with the translation stop codon is perplexing, since there are no unusual sequence features in this region which might suggest a possible basis for the termination of conversion-mediated sequence exchange. One possible scenario is that sequence exchange did not take place until these regions had diverged sufficiently to form a block to branch migration. Even small nonhomologies in human α-globin loci, for example, have been shown to block gene conversion or define boundaries of segmental conversion units (36). Alternatively, the original conversion event may have fortuitously ended in the observed region, and subsequent divergence reinforced this site as the termination site for homologous sequence interaction. Both the discrete site of initiation and the proposed mechanisms for termination of the conversion unit are consistent with the "polaron hypothesis" for gene conversion (43).

The two large segmental nonhomologies in IVS3 should serve as a block to isolate exon 4 from the conversion process, if indeed conversion was initiated at the site proposed above. However, exon 4 shows almost complete sequence identity. If the degree of divergence in the IVSs of the 3′ domain accurately reflects the time since the last conversion, then the divergence in silent codon positions in exon 4 would be expected to approach 15% instead of the observed 0.82%. To explain this discrepancy, we must postulate either that there is a mechanism which converts exon sequences without entering the IVSs, or that the silent codon positions in exon 4 are selected almost as stringently as replacement positions. We favor the latter interpretation, particularly in light of the strong codon usage bias in the genes which use C or G in the third codon position 9.6 times as frequently as A or T (60). This would also explain the discrepancy between divergence in silent codon positions and IVSs also seen in the 5′ domain (Table 1). Although we find this conclusion surprising, it is consistent with other studies which have found silent codon positions not to be selectively neutral (37). Thus, we postulate that the correction process which led to the synchronization of the 3′ sequences of β1 and β2 occurred most recently about 20×10^6 years ago and that subsequent divergence through segmental mutation in IVS3 has formed a block to further extensive sequence exchange.

The major functional consequence of the evolutionary interactions between the β1 and β2 genes has been the generation of two structurally equivalent β-tubulin genes which differ significantly in their regulatory properties. Since the two genes share extensive homology in the immediate 5′ flanking sequences, including the TATA transcriptional initiation signal, the factors which control the differential expression of β1 and β2 must lie elsewhere. These may be in promoter sequences 5′ to the homology regions or perhaps in sequences 3′ to the protein coding regions.

In any event, these results demonstrate that evolution of protein coding genes can occur at two levels which need not necessarily be coupled. At one level is the structural evolution of polypeptide coding regions, which evolve under different selective environments within multicellular organisms during the course of evolution, generating new modes of function or interaction for a system. At another level is the evolution of the control elements which regulate the expression of individual genes and integrate this expression into concerted programs of gene activity. The results presented here appear to document this sort of regulatory evolution, demonstrating the conversion of a protein sequence under the control of two different sets of regulatory signals. A converse situation has previously been seen among the chicken embryonic β-like globin loci, where conversion of control sequences has brought two different polypeptides under control of the same regulatory system (49). Collectively, these findings clearly suggest that the processes of gene conversion and intergenic recombination, which result in the concerted evolution of multigene families, may play an important role in evolution by allowing the coupling and uncoupling of structural genes and their regulatory elements.

ACKNOWLEDGMENTS

We thank D. Shortle and A. Efstradiatis for their critical reading of the manuscript and for their helpful suggestions. We also thank J. Havercroft for providing the chick mRNA samples used in the S1 analyses and F. Benko for excellent technical assistance.

This work has been supported by grants from the National Institutes of Health and March of Dimes to D.W.C., who is also the recipient of an National Institutes of Health Research Career Development Award. K.F.S. and J.T.Y.L. were both supported in part by National Institutes of Health postdoctoral fellowships.

