A survey of the α -tubulin gene family in chicken: unexpected sequence heterogeneity in the polypeptides encoded by five expressed genes

Lisa F.Pratt¹ and Don W.Cleveland

Department of Biological Chemistry, The Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205, USA

¹Present address: Division of Clinical Immunology, Scripps Clinic and Research Foundation, 10666 N. Torrey Pines Road, La Jolla, CA 92037, USA

Communicated by K.Weber

To characterize the α -tubulin gene family in chicken, we have isolated five chicken α -tubulin genes and determined the majority of the sequences of the encoded polypeptides. Three of these ($c\alpha 3$, $c\alpha 5/6$ and $c\alpha 8$) encode novel, expressed α -tubuling that have not previously been analyzed, whereas one gene segment is a pseudogene and another appears capable of encoding a functional subunit (although we were unable to document its expression in a survey of chicken tissues). Together with two additional expressed, functional α -tubulins reported earlier, we conclude that the chicken α -tubulin family is comprised of at least five functional genes whose polypeptide products are substantially more heterogeneous than found in preceding analyses of vertebrate α -tubulins. Comparison of the amino acid sequences reveals that the five polypeptides are between 96 and 83% identical, with the extreme carboxy-terminal residues representing a highly heterogeneous variable domain. Since some α -tubulins undergo cyclic post-translational removal and readdition of a carboxy-terminal tyrosine, the notable sequence divergence in this domain suggests that individual tubulins probably participate to different extents in this modification cycle.

Key words: α -tubulin/chicken/gene family/polypeptide sequence heterogeneity

Introduction

In the 16 years since α - and β -tubulin were first determined to be the principal subunits of eukaryotic microtubules, it has become abundantly clear that these polymers contribute at least two essential functions to all higher eukaryotes. First, microtubules are responsible for accurate chromosome segregation at cell division. Second, in concert with actin filaments and intermediate filaments, microtubules play a major role in establishing and maintaining the dynamic spatial organization of the cytoplasm (Heuser and Kirschner, 1980). In addition, microtubules also participate in a spectrum of specialized roles as major components of eukaryotic cilia and flagella (Gibbons, 1975), in establishing the highly asymmetric morphology of neurons, and in transport of vesicles and organelles within the cytoplasm (Hayden and Allen, 1984; Vale *et al.*, 1985).

Analysis of the molecular genetics of tubulin has lead to recognition that in higher eukaryotes small multigene families encoding both α - and β -tubulin lay beneath this heterogeneity of microtubule utilization (see Cleveland and Sullivan, 1985; Cleveland, 1987). For β -tubulin, extensive analyses in chicken (Sullivan and Cleveland, 1984; Sullivan *et al.*, 1985, 1986a,b; Murphy *et al.*, 1987; Montiero and Cleveland, 1988), mouse (Lewis *et al.*, 1985a; Sullivan and Cleveland, 1986; Wang *et al.*, 1986) and human (Hall *et al.*, 1983; Lee *et al.*, 1983; Lewis *et al.*, 1985b; Sullivan and Cleveland, 1986) have revealed six functional vertebrate β -tubulins. At least four [and probably five—see Lopata and Cleveland (1987)] of these represent distinct β -tubulin isotypes whose sequences (particularly within a carboxy-terminal variable domain) have been evolutionarily conserved (Sullivan and Cleveland, 1986; Wang *et al.*, 1986).

For α -tubulin, the analyses are less complete. In part this is the consequence of an abundant number of tubulin pseudogenes that are present in mammalian genomes (e.g. Lemischka and Sharp, 1982; Wilde et al., 1982; Lee et al., 1983). In fact, despite reports of cDNA clones for several α -tubulins (Cleveland et al., 1980; Ginzburg et al., 1981; Lemischka et al., 1981; Cowan et al., 1983; Lewis et al., 1985a; Elliott et al., 1986; Villasante et al., 1986; Dobner et al., 1987; Pratt et al., 1987), analysis of only one rat α -tubulin (Lemischka and Sharp, 1982) and two human α -tubulin (Hall and Cowan, 1985; Dobner et al., 1987) genes have been reported. To surmount such problems, Cowan and collaborators have reported an extensive analysis of mouse α tubulins by isolation of cDNA clones (Lewis et al., 1985a; Villasante et al., 1986). Six functional genes encoding five different polypeptides have been identified by isolation of cloned cDNA sequences from brain and testis libraries. Whether these represent a complete set of mouse α -tubulins is not certain inasmuch as the presence of numerous pseudogenes precludes simple genomic analysis, and the possibility of α -tubulins that are not expressed in the tissues from which the cDNA libraries were constructed cannot be excluded.

In an attempt to identify a complete set of α -tubulins encoded in a vertebrate genome, we have now analyzed α tubulin genes encoded in the chicken genome. We initially chose this approach for two reasons. First, no evidence for tubulin pseudogenes had been observed for this genome [indeed, unlike mammalian genomes, few chicken pseudogenes for any protein had been documented (Alsip and Konkel, 1986; Robins et al., 1986)]. Second, genomic blots had indicated the presence of only four or five gene segments for α -tubulin (Cleveland et al., 1980). From isolation and analysis of five chicken α -tubulin genes, we now show that three encode functional α -tubulins, one is an unexpected pseudogene and one appears capable of encoding a functional α -tubulin, although we could not demonstrate that it is actually expressed. Combined with two α -tubulins identified by cDNA cloning (Valenzuela et al., 1981; Pratt et al., 1987), a picture for the chicken α -tubulin gene family emerges in which at least five functional genes encode five

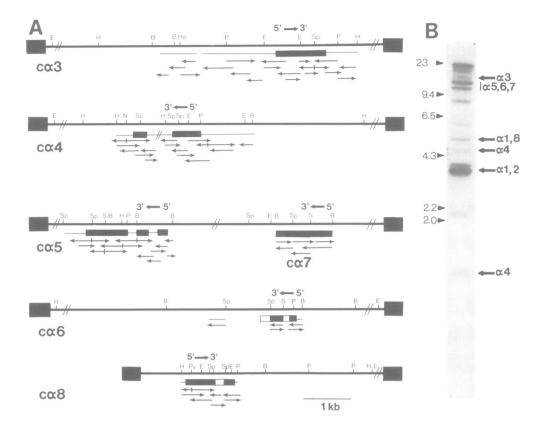


Fig. 1. Characterization of six α -tubulin gene segments. (A) Schematic diagrams of genes $c\alpha 3 - c\alpha 8$. Partial restriction maps of each of five recombinant Charon 30 phages carrying the $c\alpha 3 - c\alpha 8$ genes are shown. The thick black lines beneath each map represent the coding regions of the corresponding α -tubulin gene segment, whereas the thin lines denote the regions from which doubled stranded DNA sequence was obtained. The arrows at the botton of each gene display the sequencing strategy with which the sequence was determined. E, *Eco*RI; H, *Hind*III; B, *Bam*HI; P, *Pst*I; Sp, *Sph*I; S, *Sal*I; N, *Nde*I, Pv, *Pvu*II. (B) The chicken α -tubulin gene family visualized on a blot of genomic DNA digested with *Hind*III. Genomic chicken DNA (5 μ g) was digested with *Hind*III, blotted as in Materials and methods and hybridized to a ³²P-labeled probe corresponding to the coding sequences from c α 1. Parallel blots were probed with labeled probes prepared from the subcloned genomic fragments of $c\alpha 3 - c\alpha 8$ (see Materials and methods). Resultant assignments of specific fragments containing α -tubulin sequence homology to individual genes is shown at the left.

distinct polypeptide isotypes. These isotypes are remarkably divergent particularly in their carboxy-terminal domains, but unlike the β -tubulin family these do not fall into obvious classes that have been conserved among vertebrates.

Results

Isolation and characterization of λ bacteriophage carrying six $\alpha\text{-tubulin gene segments}$

Previously, blot analysis of genomic DNA has been used to demonstrate the presence in the chicken genome of at least four or five segments that appeared to contain complete α tubulin genes (Cleveland et al., 1980, 1981). To analyze this gene family in detail, 1.6×10^6 recombinants from a genomic library constructed in Charon 30 (library complexity 4×10^{5}) were screened by DNA hybridization for DNA segments homologous to $c\alpha 1$, a cDNA clone carrying most of the coding region of a widely expressed α -tubulin (Pratt. 1987). Of 67 positive phage, 20 were plaque purified and DNA prepared. Preliminary restriction maps revealed that the cloned fragments represented five distinct genomic segments, one of which contained two α -tubulin genes separated by 2 kb. A partial restriction map of five representative phages containing these six α -tubulin segments is shown in Figure 1A. The regions that contained α -tubulin sequence homology (named $c\alpha 3 - c\alpha 8$) were determined by

hybridization to $c\alpha 1$ sequences and were subcloned into plasmid vectors (see Materials and methods). Blot hybridization with probes corresponding to the amino-terminal codons (amino acids 41-133) or the carboxy-terminal codons (amino acids 376-453) established the apparent direction of transcription of each gene and further suggested that all six segments contained complete α -tubulin-coding regions. Hybridization to 3' untranslated region probes from the cDNA clones of $c\alpha 1$ (Valenzuela *et al.*, 1981) and $c\alpha 2$ (Pratt *et al.*, 1987) demonstrated that none of the genes encoded $c\alpha 1$ or $c\alpha 2$ (not shown).

Blot analysis of genomic DNA hybridized to a probe containing a nearly complete α -tubulin-coding segment revealed 13 fragments that were distinguishable by size in DNA digested with *Hin*dIII (Figure 1B). To ascertain what proportion of the genomic fragments with high sequence homology to α -tubulin were represented in our six genes and in the c α 1 and c α 2 cDNAs, DNA segments were prepared from each cloned gene/cDNA and used as probes of parallel genomic blots. (Each probe contained not only α -tubulincoding regions but also adjacent flanking sequences so that preferential hybridization to the parental gene would be observed.) The analysis yielded the assignments shown in Figure 1B (for details see Pratt, 1987). In almost all cases, the size of the authentic genomic fragment(s) was indistinguishable from the corresponding fragment in the clon-

1:	26				130										140																									ACT
		Asp	Gln	Cys			Leu	Gln	Gly	Phe	Leu	Val	Phe	His		Phe	G1 v	Gly	Glv	Thr	Glv	Ser	61 v	Phe	150 Thr	Ser	1 au	Len	Mat	61 11	4=0	1.00	Sen	Ve 1	160			1	1	
G	CT	GAC	CAG	TGC	AGT	GGC	CTC	CAG	GGG	TTC	CTG	GTC	TTC	CAC	AGC	TTC	GGG	GGC	GGC	ACA	GGC	TCG	GGG	TTC	ACC	TCC	CTC	CTC	ATG	GAG	000	CTC	TOT	CT A	GAG	TAC	100	AAG		с т
G	СТ	GAT	GCC	TGT	TCT	GGA	CTG	CAG	GGA	TTC	CTC	ATC	TTC	CAC	AGC	TTT	GGA	GGC	GGC	ACC	GGC	TCT	GGA	TTT	ACC	TCT	TTG	TTG	ATG	GAG	COC	CTC	TCT	GTG	GAT	TAT	GGA	AAG	AA	Ă T
			Ala		Ser							Ile																							Asp		G1 y			
		1	C 1	Dhe	170				•					_	180				_	_		_		_	190										200					
Ň	ÅG I	CTG	GAG	TTC	TCC	GTC	TAC	CCA	GAC	CGT	CAG	GTA	TCC	ACA	ALB	CT.	CTC	Glu GAG	Pro	Tyr	Asn	Ser	ATC	Leu	Thr	Thr	HIS	Thr	Thr	CTC	Glu	His	Ser	Asp	Cys	Ser	Phe	Met	Val	
N		CTA	GAG	TTT	GCC	ATC	TAC	CCA	GCT	CCT	CAG	GTC	TCC	ACT	GCT	GTG	GTA	GAA	ccc	TAC	AAT	TCC	ATC	CTT	ACC	ACA	CAC	ACC	ACT	CTG	GAA	CAC	TCA	GAC	TGT	OCC.	TTC	ATG	GTO	G (
					Ala	Ile			Ala	Pro																										Ala				
		~ `			210										220										230										240					
**	an Ac	GIU	Ala	11e	Tyr	Asp	Ile	Cys	Asn	Arg	Asn	Leu	Asp	Ile	Glu	Arg	Pro	Thr	Tyr	Thr	Asn	Leu	Asn	Arg	Leu	Ile	Gly	Gln	Ile	Val	Ser	Ser	Val	Thr	Ala	Ser	Leu	Arg	Phe	8 1
ñ	AT	GAG	GCT	ATA	TAT	GAC	ATC	TGT	TGC	AGA	AAC	CTA	GAC	ATT	GAG	CGC	CCA	ACA ACG	TAC	ACC	AAC	CTC	AAC	AGG	CTC	ATT	GGG	CAG	ATT	GTC	TCC	TCT	GTT	ACT	OCC.	TCT	TTA	AGA	TT	: :
									Cys			•••			unn		COA	AUU	140		AAC	010			010		Ser	CAU		010	ICC		Ile	ALL	GUA	100	110	COC	111	1
					250										260										270										280					
G.	ly CT	Ala	Leu	Asn	Val	Asp	Leu	Ile	Glu	Phe	Gln	Thr	Asn	Leu	Val	Pro	Tyr	Pro	Arg	Ile	His	Phe	Pro	Leu	Thr	Thr	Tyr	Ala	Arg	Ile	Ile	Ser	Ala	Glu	Lys	Ala	Tyr	His	Glu	a.
G	GT	GCC	CTC	AAC	GTG	GAT	CTG	ACA	GAA	TTC	CAG	ACC	AAC	CTG	GTG	CCC	TAC	CCA CCA	CGG	ATA	CAC	TTC	CCC	CTC	ACC	ACC	TAT	OCA	CGC	ATC	ATC	TCG	GCA	GAG	***	000	TAC	CAC	GAC	3
								Thr						••••							U.L.U				Val	AUU			Pro	AIC	A1 1		Ser	UAA	Arg	ULA	141	LAL	un ('
					290										300										310										320					
	eu -	Ser	Val	Pro	Glu	Ile	Thr	Asn	Ala	Cys	Phe	Glu	Phe	Ser	Asn	Gln	Met	Val	Lys	Cys	Asp	Pro	Arg	Arg	Gly	Lys	Tyr	Met	Ala	Cya	Cys	Leu	Leu	Tyr	Arg	Gly	Asp	Val	Va]	L
C1	TC	TCG	OTG	OCT	GAG	ATC	ACC	AAT	GCT	TGC	TTT	GAG	TTC	TCC	AAC	CAG	ATG	GTG GTG	***	TGT	GAC	CCA	CGG	CGT	GGC	AAG	TAC	ATG	OCG	TOC	TOC	CTG	CTG	TAC	CGG	GGT	GAT	GTG	GTO	3
				Ala				Ser	Ser		•••		Pro	Asn	ANU	UNG			лла		unu			His	000	AAU		AIU	001	100		Net		1.81	CGI	601	GAI	GTA	Gri	•
					330										340										350										360					
Ŀ	ya .	Asp	Val	Asn	Ala	Ala	Ile	Ala	Ala	Ile	Lys	Thr	Arg	Arg	Ser	Ile	Gln	Phe	Val	Asp	Trp	Cys	Pro	Thr	Gly	Phe	Lys	Val	Gly	Ile	Asn	Tyr	Gln	Pro	Pro	Thr	Val	Val	Pro	,
		GAT	GTC	AAC	GUG	GCL	ATT	OCG	GCC	ATC	AAA	ACC	CGC	CGC	TCT	ATC	CAG	TTT TTT	GTG	GAC	TGG	TGC	CCC	ACG	GGC	TTC	AAG	GTG	GGC	ATC	AAC	TAT	CAG	CCC	CCA	ACG	GTG	GTG	CCT	2
_					Val	GUA				AIC	ANG	ACC	Lys		Asn	AIC	CAU		911							Val								_			GTA	GTT	CCI	
																				GTGA	GAAT	GCCA	TGAG	ATTT	ACTT		∿77 co#i			ACTT	CTCT	CATA	TTCC	TACT	TCCA	Ū.				
C1	1.		1	41.	370	Ve 1	61 m	470	41.	V-1	· · · -	Mat	1		380		m b		•••		.				390									_	400					
G	GG	GAC	CTG	GCC	AAG	GTG	CAG	CGG	OCT	GTG	TGC	ATG	CTG	AGC	ASD	ACC	ACG	Ala GCC	ATC	A1a OCC	GAG	ALA	TCC	Ala OCA	Arg	CTC	ASP GAC	His	Lys :	Phe	Asp	Leu	Met	Tyr	Ala	Lys	Arg	Ala	Phe)
GC	GA	GAC	TTA	OCC	CAG	GIT	CAG	CGA	GCT	CTC	TOC	ATG	CTC	AGC	AAC	ACC	ACA	GCC	ATT	GCA	GAG	GCC	TGG	GCA	AGG	CTC	GAC	CAC	AAG	TTT	GAT	CTT	ATG	TAT	GCC	AAG	AGA	000	TTC	•
					Gln																																			
		_	_		410										420										430										440					
E1 CI		Trp	Tyr	Val	Gly	Glu	Gly	Met	Glu	Glu	Gly	Glu	Phe	Ser	Glu	Ala	Arg	Glu GAG	Asp	Leu	Ala	Ala	Leu	Glu	Lys .	Asp	Tyr	Glu	Glu	Val	Gly	Arg	Asn	Ser	Ala	Asp	Gly	G1 y	Glu	ł
či	AT :	TGG	TAT	GTA	AGT	GAA	GGC	ATG	GAG	GAA	GGA	GAA	TTT	GCA	GAG	GCC	CGA	GAG	GAC	CTG	GCT	GCC	CTG	GAG	AAG	GAT '	TAT TAT	GAG	GAG GAA	OTT CTC	GGA	AGG	GAC	TCA	GCA	GAT	GGG	GGG	GAG	Ľ
					Ser									Ala													••••					Thr				Glu				
			44.	8																																				
											-																													
	UAG	GAA	TA TA	A CT	CAC	TAT	Nith	GGCI	GCTT	GGGI	Tete	NIATI	GCTT	TTG	ATCT	TTGI	CTTC	TATA	GGAG	ICT GG	AAGO	GAGT	TAGG	GGCT	GGAT	TTGA	TGCA	GTGT	GACT	GGCA	AACT	CCAC	AGAT	TGGC	CAGA	AGCA	ACTO	TTTC	CT A	σ

Fig. 2. The DNA and inferred encoded polypeptide sequences for genes α^3 and α^4 . The sequence corresponding to a portion of intron 3, followed by all of the coding sequences from codon 126 through to the translation termination codon for both α^3 and α^4 were determined on both strands by the strategies diagrammed in Figure 1A. The asterisk (*) represents the translation termination codon. A putative consensus signal for polyadenylation in the α^3 sequence is underlined. The predicted polypeptide encoded by α^4 is shown only where it differs from α^3 . A fourth intervening sequence (~800 bases in length) disrupts the α^4 sequence between codons 352 and 353. The 5' and 3' sequences (not shown) of this intron conform to the consensus signals of 5' and 3' splice junctions (5' GTGAGAA . . . and 3' . . . CCTACTTCCAG).

ed gene, veryifying that the cloned segment did not contain a major cloning artefact. In one instance, however, the map of the cloned $c\alpha 3$ gene predicted that a $c\alpha 3$ *Hin*dIII fragment of ~6 kb should arise on the genomic blot, whereas in fact a much larger fragment is observed (Figure 1B). While this might indicate a cloning artefact, we believe that it is due to restriction site polymorphism in the chicken population because four independent phage-carrying portions of the $c\alpha 3$ gene (each derived from a different cloning event) possess the same structure within the regions of overlap.