LITERATURE CITED

1. **Alexandraki, D., and J. V. Ruderman.** 1984. Evolution of α- and β-tubulin genes as inferred by the nucleotide sequences of sea urchin cDNA clones. J. Mol. Evol. **19**:397–410.
2. **Baltimore, D.** 1981. Gene conversion: some implications for immunoglobulin genes. Cell **24**:592–594.
3. **Bentley, D. L., and T. H. Rabbitts.** 1983. Evolution of immunoglobulin V genes: evidence indicating that recently duplicated human V kappa sequences have diverged by gene conversion. Cell **32**:515–523.
4. **Berk, A., and P. Sharp.** 1978. Spliced early mRNAs of SV40. Proc. Natl. Acad. Sci. U.S.A. **75**:1274–1278.
5. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–383.
6. **Burnside, B.** 1973. Microtubules and microfilaments in amphibian neurulation. Am. Zool. **13**:989–1006.

7. **Calos, M., and J. H. Miller.** 1980. Transposable elements. Cell **20**:579–595.
8. **Chan, S. J., V. Episkopu, S. Zeitlin, S. K. Karathanasis, A. MacKrell, D. F. Steiner, and A. Efstratiadis.** 1984. Guinea pig preproinsulin gene: an evolutionary compromise? Proc. Natl. Acad. Sci. U.S.A. **81**:5046–5050.
9. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **24**: 5294–5299.
10. **Cleveland, D. W.** 1983. The tubulins: from DNA to RNA to protein and back again. Cell **34**:330–332.
11. **Cleveland, D. W., S. H. Hughes, E. Stubblefield, M. W. Kirschner, and H. H. Varmus.** 1981. Multiple α and β tubulin genes represent unlinked and dispersed gene families. J. Biol. Chem. **256**:3130–3134.
12. **Cleveland, D. W., M. A. Lopata, R. J. MacDonald, N. J. Cowan, W. J. Rutter, and M. W. Kirschner.** 1980. Number and evolutionary conservation of α and β tubulin and cytoplasmic β and γ actin genes using specific cloned cDNA probes. Cell **20**:95–105.
13. **Cleveland, D. W., and K. F. Sullivan.** 1985. Molecular biology and genetics of tubulin. Annu. Rev. Biochem. **54**:331–365.
14. **Cowan, N. J., P. Dobner, E. V. Fuchs, and D. W. Cleveland.** 1983. Expression of human α tubulin genes: interspecies conservation of 3′ untranslated regions. Mol. Cell. Biol. **3**: 1738–1745.
15. **Efstratiadis, A., J. W. Pasakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
16. **Farmer, S. R., J. F. Bond, G. S. Robinson, D. Mbangkollo, M. J. Fenton, and E. M. Berkowitz.** 1984. Differential expression of the rat β-tubulin multigene family, p. 333–342. *In* G. G. Borisy, D. W. Cleveland, and D. B. Murphy (ed.), The molecular biology of the cytoskeleton. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
17. **Fulton, C., and P. A. Simpson.** 1976. Selective synthesis and utilization of flagellar tubulin: the multitubulin hypothesis, p. 987–1005. *In* R. Goldman, T. Pollard, and J. Rosenbaum (ed.), Cell motility. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
18. **Goad, W. G., and M. I. Kanehisa.** 1982. Pattern recognition in nucleic acid sequences. I. A general method for finding local homologies and symmetries. Nucleic Acids Res. **10**:247–263.
19. **Hall, J. L., and N. J. Cowan.** 1985. Structural features and restricted expression of a human α-tubulin gene. Nucleic Acids Res. **13**:207–223.
20. **Hall, J. L., L. Dudley, P. R. Dobner, S. A. Lewis, and N. J. Cowan.** 1983. Identification of two β tubulin isotypes. Mol. Cell. Biol. **3**:854–862.
21. **Havercroft, J. C., and D. W. Cleveland.** 1984. Programmed expression of β tubulin genes during development and differentiation in the chicken. J. Cell Biol. **99**:1927–1935.
22. **Hayden, J. H., and R. D. Allen.** 1984. Detection of single microtubules in living cells: particle transport can occur in both directions along a single microtubule. J. Cell Biol. **99**:1785–1793.
23. **Hughes, S. H., P. R. Shank, D. H. Spector, H.-J. Jung, J. M. Bishop, H. E. Varmus, P. K. Vogt, and M. L. Breitman.** 1978. Proviruses of avian sarcoma virus are terminally redundant, co-extensive with unintegrated linear DNA and integrated at many sites. Cell **15**:1397–1410.
24. **Jones, C. W., and F. C. Kafatos.** 1982. Accepted mutations in a gene family: evolutionary diversification of duplicated DNA. J. Mol. Evol. **19**:87–103.