We conclude that the six genomic recombinants in conjunction with the $\alpha 1$ and $\alpha 2$ cDNA clones appear to contain ~75% of the chicken genomic segments that have high nucleotide sequence homology to cloned α -tubulin cDNAs. Thus, while it is clear that analysis of these eight segments does not comprise the entirety of the chicken α -tubulin gene family, analysis of them should provide insight into the majority of α -tubulin gene complexity in this organism.

Analysis of $c\alpha 3$: an expressed α -tubulin gene without a coded carboxy-terminal tyrosine

The DNA sequence of the region of clone $c\alpha 3$ that is homologous to previously isolated α -tubulin-coding segments was determined according to the strategy diagrammed in Figure 1A. As shown in Figure 2, this yielded an open reading frame of 966 bases that, when translated into protein and compared with the known sequence of other α tubulins, corresponded to amino acid positions 126 through to the translation termination codon (at position 448). This coding sequence was followed by a putative 3' untranslated region that contained a consensus sequence for polyadenylation 152 bases 3' to the translation termination codon. By analogy to the structure of three previously characterized vertebrate α -tubulin genes (in which three introns interrupt the coding sequence at amino acid positions 1/2, 76 and 125/126), we presume that the c α 3 sequence corresponds to exon 4 of this gene.

Unfortunately, efforts to locate sequences corresponding to exons 1-3 were unsuccessful. Despite the fact that the $c\alpha$ 3 clone contained 9 kb of chicken sequence 5' to codon 126, hybridization analysis with probes containing the majority of exons 2 and 3 [from either $c\alpha$ 1 or from a human α -tubulin cDNA, $k\alpha$ 1 (Cowan *et al.*, 1983)] failed to detect the presence of the upstream exons. Since it was still possible that the upstream exons were present but sufficiently divergent to preclude their detection by hybridization, we also searched for the upstream exons by determining the sequence of 2 kb upstream of codon 126 (see Figure 1A). However, no sequences corresponding to those expected for exons 2 and 3 were found. These data indicate that $c\alpha$ 3 either

ca 5 GGATCCCCGGCTGATCCCCCAGGTGCGACCGCGGCGGGCG	AAGAGGGGGATGGGCTCACCGTCGGGCCCGGGCGCTATATTTAGCGGCGCGGGGCCCTTCGTCTCGCCGGGATG
ca5 ACCCTCTCCCTCTCCCCGCAGCTGCAATCCTGGCGCACCCCCAGCCCTGCCGATGCCGCGCGCG	GCAGGCTTTGCAGCGCGGACGTGGCCCGTAGCCCTCAGCGATGCCTAGTCTGATCCTGCCTCCCCCAG
2 Arg Glu Cys Ile Ser Val His Val Gly Gln Ala Gly Val Gln Met Gly Asn Thr Cys Trp Glu I	40 eu Tyr Cys Leu Glu His Gly Ile Gln Pro Asp Gly Gln Met Pro Ser Asp Lys Thr.
COS CGC GAG TGC ATC TCC GTC CAC GTC GGC CAG GCC GGC GTC CAG ATG GGC AAC ACC TGC TGG GAG C	TTG TAC TGC CTG GAG CAC GGC ATC CAG CCC GAC GGG CAG ATG CCC AGC GAC AAG ACC
50 Ile Gly Gly Asp Asp Ser Phe Thr Thr Phe Phe Cys Glu Thr Gly Ala Gly Lys His Val B co5 ATC GGT GGG GGG GAC GAC TCC TTC ACC ACT TTC TTC TGT GAG ACC GGC GCG GAA G CAC GTC G	70 76 Pro Arg Ala Ile Phe Val Asp Leu Glu Pro Thr Val Ile A
ca5 ATC GGT GGG GGG GAC GAC THE THE ACC ACT THE THE THE THE THE GAC GCC GGC GCC GGG GAC GAC GAC GAC GAC	
ca5 ITGTGGCACGGGGGGAGAAACCAGCCGGCCGGCCGGCGCGCGGCG	AGCTGAGCTCCCGTGCAGGGAGGCGCACAGACTCCGCTGGCAGAGAAACCTGGCACCGGCACTGCATGGCCCCACTG
	76 80 sp Clu Val Arg Ala Cly Ile Tyr Arg Gln Leu Phe
$\texttt{ca5} \ acgtacggtctggcaggaggaggaggaggaggagggaggg$	CAGCCTCCTTTGCTCTCTCCTAGAT GAG GTG CGG GCT GGG ATC TAC CGG CAG CTC TTC
90 His Pro Glu Gln Leu Ile Thr Gly Lys Glu Asp Gly Ala Asn Asn Tyr Ala Arg Gly His Tyr 1	110 125 hr Ile Gly Lys Glu Ile Ile Asp Gin Val Leu Asp Arg Ile Arg Lys Leu
Cu5 CAC CCC GAG CAA CTC ATC ACT GGC AAG GAG GAT GGT GCC AAC AAC TAT GCC CGT GGG CAC TAC A	ICC ATC GGC AAG GAG ATC ATC GAC CAA GTG CTG GAC AGG ATC CGG AAG CTG GTGAGTA
Ca5 ACCACAAACTGGGATCACATGTCTCTGGGTGGGCTAGGTCGTGGTGCTGGGGCTCGGAGCCTGATTGTGCCAAGAGAGGGATTTGG	CAGCTGTCCGTGTCAGGGATGGAACAAGTCCTGAAGGACTTCCAGAGTCAGATGTTGACAGGAATATCTGTGGC
126 130 Ala Asp Gln Cys Thr Gly Leu Gln Gly Phe Leu Val Phe F	140 150 Iis Ser Phe Gly Gly Gly Thr Gly Ser Gly Phe Thr Ser Leu Leu Met Glu Arg Leu
Ca5 CAGAGCAATCTCATGTTTTTCTCTCTCTGCAG GCT GAC CAG TGC ACG GGC CTC CAA GGC TTC CT TGT TTC C Ca7ACGCGCCACTTACCTCTCTCTCTCTGCAG GCG GAC CAG TGC ACA GGC CTC CAA GGC TTC CTT GTG TTC C	CAC AGT TTT GGA GGT GGC ACT GGC TCA GGA TTC ACC TCC CTG ATG GAA CGA CTC CCC AGT TTT GGA GGT GGC ACT GGC TCA GGA TTC ACC TCC CTG ATG GAA CGA CTC
160 170	ing 180 190
Ser Gly Asp Tyr Gly Lys Lys Ser Lys Leu Glu Phe Ser Ile Tyr Pro Ala Pro Gln Val Ser T ca5 TCC GGT GAC TAT GGC AAG AAG TCC AAG CTT GAG TTC TCC ATC TAC CCG GCA CCA CAG GTC TCC A	ACT GCT GTG GTG GAG CCC TAC AAC TCC ATC CTC ACC ACC CAC ACC CTG GAG CAC
CA7 TCC GTT GAC TAT GGC AAG AAG TCC AAG CTC GAG TTC TCC ATC TAC CCG GCC CCA CAG GTC TCC A Val	Ser
200 Ser Asp Cys Ala Phe Met Val Asp Asn Glu Ala Ile Tyr Asp Ile Cys Arg Arg Asn Leu Asp I	220 230 le Glu Arg Pro Thr Tyr Thr Asn Leu Asn Arg Leu Ile Ser Gln Ile Val Ser Ser
COS TCA GAC TOT GCC TTC ATG GTG GAC AAC GAG GCC ATC TAC GAC ATC TGC CGC AGG AAC CTG GAC A COT TCA GAC TGT GCC TTC ATG GTG GAC AAC GAG GCC ATC TAC GAC ATC TGC CGC AGG AAC CTG GAC A	TO GAG CGC CCA ACC TAC ACC AAC CTC AAC CGC CTC ATC AGT CAG GTC GTC TCC TCC
240 250	Val 260 270
Ile Thr Ala Ser Leu Arg Phe Asp Gly Ala Leu Asn Val Asp Leu Thr Glu Phe Gln Thr Asn L co5 ATC ACC GCC TCT CTG AGG TTC GAT GGG GCA CTC AAT GTC GAC CTC ACC GAG TTC CAG ACC AAC	TG GTG CCC TAC CCT CGC ATC CAC TTC CCC CTG GCC ACC TAC GCC CCT GTC ATC TCT
CAT ATC ACC GCC TCT CTG AGG TTC GAT GGG GCA CTC AAC GTC GAC CTC ACC GAG TTC CAG ACC AAC C	Val
280 Ala Glu Lya Ala Tyr His Glu Gln Leu Ser Val Ala Glu Ile Thr Asn Ser Cys Phe Glu Pro A cn5 GCA GAA AAA GCC TAC CAC GAG GAG CTG TG GG GTG GGG ATC ACC AAC TCC TGC TTC GAG CCG G	la Asn Gln Met Val Lys Cys Asp Pro Arg His Glv Lys Tyr Met Ala Cys Cys Leu
cas gea gaa aaa gee tae cae gag eng eig ieg gig get gag ate ace aac teg toe the the gag eeg g ca7 gee gag agg get tat cat gag eag etg tea gtg get gag ate ace aac teg tge tte gag eeg g	CC AAC CAG ATG GTG AAG TGC GAC CCT CGG CAC GGC AAG TAC ATG GCG TGC TGC CTG
320 Leu Tyr Arg Cly Asp Val Val Pro Lys Asp Val Asn Ala Ala Ile Ala Thr Ile Lys Thr Lys A	340 350 rg Ser Ile Gin Phe Vel Asn Trn Cys Pro Thn Giy Phe Lyn Vel Ciy Ile Asn Tyn
ca5 CTG TAC CGC GGG GAC GTG GTG CCC AAG GAC GTC AAC GCC GCC ATC GCC ACC ATC AAG ACC AAG C ca7 CTG TAC CGC GGG GAC GTG GTG CCC AAG GAC GTC AAC GCC GCC ATC GCC ACC ATC AAG ACC AAG C	GC AGC ATC CAG TTT GTG GAC TGG TGC CCA ACT GGT TTC AAG GTG GGC ATC AAC TAC
	380 390
Gin Pro Pro Thr Val Val Ala Giy Giy Asp Leu Ala Lys Val Gin Arg Ile Val Cys Met Leu S ca5 CAG CCC CCC ACC GTG GTT GCC GGG GGA GAC CTG GCC AAG GTG CAG CGA ATC GTC TGC ATG CTG A	er Asn Thr Thr Ala Ile Ala Glu Ala Trp Ala Arg Leu Asp His Lys Phe Asp Leu
CAS CAG CCG CCC ACG GTG GTG GCC GGG GGA GAC CTG GCC AAG GTG CAG CGA ATC GTC TGC ATG CTG A	GC AAC ACC ACG GCC ATC GCC GAG GCG TGG GCG CGC CTG GAC CAC AAG TTC GAC CTG
400 Net Tyr Ala Lys Arg Ala Phe Val His Trp Tyr Val Gly Glu Gly Met Glu Glu Gly Glu Phe S	420 er Glu Ala Arg Glu Asp Met Ala Ala Leu Glu Lys Asp Tyr Glu Glu Val Glv Leu
Ca5 ATG TAC GCC AAG CGC GCC TTC GTG CAC TGG TAC GTG GGC GAG GGC ATG GAG GAC GGG GAG TTC T Ca7 ATG TAC GCC AAG CGC GCC TTC GTG CAC TGG TAC GTG GGC GAG GGC ATT GGA GGA GGG GGA GTT C	CC GAG GCG CGG GAG GAC ATG GCC GCC CTG GAG AAG GAT TAC GAG GAG GTG GGC CTG
Ile Gly Gly Gly Gly Val L 440 449	eu Arg Gly Ala Gly Gly His Trp Pro Pro Trp Arg Arg Ile Thr Arg Arg Trp Ala
Asp Ser Tyr Glu Asp Glu Glu Glu Glu Glu Glu Glu . •	ATGTCACCTGCGGAAACCCTTTGC <u>AATAAA</u> CCGTTTGAGAACCTCCGTGTCTCCCAGTGACCCCCCCA
ca7 TGG ACT CCT ACG AGG ACG AAG AGG AGG GGG GGG AGG AGT AGAGCGGCCGCCAGCGCGGGGACCTCGCTCGCTTCCC Trp Thr Pro Thr Arg Thr Lys Arg Arg Ala Arg Ser ArgAlaAlaAlaGlnArgGlyThrSerLeuLeuSerH	ATGTCACCTGCGGAAACCCTTTGC <u>AATAAA</u> CCGTTTGAGAACCTCCGTGTCTCCCAGTGACCCCCCCA

Fig. 3. The DNA and inferred encoded polypeptide sequences for genes $c\alpha 5$ and $c\alpha 7$. The DNA sequences for $c\alpha 5$ and $c\alpha 7$ were determined on both strands according to the strategies diagrammed in Figure 1A. The amino acid sequence for $c\alpha 5$ is shown above the $c\alpha 5$ DNA sequence. Only those amino acid positions that differ from $c\alpha 5$ are shown for $c\alpha 7$. Asterisks (*) represent the translation termination codons. Putative consensus signals for polyadenylation are underlined.

represents a truncated pseudogene in which the 5' exons have been lost, those exon sequences lie >9 kb upstream, or those exons are very divergent and lie >2 kb upstream.

In any event, determination of the size of the $c\alpha3$ mRNA using RNA blotting with an oligonucleotide probe complementary to a portion of the putative 3' untranslated region (see Figure 5A and next section) revealed a doublet of RNAs whose 1.8-kb size is consistent both with the known features of the $c\alpha3$ gene and with it encoding an authentic α -tubulin polypeptide [e.g. 50-base 5' untranslated region, 1350-base coding region, 180-base 3' untranslated region, 100- to 200-base poly(A)tail].

We conclude that $c\alpha 3$ represents a functional, expressed α -tubulin gene and that the 5' exons must lie >9 kb 5' to exon 4 or are very divergent and lie >2 kb away.

Comparison of the inferred polypeptide sequence of $c\alpha 3$ with $c\alpha 1$ or with the highly conserved consensus sequence derived from six known mammalian α -tubulins (see Figure 6) revealed that 30 of the 325 residue positions determined (10%) are divergent. The highest concentration of these differences is within the carboxy-terminal region where, compared to $c\alpha 1$, substitutions or deletions have occurred in nine of the last 15 amino acid positions. Of further interest was

934

the absence of an encoded, terminal tyrosine which for many α -tubulin polypeptides undergoes a cyclic post-translational removal and readdition.

DNA sequence analysis of $c\alpha 4$: an α -tubulin gene with an additional intron interrupting codon 352

DNA sequence of the $c\alpha 4$ gene was obtained as shown in the strategy in Figure 1A. The actual sequence is shown along with $c\alpha 3$ in Figure 2. As with the $c\alpha 3$ gene, sequence of all segments with detectable homology to the $c\alpha 1$ coding sequence probe yielded an incomplete gene sequence that began at codon 126 and terminated at the carboxy terminus of α -tubulin. Sequence analysis extending 1 kb upstream did not reveal the presence of the remaining exons. Inspection of the sequence data did reveal, however, that within the sequenced domain an apparent, additional intron segment interrupted the coding sequences between codons 352 and 353. Following this 800-base non-coding sequence, a new fifth exon emerged (in the correct translation frame) from a normal splice junction and continued on to the translation termination codon at position 448. Unlike $c\alpha 3$, no consensus signal for polyadenylation lies within 245 bases 3' of the translation termination codon.

Like $c\alpha 3$, $c\alpha 4$ displays 10% sequence divergence with the $c\alpha 1$ and mammalian consensus α -tubulins (Figure 6), with the highest degree of divergence localized in the carboxy terminus. Unfortunately, all attempts to demonstrate expression of $c\alpha 4$ were unsuccessful. One such example using RNA blotting with an oligonucleotide complementary to the putative 3' untranslated region of $c\alpha 4$ failed to detect any RNA transcripts within the 12 tissues and cell lines examined (Figure 5A). Similarly, no $c\alpha 4$ RNAs could be detected in spleen, bursa, thymus or pancreas RNAs (not shown).

Thus, although in the domain sequenced $c\alpha 4$ has retained an open reading frame that could encode an α -tubulin, we are unsure if it is a functional, expressed gene. If it is expressed, either its level of expression is very low or it possesses a restricted range of expression in tissues we have not yet examined.

Analysis of $c\alpha 5$ and $c\alpha 6$: allelic genes encoding a conservative, expressed α -tubulin without an encoded carboxy-terminal tyrosine

The sequence of the $c\alpha 5$ gene (determined by the strategy shown in Figure 1A) is presented in Figure 3. From the 2500 nucleotides of sequence that were determined, we deduced the sequence for 447 encoded amino acids. Like the rat and human α -tubulin genes previously analyzed, $c\alpha 5$ is interrupted by three introns at positions between codons 1 and 2, within codon 76 and between codons 125 and 126. (Since exon 1 encodes only the initiator methionine, it is not possible to identify where in the 5' sequences this codon and the adjacent 5' untranslated region are located.) The relatively short introns 2 and 3 (442 and 211 bases respectively) are flanked by appropriate consensus splicing sequences. There are no lesions (translation termination codons etc.) that would suggest that $c\alpha 5$ is other than a functional gene.