25. **Kalfayan, L., and P. T. Wensink.** 1981. α Tubulin genes of Drosophila. Cell **24**:97–106.
26. **Karfunkel, P.** 1974. The mechanisms of neural tube formation. Int. Rev. Cytol. **38**:245–271.
27. **Katakoa, T., M. Takashi, and H. Tasuku.** 1981. Repetitive sequences in class-switch recombination regions of immunoglobulin heavy chain genes. Cell **23**:357–368.
28. **Kemphues, K., T. C. Kaufman, R. A. Raff, and E. C. Raff.** 1982. The testis specific β tubulin subunit in Drosophila melanogaster has multiple functions in spermatogenesis. Cell **31**:655–670.
29. **Kirschner, M. W.** 1978. Microtubule nucleation and assembly. Int. Rev. Cytol. **54**:1–71.
30. **Krauhs, E., M. Little, T. Kempf, R. Hofer-Warbinek, W. Ade, and H. Ponstingl.** 1981. Complete amino acid sequence of β tubulin from porcine brain. Proc. Natl. Acad. Sci. USA **78**:4156–4160.
31. **Lee, M. G.-S., S. A. Lewis, C. D. Wilde, and N. J. Cowan.** 1983. Evolutionary history of a multigene family: an expressed human β tubulin gene and three processed pseudogenes. Cell **33**:477–487.
32. **Lopata, M. A., J. C. Havercroft, L. T. Chow, and D. W. Cleveland.** 1983. Four unique genes required for β tubulin expression in vertebrates. Cell **32**:713–724.
33. **Max, E. E., J. G. Seidman, and P. Leder.** 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin kappa constant region gene. Proc. Natl. Acad. Sci. USA **76**:3450–3454.
34. **McClain, D. A., P. D'Eustachio, and G. M. Edelman.** 1977. Role of surface modulating assemblies in growth control of normal and transformed fibroblasts. Proc. Natl. Acad. Sci. USA **74**:666–670.
35. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either strand of double-digested restriction fragments. Gene **19**:269–276.
36. **Michelson, A. M., and S. H. Orkin.** 1983. Boundaries of gene conversion within the duplicated human α-globin genes—concerted evolution by segmental recombination. J. Biol. Chem. **258**:15245–15254.
37. **Modiano, G., G. Battistuzzi, and A. G. Motulsky.** 1981. Nonrandom patterns of codon usage and of nucleotide substitutions in human α- and β-globin genes: an evolutionary strategy reducing the rate of mutations with drastic effects? Proc. Natl. Acad. Sci. USA **78**:1110–1114.
38. **Murphy, D. B., and L. G. Tilney.** 1974. The role of microtubules in the movement of pigment granules in teleost melanophores. J. Cell Biol. **61**:757–779.
39. **Nasmyth, K. A.** 1982. Molecular genetics of yeast mating type. Annu. Rev. Genet. **16**:439–500.
40. **Nikaido, T., S. Nakai, and H. Tasuku.** 1981. Switch region of immunoglobulin Cμ gene is composed of simple tandem repetitive sequences. Nature (London) **292**:845–848.
41. **Ollo, R., and F. Rougeon.** 1983. Gene conversion and polymorphism: generation of mouse immunoglobulin gamma 2a chain alleles by differential gene conversion by gamma 2b chain gene. Cell **32**:515–523.
42. **Pease, L. R., D. H. Schultze, G. M. Pfaffenbach, and S. G. Nathenson.** 1983. Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major histocompatibility complex. Proc. Natl. Acad. Sci. USA **80**:242–246.
43. **Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dodgson.** 1980. The evolution of genes: the chicken preproinsulin gene. Cell **20**:555–565.
44. **Poncz, M., D. Solowiejczyk, M. Ballantine, E. Schwartz, and S. Surrey.** 1982. 'Nonrandom' DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc. Natl. Acad. Sci. USA **79**:4298–4302.
45. **Ponstingl, H., E. Krauhs, M. Little, and T. Kempf.** 1981. Complete amino acid sequence for α tubulin from porcine brain. Proc. Natl. Acad. Sci. USA **78**:2757–2761.
46. **Potter, S. S.** 1982. DNA sequence of a foldback transposable element in *Drosophila*. Nature (London) **297**:210–214.
47. **Radding, C. M.** 1982. Homologous pairing and strand exchange in genetic recombination. Annu. Rev. Genet. **16**:405–437.
48. **Raff, E. C.** 1984. The genetics of microtubule systems. J. Cell Biol. **99**:1–10.
49. **Roninson, I. B., and V. M. Ingram.** 1982. Gene evolution in the chicken β-globin cluster. Cell **28**:515–521.
50. **Rossignol, J.-L., N. Paquette, and A. Nicolas.** 1978. Aberrant 4:4