Partial DNA sequence analysis of the $c\alpha 6$ gene was also performed according to the strategy shown in Figure 1A. However, comparison of these sequences with $c\alpha 5$ revealed almost complete identity, including those portions within introns (e.g. intron 2 diverged in only three positions). Although restriction mapping had initially revealed many differences between $c\alpha 5$ and $c\alpha 6$, some clear homologies in the maps could also be observed (see Figure 1A). These observations raised the possibility that $c\alpha 5$ and $c\alpha 6$ represented alleles of the same gene rather than independent genetic loci. To test this, for both genes we sequenced an SspI-BamHI fragment that lies 3' to the tubulin homology in both genes (see strategy for $c\alpha 6$ —Figure 1A). Both sequences were identical, unambiguously indicating that the flanking domain was highly homologous in the two clones. For this to be the case, either $c\alpha 5$ or $c\alpha 6$ arose by recent gene duplication of a large segment of DNA or $c\alpha 5$ and $c\alpha 6$ represent alleles of the same gene which display numerous restriction site polymorphisms. In either case, we cannot distinguish between the RNAs encoded by the two genes and will refer to the corresponding RNA/protein as $c\alpha 5$

Of all the genes analyzed, $c\alpha 5$ is most closely related to $c\alpha 1$, both at the nucleotide and protein levels. Although 4% of the amino acids are divergent, most of the divergence is concentrated in exons 2 and 3 and at the carboxy terminus. Just as was the case for $c\alpha 3$ and $c\alpha 4$, an encoded carboxy-terminal tyrosine residue is not found.

Analysis of $c\alpha$ 7: a chicken α -tubulin pseudogene

The $c\alpha$ 7 gene segment was isolated as a gene linked within 2 kb, 5' to the $c\alpha 5$ genomic fragment. Both genes are oriented in the same transcriptional direction. Nucleotide sequence of $c\alpha 7$ (obtained by the strategy in Figure 1A) is shown in Figure 3 along with $c\alpha 5$. $c\alpha 7$ shows a very strong nucleotide sequence homology to $c\alpha 5$ beginning 15 nucleotides 5' to the beginning of exon 4 and continuing into the 3' untranslated region. However, no homology is detected in sequences further 5' nor can sequences homologous to exons 2 and 3 be detected by hybridization within ~ 6 kb upstream. Thus, if the 5' exons exist they must be very divergent in sequence or lie >6 kb away. Further, at amino acid position 413 and again at 426 an extra thymidine residue has been inserted. In each instance these insertions change the translation frame. As a consequence, the predicted polypeptide terminates with 24 'extra' carboxyterminal amino acids encoded by sequences that lie within the 3' untranslated region of $c\alpha 5$. In view of the frameshift lesions and the possible loss of the 5' exons, we conclude that $c\alpha 7$ is a pseudogene, the apparent product of an unequal crossover event during recombination. [Note that if $c\alpha 7$ were actively transcribed, its RNA could not be distinguished from $c\alpha 5$ by 3' untranslated region probes. However, since $c\alpha 5$ contains no apparent lesions that would disrupt its expression, we believe that any mRNA detected (see below and Figure 5A) is encoded by $c\alpha 5$.]

Analysis of $c\alpha 8$: a functional gene with an additional intron between codons 352 and 353

 $c\alpha 8$ was subjected to DNA sequence analysis (Figure 1A) and the sequences determined are shown in Figure 4. Again, only sequences 3' to codon 126 were determined [the $c\alpha 8$ gene segment contains only 1 kb of sequences (presumably intron 3) lying 5' to codon 126]. Immediately apparent from inspection of the figure is that this gene segment also contains an apparent, additional 201-base intron between codons 352 and 353 (the same position as the extra intron in $c\alpha 4$).

Comparison of the predicted polypeptide sequence encoded by $c\alpha 8$ to that of the other genes (Figure 6) reveals that $c\alpha 8$ most closely resembles $c\alpha 2$ and $c\alpha 4$ in the type of amino acid substitutions made, but overall is still more closely related to $c\alpha 1$ (differing in only 4% of positions). Once again, a major cluster of divergence appears in the carboxyterminal region of the coding sequence, where six substitutions and/or deletions have been accepted in the last 15 amino acid positions. Alone among $c\alpha 3 - c\alpha 8$, $c\alpha 8$ contains the encoded terminal tyrosine found in most previously analyzed α -tubulins.

Determination of the pattern of expression of $c\alpha 3 - c\alpha 8$

To determine where $c\alpha 3 - c\alpha 8$ were expressed in chicken tissues, ³²P-labeled probes that would specifically identify the corresponding mRNAs were prepared from the 3' untranslated regions of $c\alpha 3$, $c\alpha 4$, $c\alpha 5/6$ and $c\alpha 8$. These were used to detect the corresponding RNA transcripts on parallel blots of RNAs isolated from a series of chicken tissues or cell lines. Autoradiograms of the resultant blots are shown in Figure 5A along with an additional blot probed with a coding sequence probe (from $c\alpha 1$) that we believe detects all α -tubulin mRNAs. As expected, α -tubulin mRNAs were detected in all cells and tissues using the coding sequence

126 130 Ala Asp Gin Cys Thr Giy Leu Gin Giy Phe Leu Ile AAGCTTGATCAGTTGGTATGTCAGTTGAGTGCTTACATGAGTCTGGAGTAGGCTTGTTTTGGGATTCTTGGTGTCTGTACCACTGCTGTAACACATGTATTTTCTCTTTCAGGTC GAC CAA TGC ACT GGC TTG CAG GGT TTC CTG ATT
140 140 140 170 160 170 170 170 170 170 170 170 17
180 Ser Thr Ala Val Glu Pro Tyr Asn Ser Ile Leu Thr Thr His Thr Thr Leu Glu Gln Ser Asp Cys Ala Phe Met Val Asp Asn Glu Ala Ile Tyr Asp Ile Cys Arg Arg Asn Leu TCA ACA OCT GTT GTT GAG CCA TAC AAC TCC ATC CTG ACC ACC CAC ACC CTG GAG CAG TCT GAC TGT GCC TTT ATG GTT GAC AAT GAG GCC ATC TAT GAC ATT TGC CGT AGG AAT CTG
220 250 Asp Ile Glu Arg Pro Thr Tyr Thr Asn Leu Asn Arg Leu Ile Gly Gln Ile Val Ser Ser Ile Thr Ala Ser Leu Arg Phe Asp Gly Ala Leu Asn Val Asp Leu Thr Glu Phe Gln Thr GAC ATT GAA CGC CCG ACA TAC ACT AAC CTC AAT CGC CTC ATT GGT CAG ATT GTG TCT TCC ATT ACT GCT TCC CTC AGA TTC GAT GGT GCC TTA AAT GTG GAT CTC ACT GAA TTC CAA ACT
260 290 Asn Leu Val Pro Tyr Pro Arg Ile His Phe Pro Leu Val Thr Tyr Ser Pro Ile Ile Ser Ala Glu Lys Ala Tyr His Glu Gln Leu Ser Val Ser Glu Ile Thr Asn Ala Cys Phe Glu AAC TTG GTG CCT TAC CCT CGT ATC CAC TTC CCA CTG GTA ACG TAT TCC CCA ATT ATT TCA GCA GAG AAA GCC TAT CAT GAG CAG CTC TCT GTG TCT GAG ATA ACC AAT GCA TGC TTT GAG
300 Pro Ser Asn Gln Met Val Lys Cys Asp Pro Arg His Gly Lys Tyr Met Ala Cys Cys Met Leu Tyr Arg Gly Asp Val Val Pro Lys Asp Val Asn Ala Ala Ile Ala Ala Ile Lys Thr CCT TCC AAT CAG ATG GTG AAG TGT GAC CCT CGC CAT GGC AAG TAC ATG GCC TGC TGC TGC GT GT GTG CGG GAT GTT GTC CCC AAG GAT GTC AAC GCT GCT ATT GCT GCT ATC AAA ACC
340 350 352 Lys Arg Thr Ile Gln Phe Val Asp Trp Cys Pro Thr Gly Phe Lys AAG CGC ACG ACG ACG CAG TTT GTG GAC TGG TGC CCT ACT GGT TTT AAG GCAAGTCTCAGTTATATCATGAGTCTTTTTTTTTT
353 Val Gly Ile Asn AATCTGTAGGCAGGTACCTGATTTGTGTTGAGAAATGAGACCCTTAATGCTTTTGTGCAGAGATGCTTTCTCTGGCTTCTTTGACTATGGTCACTGTGGGAAACTCATGGTCTGTGTGCTTCTCCTGTAG GTT GGC ATC AAT
360 390 Tyr Gln Pro Pro Thr Val Val Pro Gly Gly Asp Leu Ala Lys Val Gln Arg Ala Val Cys Met Leu Ser Asn Thr Thr Ala Ile Ala Glu Ala Trp Ala Arg Leu Asp His Lys Phe Asp TAC CAG CCA CCA ACA GTG GTT CCT GGT GGT GGC CTG GCC AAA GTG CAG GGG GCA GTG TGC ATG GTG AGC AAC ACT ACA GCC ATT GCT GAA GCA TGG GCT CGT CTT GAC CAC AAG TTT GAT
400 430 Leu Met Tyr Ala Lys Arg Ala Phe Val His Trp Tyr Val Gly Glu Gly Met Glu Glu Glu Glu Glu Phe Ser Glu Ala Arg Glu Asp Leu Ala Ala Leu Glu Lys Asp Tyr Glu Glu Val Gly CTG ATG TAT GCC AAA CGT GCC TTT GTT CAC TGG TAT GTT GGA GAA GGG ATG GAG GAG GGG GAA TTC TCA GAG GCT CGG GAA GAT TTG GCT GCA CAT AAA GAT TAT GAA GAA GTG GGC
440 Thr Asp Ser Met Asp Gly Glu Asp Glu Glu Glu Glu Clu Tyr ● ACA GAC TCG ATG GAT GAA GAT GAA GAT GAA GAG TAC TGA ATGCAGCCAAAGTGAATTGCAGGTCTTCACCAGTGTACTGCCTTCTTATTAGCAGTGTAAATCATGTCTAATGATTGG <u>AATAAA</u> CTTCTTTAAGACATCCAGT
AGT OT CC AAACACTAT GT ACCTT CAT CA OCT GT AT A CAAACAGT CC GT AGTCT AAGAT AC CAT TT CT TT CT GAGAAAGGT AGTC GAGAAAGGT AGTC AT GCT GCG G

Fig. 4. The DNA and inferred encoded polypeptide sequences for gene $c\alpha 8$. The DNA sequence of $c\alpha 8$ beginning in intron 3 and continuing through the putative site for polyadenylation was determined by the strategy diagrammed in Figure 1A. The corresponding amino acid sequence is shown above the DNA sequence. The asterisk (*) represents the translation termination codon. The putative consensus sequence for polyadenylation is underlined.

probe. A distinctly different pattern was seen for $c\alpha 3$. For this gene, a ³²P-labeled oligonucleotide complementary to a portion of the putative 3' untranslated region identified a doublet of RNAs of ~ 1.8 kb in length in RNAs derived from several tissues, most prominently intestine and liver. (We cannot distinguish among the possibilities that could yield RNAs of slightly different size.) On the other hand, no RNAs could be detected with an oligonucleotide probe complementary to the $c\alpha 4$ 3' untranslated region, either in the 12 samples analyzed in Figure 5A or in spleen, bursa, thymus or pancreas (not shown). Thus, despite the fact that the $c\alpha 4$ gene contains no lesions that would suggest it to be other than a functional gene, we could not document that it is actually expressed. For $c\alpha 5$, a relatively high level of expression was found in two of three lymphocyte lines and in thymus. Low levels of expression were found in brain and testis, the only other cells or tissues in which $c\alpha 5$ transcripts were detected. Lastly, $c\alpha 8$ RNAs were detected ubiquitously at low levels in all 14 tissues surveyed, and at substantially higher levels in brain and in two of three lymphocyte lines. A summary of these hybridization results is presented in Figure 5B.

Discussion

In conjunction with two previously analyzed chicken α tubulin cDNAs, the characterization of the $\alpha 3 - \alpha 8$ genes has provided a survey of the α -tubulin gene family in one vertebrate genome. $\alpha 1 - \alpha 8$ appear to account for $\sim 75\%$ of the genomic α -tubulin segments. Within this family of sequences, five genes appear encode functional tubulins ($\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 5$ and $\alpha 8$), whereas one is almost cer-

936

tainly a pseudogene ($c\alpha$ 7) and one ($c\alpha$ 4) may yet be a functional gene (although we were unable to document its expression). Potentially, several more functional α -tubulins may be encoded by genes not yet isolated.

Detailed comparison of the polypeptides encoded by the genes now identified to be functional is presented in Figure 6. (Also included is $c\alpha 4$ since it may be functional.) Only one $(c\alpha 1)$ is highly homologous in amino acid sequence to the consensus found in most previously sequenced vertebrate α -tubulins (Figure 6). For the portions of the sequences that have been determined, the remaining four ($c\alpha 2$, $c\alpha 3$, $c\alpha 5$ and $c\alpha 8$) diverge from the consensus sequence in 17, 10, 4 and 4% of residue positions. Most striking is that all but $c\alpha 1$ show substantial sequence heterogeneity in the carboxyterminal positions beyond residue 435, where substitutions and deletions have been accepted in 100% of the positions (these data are summarized in Figure 7). This is analogous to the situation for β -tubulin where the major differences among β -tubulin isotypes are within the carboxy-terminal 15 residues (e.g. Sullivan and Cleveland, 1986).

Analysis of the chicken α -tubulins represents the second detailed look at a vertebrate α -tubulin family. Previously, Cowan and collaborators (Villasante *et al.*, 1986) identified six different mouse α -tubulin cDNAs and showed them to encode five different polypeptides. These included a polypeptide ubiquitously expressed, a sister gene whose expression is a subset of the first, two genes expressed only in testis, and one gene expressed at low levels in many tissues and which yields only an aberrant transcript in brain and testis. Although there are some similarities, the patterns of expression of the five functional chicken α -tubulins (summarized in Figure 5B) do not obviously parallel those in mouse. Fur-

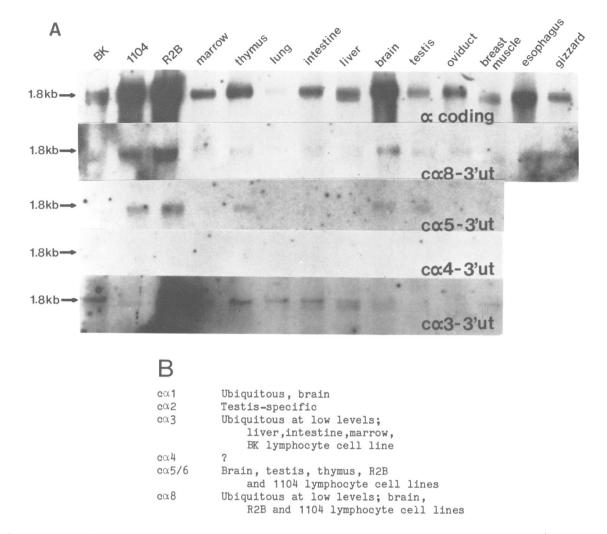


Fig. 5. RNA blot analysis of the expression of genes $\alpha 1 - \alpha 8$ in chicken tissues and cell lines. (A) Equal amounts of $poly(A)^+$ RNA from various chicken tissues and cell lines were electrophoresed under denaturing conditions on five parallel gels and blotted to nitrocellulose. Filters were hybridized with a coding sequence probe from $\alpha 1$ that is believed to hybridize efficiently to all α -tubulin mRNAs (top panel) or with probes (see Materials and methods) that specifically recognize $\alpha 8$, $\alpha 5$, $\alpha 4$ or $\alpha 3$ RNAs respectively. BK, 1104 and R2B are chicken lymphocyte cell lines. (B) Summary of the patterns of expression of $\alpha 1 - \alpha 8$ as deduced from RNA blots. Data for expression of $\alpha 1$ and $\alpha 2$ are from Pratt (1987) and Pratt *et al.* (1987) respectively.

ther, the divergence in sequence among chicken α -tubulins is substantially greater than the 4% maximum divergence found thus far in mouse.

Consideration of the heterogeneity within at least the chicken α -tubulin carboxy-terminal sequences suggests the possibility that the polypeptides may be functionally distinguishable. Only $\alpha 1$ and $\alpha 8$ contain encoded terminal tyrosine residues, whereas all but one α -tubulin previously analyzed in vertebrates (M α 4—Villasante *et al.*, 1986) has contained an encoded carboxy-terminal tyrosine. In many tissues α -tubulins are known to undergo a cyclic process of detyrosination (through the action of a carboxypeptidase) followed by re-tyrosination (by a specific α -tubulin—tyrosine ligase). Further, a subset of microtubules in cultured cells that is enriched in detyrosinated tubulin are also more stable (Kreis, 1987; Wehland and Weber, 1987), although detyrosination is probably not the cause of the increased stability (Khawaja *et al.*, 1988).

Despite this, the determinants that specify α -tubulin as the unique substrate for the tyrosine ligase have not been identified. However, the divergence of the carboxy-terminal

residues among the chicken polypeptides clearly suggests the possibility that one or more of these α -tubulins does not participate in this cycle. Since it has been demonstrated (Gundersen *et al.*, 1987; Wehland and Weber, 1987) that (at least for the consensus α -tubulins such as $c\alpha 1$) it is the tyrosinated subunit that is the major subunit used for assembly and that detyrosination occurs post-polymerization, carboxy-terminal sequence heterogeneity could easily produce functional distinctions among α -tubulin isotypes.