asci, disparity in the direction of conversion and frequencies of conversion in *Ascobolus immersus*. Cold Spring Harbor Symp. Quant. Biol. **43**:1343–1352.

51. **Sanchez, F., J. E. Natzle, D. W. Cleveland, M. W. Kirschner, and B. J. McCarthy.** 1980. A dispersed multigene family encoding tubulin in *Drosophila melanogaster*. Cell **22**:845–854.
52. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–169.
53. **Sanger, F., S. Nicken, and A. R. Coulsen.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
54. **Schon, E. A., S. M. Wernke, and J. B. Lingrel.** 1982. Gene conversion of two functional goat α globin genes preserves only minimal flanking sequence homology. J. Biol. Chem. **257**:6825–6835.
55. **Slightom, J. L., A. E. Blechl, and O. Smithies.** 1980. Human fetal Gγ- and Aγ-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**:627–638.
56. **Sobell, H. M.** 1975. A mechanism to activate branch migration between homologous DNA molecules in general recombination. Proc. Natl. Acad. Sci. USA **72**:279–283.
57. **Sullivan, K. F., and D. W. Cleveland.** 1984. Sequence of a highly divergent β tubulin gene reveals regional heterogeneity in the β tubulin polypeptide. J. Cell Biol. **99**:1754–1760.
58. **Sullivan, K. F., J. C. Havercroft, and D. W. Cleveland.** 1984. Primary structure and expression of a vertebrate β tubulin gene family, p. 321–332. *In* G. G. Borisy, D. W. Cleveland, and D. B. Murphy (ed.), Molecular biology of the cytoskeleton. Cold Spring Harbor Press, Cold Spring Harbor, N.Y.
59. **Szostak, J. W., and R. Wu.** 1980. Unequal crossing over in the ribosomal DNA of *Sacchromyces cerevisiae*. Nature (London) **284**:426–430.
60. **Valenzuela, P., M. Quiroga, J. Zaldivar, W. J. Rutter, M. W. Kirschner, and D. W. Cleveland.** 1981. Nucleotide and corresponding amino acid sequences encoded by α and β tubulin mRNAs. Nature (London) **289**:650–655.
61. **Weiss, E. H., A. Mellor, L. Golden, K. Fahrner, E. Simpson, and J. Hurst.** 1983. The structure of a mutant H-2 gene suggests that the generation of polymorphism in H-2 genes may occur by gene conversion-like events. Nature (London) **301**:671–674.
62. **Wilbur, W. J., and D. J. Lipman.** 1983. Rapid similarity searches of nucleic acid and protein data banks. Proc. Natl. Acad. Sci. USA **80**:726–730.
63. **Yousaf, S. I., A. R. Carroll, and B. E. Clarke.** 1984. A new and improved method for 3′-end labelling DNA using [α-^{32}P]ddATP. Gene **27**:309–313.