So what can we conclude concerning the significance of the α -tubulin multigene family? Unlike the case for β -tubulin where evolutionarily conserved polypeptide isotypes have been identified both by their unique carboxy-terminal sequences and programmes of expression, comparison of the carboxy-terminal sequences from all available vertebrate α tubulins (compared in Figure 7) provides little support for the existence of conserved isotypic classes of α -tubulin polypeptides. (In fact, except for $c\alpha 1$, no other chicken isotype corresponds in sequence to any of the remaining four mouse α -tubulins.) Nor does consideration of the pattern of expression offer firm support for such a hypothesis. The

MAMMALIAN Composite ca1		IO VGQAGVQIO	20 GNACWELY	30 Clehgiqp	DGQMP	40 SDKTIG	50 GGDDSFNT	FFSE	60 TGAGKHV	70 PRAVFVDLE	80 EPT VIDE VRTGT	90 Y RQLFHPEQL	ITG	100 (EDAANNYA)	110 RGHYTIGKEII	120 DLVLD
ca5 ca3 ca8 ca4		м	т			L	Т	C		I	A I			G		Q
ca2	L V	I	F	S	TF-	PP-S	S- A	R	SMS Y	IM	v	H			V DKV	MS
MAMMALIAN Composite cai				150 Igsgftsl		160 SVDYGKI	170 KSKLEFSI	YPAP	180 QVSTAVV	190 EPYNSILTT	200 THTTLEHSDCAF	210 MVDNEAIYDI	CRRI	220 NLDIERPTY	230 S TNLNRLIGQIN G	240 Issita
cα5		L.						R							Ğ	
ca3	1	ŝ				ER	т V	DR			S		N		G	V
ca8	i	LI	-					2.			-				G	
ca4	i .						LA						С		S	
ca2	, n S					Е	A		A	v	v		Ĥ		s	
CUL	5	5 51	•			2	А		n	•	•				-	
MAMMALIAN Composite ca1	25 SLRFDGALN		260 NLVPYPRI	270 HFPLATY		280 AEKAYHI	290 Eqlsvaei	TNAC	300 Fepanqm	310 WKCDPRHGK	320 CYMACCLLYRGD	330 VVPKDVNAAI	ATI	340 KTKRTIQFV	350 DWCPTGFKVG	360 INYQPP
cas								s						S		
cα3	N	I		T	RI		P		FS	R			A	RS		
ca8		-		v	SI		s		s		м		Ā			
ca4				• v		SR	5	SS	N		M	v	Å	N		
			F	v		SDR		SS	N	00	M	v	~	N SL		
cα2			r	v	1	SDR		22	N	44	м	v	A	N BL		
MAMMALIAN	31		380	390		400	410		420	430	440	450				
COMPOSITE	TVVPGG9L	IK VQRA VCM	ILSNITALA	AEAWARLD	HKFDL	MI AK KA	FVHWIVGE	GMEE	GEFSEAN	EDMAALEKL	DYEEVGVDSVEG	EGEEEGEEI*				
cα1		_														
ca5	A	I								-	LYD)				
ca3										L	RN AD	- F •				
ca8										L	T MD	D	r			
ca4		Q					S		A	L	TFD) ND *				
ca2	IPT	Q					S		A	L	DATLFD) N AGS#				

Fig. 6. Comparison of the polypeptide sequences of the known chicken α -tubulins. The one-letter amino acid code is utilized to display the sequences of $c\alpha_1 - c\alpha_8$. A composite mammalian sequence [represented by human genes $b\alpha_1$ and $k\alpha_1$ (Cowan *et al.*, 1983), a porcine sequence (Ponstingl *et al.*, 1981), a rat α -tubulin (Lemischka *et al.*, 1981) and two Chinese hamster α -tubulins (Elliott *et al.*, 1986)] is displayed on the top line. The $c\alpha_1 - c\alpha_8$ sequences are shown beneath the composite, explicitly showing only the residue positions that differ with the composite sequence. Dashes (-) represent deletions introduced to maintain homology. Translation termination codons are denoted by (*). Portions of sequence that have not yet been determined (the amino-terminal 125 amino acids of $c\alpha_3$, $c\alpha_4$ and $c\alpha_8$ and the amino-terminal 39 residues of $c\alpha_1$) are denoted on the figure by the areas contained within brackets.

absence of such evidence may simply be the result of failure to isolate the homologous sequences from multiple species. Alternatively, it may indicate that the carboxy-terminal sequences of α -tubulin can tolerate substantial sequence divergence without disruption of function as has been demonstrated for an amino-terminal variable domain between residues 30 and 50 (Schatz *et al.*, 1987). Although the evolutionarily conserved cyclic process of carboxy-terminal tyrosination/detyrosination argues against this latter possibility, additional analyses will be necessary to settle this point. What is clear is that significantly more sequence heterogeneity is present in vertebrate α -tubulins than had previously been recognized.

Materials and methods

Genomic library screening

A genomic library of embryonic chicken DNA was obtained from Dr Bjorn Vennstrom (EMBL, Heidelberg–Vennstrom and Bishop, 1982). The library was constructed by partial digestion of chicken DNA with *Mbo*I and selection of DNA fragments of ~15–20 kb in length. These were then ligated into Charon 30 which had been digested with *Bam*HI. Approximately 1.6 × 10⁶ phage (four times the complexity of the library) were screened for α -tubulin sequences by the protocol of Maniatis *et al.* (1978). A ³²P-labeled probe was prepared from an embryonic chick brain cDNA clone originally called pT1, renamed c α 1 (Pratt *et al.*, 1987), that contains the coding sequence for α -tubulin beginning at amino acid 42 and continuing through the 3' complete untranslated region (Valenzuela *et al.*, 1981). The probe was prepared by the random priming method of Shank *et al.* (1978). Conditions for hybridization and washing were as given below for DNA blots.

Restriction maps of the cloned genomic DNA were prepared using restriction endonucleases according to suppliers' recommendations.

Gel electrophoresis and blotting

Gel electrophoresis of DNA was performed on vertical slab gels of 0.8% agarose containing 40 mM Tris-HCl, pH 8.1, 20 mM sodium acetate,

cal Li	EKDYEE	VGVI	OSVE	GEGEEEGE	EY #
ca5		L	Y	D	×
ca3		RI	N AD	- F	#
ca8		Т	MD	D	*
ca4		Т	F	D ND	#
ca 2	D	AT	LF	D N AGS	ł
ra1, raT26, h-ka1, h-ba	1				
Ma1/2,CHOal,II,pa1					¥
Μα3/7				A A-	¥
Ma4,ha44		I	Y	D	¥
Mad,CHOaIII		A	A	DD	Ħ

Fig. 7. Summary of carboxy-terminal sequences of chicken $c\alpha 1 - c\alpha 8$ tubulins. The carboxy-terminal sequences for all known vertebrate α -tubulins are aligned. Asterisks (*) represent the translation termination codons. Dashes (—) represent deletions introduced so as to maintain homology. $r\alpha 1$ and $r\alpha T26$ are taken from Lemischka *et al.* (1981) and Ginzburg *et al.* (1981) respectively; h-k $\alpha 1$ and h-b $\alpha 1$ are from Cowan *et al.* (1983); CHO I, II and III are from Elliott *et al.* (1986); $p\alpha 1$ is from Ponstingl *et al.* (1981); $M\alpha 1 - M\alpha 7$ are from Villasante *et al.* (1986); h $\alpha 44$ is from Dobner *et al.* (1987).

and 2 mM EDTA. Gels were stained with 1 μ g/ml ethidium bromide to visualize the DNA bands. DNA was transferred to nitrocellulose filters according to the method of Southern (1975).

Gel electrophoresis of RNA was carried out on vertical slab gels cast from 1.0% agarose containing 2.2 M formaldehyde (Boedtker, 1971). RNA was transferred to nitrocellulose by the method of Thomas (1980).

Hybridization conditions for DNA and RNA blots

DNA and RNA blots were prehybridized for 2-16 h at 41° C in 50% formamide, $5 \times SSC$ ($1 \times SSC = 150$ mM NaCl, 15 mM sodium citrate), 20 mM Hepes (pH 7.4), 0.4 mg/ml yeast tRNA, 0.06 mg/ml sonicated, denatured herring sperm DNA, 0.06% bovine serum albumin, 0.06% Ficoll, 0.06% polyvinylpyrrolidone. The filters were then hybridized in the same solution containing 10^{6} d.p.m./ml of a 32 P-labeled probe that had been denatured by boiling for 3 min. These filters were washed in three changes of 0.1 × SSC, 0.1% SDS at 51°C for a total wash time of 1 h.

Hybridization conditions for DNA and RNA blots probed with oligonucleotides

Filters were prehybridized for 24 h in 20% formamide, 5 × SSC, 20 mM phosphate buffer (pH 6.5) and 0.06% albumin, 0.06% Ficoll and 0.06% polyvinylpyrrolidone. The blots were then hybridized in the same solution containing 10⁶ d.p.m./ml of the appropriate ³²P-labeled probe (labeled with polynucleotide kinase and [γ -³²P]ATP) for a minimum of 24 h. The blots were washed to a final stringency of 2 × SSC at 60°C according to Villasante *et al.* (1986).

Subcloning isolated chicken α -tubulin genes

Regions of phage DNA containing each individual α -tubulin hybridizing region were subcloned in pBR322 as follows.

pc α 3: a 6.0 kb genomic *Hind*III fragment was subcloned into the *Hin*dIII site of pBR322, such that the direction of transcription of the α -tubulin segment would be in the clockwise direction in the normal presentation of pBR322.

 $pc\alpha 4$: a 5.0-kb *Bam*HI fragment was subcloned into the *Bam*HI site of pBR322 in the counterclockwise orientation to the pBR322 vector.

 $pc\alpha 5$: a 9.4-kb *Eco*RI fragment was subcloned into the *Eco*RI site of pBR322 in the counterclockwise orientation to the pBR322 vector.

 $pc\alpha$?: a 5.0-kb *Eco*RI fragment was subcloned into the *Eco*RI site of pBR322 in the counterclockwise orientation to the pBR322 vector.

 $pc\alpha 6$: a 12-kb fragment bounded by *Eco*RI and *Hind*III was subcloned into double-cut pBR322.

 $pc\alpha 8$: a 6-kb HindIII fragment was subcloned into the HindIII site of pBR322 in the clockwise orientation.

Construction of DNA probes that hybridized with specific α -tubulin genes

c α 3: an antisense oligonucleotide probe of 29 bases complementary to the 3' non-coding region directly proximal to the translation stop codon was used as probe. The probe was 5' end-labeled with [γ -³²P]dATP and polynucleotide kinase.

 $\alpha 4$: an antisense oligonucleotide probe of 29 bases complementary to nucleotides 20-49 3' to the translation stop codon (with 1 taken as the first base 3' to the stop codon) was synthesized for use as a probe and end-labeled as for $\alpha 3$.

 $c\alpha 5$: a region of DNA from an AvaII site 18 nucleotides downstream of the stop codon to an AvaII site ~ 200 nucleotides more 3' to the gene was used as a probe.

 $c\alpha 8$: a region of DNA from a *TaqI* site at the 3' end of the coding region (30 bases upstream from the translation termination codon) extending to the distal *PstI* site 209 nucleotides downstream was used as a probe.

We confirmed that the probes were specific to the corresponding gene/RNA by hybridization of the probes to (i) a panel of blots containing all of the cloned sequences and (ii) to genomic blots of chicken DNA (not shown).

DNA sequence analysis

Three methods of obtaining M13 clones carrying appropriate α -tubulin gene segments were utilized for DNA sequence analysis. First, restriction fragments were directly subcloned into the replicative form of M13 mp18 or mp19. Second, progressive deletion series were obtained using Bal31 exonuclease as described in detail by Lewis et al. (1985a). These deletions were ligated into appropriately digested M13 with T4 DNA ligase under conditions recommended by the supplier. Third, the kilo-sequencing method of Barnes et al. (1983) was used following initial subcloning of specific gene fragments into M13. Briefly, the replicative form of the phage was nicked in a random position using DNase I cleavage in the presence of ethidium bromide. The nick was extended into a gap with exonuclease III and Bal31 nuclease was used to make a double-stranded break at the site. A restriction site proximal to the M13 sequencing priming site was then digested to release the region between the two sites and the remaining portion was religated together under conditions encouraging self-religation. Phage-carrying deletions were then sized on agarose slab gels to select an appropriate series of deletion clones.

DNA sequence was obtained from single-stranded M13 templates using the dideoxy method of Sanger *et al.* (1977). Sequencing reactions were loaded onto 9 M urea -6% acrylamide sequencing gels and run for 2 and/or 5 h. The sequence was visualized by autoradiography.

Preparation of RNA

Total RNA was prepared from tissues and cells by the method of Chirgwin *et al.* (1979). Briefly, cells or frozen tissues were disrupted in 6 M guanidine thiocyanate. Debris was pelleted by low-speed centrifugation and the supernatant layered on a cushion of cesium chloride. After centrifugation, the

RNA pellet was resuspended in diethyl pyrocarbonate (DEPC)-treated water and stored at -80° C. Poly(A)⁺ RNA was isolated from total RNA by passage over a column of oligo(dT) cellulose (Type III, Collaborative Research). Poly(A)⁺ RNA was precipitated with ethanol, resuspended in distilled water and stored at -80° C.

Acknowledgements

We thank Dr Bjorn Vennstrom for providing the chicken genomic library. This work has been supported by a grant from the NIH to D.W.C., who is also the recipient of an NIH Research Career Development Award.

References

Alsip, G.R. and Konkel, D.A. (1986) *Nucleic Acids Res.*, **14**, 2123-2138. Barnes, W.M., Bevan, M. and Son, P.H. (1983) *Methods Enzymol.*, **101**, 98-122

- Boedtker, H. (1971) Biochem. Biophys. Acta, 240, 448-453.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J.. and Rutter, W.J. (1979) Biochemistry, 18, 5294-5299.
- Cleveland, D.W. (1987) J. Cell Biol., 103, 381-383.
- Cleveland, D.W., Lopata, M.A., Cowan, N.J., MacDonald, R.J., Rutter, W.J. and Kirschner, M.W. (1980) *Cell*, **20**, 95-105.

Cleveland, D.W., Hughes, S.H., Stubblefield, E., Kirschner, M.W. and Varmus, H.E. (1981) J. Biol. Chem., 256, 3130-3134; Cell, 25, 537-546.

Cleveland, D.W. and Sullivan, K.F. (1985) Annu. Rev. Biochem., 54, 331-365.

- Cowan, N.J., Dobner, P.R., Fuchs, E.V. and Cleveland, D.W. (1983) *Mol. Cell. Biol.*, **3**, 1738-1745.
- Dobner, P.R., Kislauskis, E., Wentworth, B.M. and Villa-Komaroff, L. (1987) Nucleic Acids Res., 15, 199-218.
- Elliott, E.M., Henderson, G., Sarangi, F. and Ling, V. (1986) *Mol. Cell. Biol.*, 6, 906-913.
- Gibbons, I.R. (1975) In Inove, S. and Stephens, R.E. (eds), *Molecules and Cell Movement*. Raven Press, New York, pp. 207-232.
- Ginzburg, I., Behar, L., Givol, D. and Littauer, U.Z. (1981) Nucleic Acids Res., 9, 2691-2697.
- Gundersen, G.G., Khawaja, S. and Bulinski, J.C. (1987) J. Cell Biol., 105, 251-264.
- Hall, J.L. and Cowan, N.J. (1985) Nucleic Acids Res., 13, 207-222.
- Hall, J.L., Dudley, L., Dobner, P.R., Lewis, S.A. and Cowan, N.J. (1983) Mol. Cell. Biol., 3, 854-862.
- Hayden, J.H. and Allen, R.D. (1984) J. Cell Biol., 99, 1785-1793.
- Heuser, J. and Kirschner, M.W. (1980) J. Cell Biol., 86, 212-234.
- Khawaja,S., Gundersen,G.G. and Bulinski,J.C. (1988) J. Cell Biol., in press. Kreis, T.E. (1987) EMBO J., 6, 2597-2606.
- Lee, M.G.-S., Lewis, S.A., Wilde, C.D. and Cowan, N.J. (1983) Cell, 33, 477-487.
- Lemischka, I.R., Farmer, S.R., Racaniello, V.R. and Sharp, P.A. (1981) J. Mol. Biol., 150, 101-120.
- Lemischka, I.R. and Sharp, P.A. (1982) Nature, 300, 330-335.
- Lewis, S.A., Lee, M.G.-S. and Cowan, N.J. (1985a) J. Cell Biol., 101, 852-861.
- Lewis, S.A., Gilmartin, M.E., Hall, J.L. and Cowan, N.J. (1985b) J. Mol. Biol., 181, 11-20.
- Lopata, M.A. and Cleveland, D.W. (1987) J. Cell Biol., 105, 1707-1720.
- Maniatis, T., Hardison, R.C., Lacey, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K. and Efstratiadis, A. (1978) Cell, 15, 687-701.
- Montiero, M.J. and Cleveland, D.W. (1988) J. Mol. Biol., 199, 439-446.
- Murphy, D.B., Wallis, K.T., Machlin, P.S., Ratrie, H. and Cleveland, D.W. (1987) J. Biol. Chem., 262, 14305-14312.
- Ponstingl, H., Krauhs, E., Little, M. and Kempf, T. (1981) Proc. Natl. Acad. Sci. USA, 78, 2757-2761.
- Pratt, L. (1987) Ph.D. thesis, The Johns Hopkins University.
- Pratt,L. Okamura,S. and Cleveland,D.W. (1987) Mol. Cell. Biol., 7, 552-555.
- Robins, A.J., Wang, S.W., Smith, T.F. and Wells, J.R. (1986) J. Biol. Chem., 261, 18-20.
- Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) Proc. Natl. Acad. Sci. USA, 74, 5463-5467.
- Schatz, P., Georges, G., Solomon, F. and Botstein, D. (1987) Mol. Cell. Biol., 7, 3799-3805.
- Shank, P.R., Hughes, S.H., Kung, H.-J., Majors, J.E., Quintrell, N., Guntaka, R.V., Bishop, J.M. and Varmus, H.E. (1978) Cell, 15, 1383-1395.
- Southern, E.M. (1975) J. Mol. Biol., 98, 503-517.

- Sullivan, K.F. and Cleveland, D.W. (1984) *J. Cell Biol.*, **99**, 1754–1760. Sullivan, K.F. and Cleveland, D.W. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4327–4331.
- Sullivan, K.F., Lau, J.T.Y. and Cleveland, D.W. (1985) Mol. Cell. Biol., 5, 2454-2465.
- Sullivan, K.F., Havercroft, J.C., Machlin, P.S. and Cleveland, D.W. (1986a) *Mol. Cell. Biol.*, **6**, 4409-4418.
- Sullivan, K.F., Machlin, P.S., Ratrie, H., III and Cleveland, D.W. (1986b) J. Biol. Chem., 261, 13317-13322.
- Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA, 77, 5201-5205.
- Vale, R.D., Schnapp, G.J., Reese, T.S. and Sheetz, M.P. (1985) Cell, 40, 449-454.
- Valenzuela, P., Quiroga, M., Zaldivar, J., Rutter, W.J., Kirschner, M.W. and Cleveland, D.W. (1981) *Nature*, **289**, 650-655.
- Vennstrom, B. and Bishop, J.M. (1982) Cell, 28, 135-143.
- Villasante, A., Wang, D., Dobner, P.R., Dolph, P., Lewis, S.A. and Cowan, N.J. (1986) Mol. Cell. Biol., 6, 2409–2419.
- Wang, D., Villasante, A., Lewis, S.A. and Cowan, N.J. (1986) J. Cell Biol., 103, 1903-1910.
- Wehland, J. and Weber, K. (1987) J. Cell Sci., 88, 185-203.
- Wilde, C.D. Crowther, C.E., Cripe, T.D., Lee, M.G. and Cowan, N.J. (1982) Nature, 297, 83-84.

Received on November 3, 1987; revised on January 22, 1988