The murine immunoglobulin α gene expresses multiple transcripts from a unique membrane exon

Charlotte J. Word*, J. Frederic Mushinski¹ and Philip W. Tucker

Southwestern Medical School, Department of Microbiology, Dallas, TX 75235, and ¹Laboratory of Genetics, National Cancer Institute, National Institutes of Health, Building 8, Room 120, Bethesda, MD 20205, USA

Communicated by B. Pernis

Received on 3 March 1983; revised on 12 April 1983

Immunoglobulin heavy chains of each class exist in two forms, membrane (m) and secreted (s), which have different carboxy-terminal sequences encoded by different gene segments. We have sequenced the DNA spanning the coding region for the carboxy-terminal end of the α_m chain in a BALB/c genomic clone and have examined α mRNA expression in transformed and normal B cells. Unlike the other murine heavy chain genes, which have two membrane exons (M1 and M2), the α gene has only one membrane exon (α M1). This unusual exon encodes a highly conserved transmembrane peptide plus flanking amino acids singular to IgA. The complexity of the α gene transcription is also unusual. Downstream of α M1, three AATAAA hexanucleotides are used to terminate as many as six distinct α membrane mRNAs. These appear to differ only in length and splicing pattern of their 3'-untranslated regions, and thus, encode the same α_m protein. The α_s protein is ~4000 daltons smaller than α_m as judged by *in vitro* translation and is coded by a single 2.0-kb α_s mRNA. We propose that the expression of α_s and the multiple α_m mRNAs is differentially controlled at the level of transcriptional termination.

Key words: α heavy chain/membrane exon/mouse immunoglobulin/multiple mRNAs/transcriptional regulation

Introduction

Transcriptional regulation of eukaryotic genes occurs at several levels: initiation, elongation and termination, and internal splicing of the primary RNA transcript (reviewed in Darnell, 1982). Much attention has been given to the regulation of transcription initiation, although it has become increasingly clear that termination of transcription may be equally important in determining the protein product. Transcription continues perhaps 1000 or more bases beyond the 3' end of the mature mRNA in SV40 transcripts (Ford and Hsu, 1978; Lai *et al.*, 1978), in adenovirus early and late RNA transcripts (Nevins and Darnell, 1978; Nevins *et al.*, 1980) and in the mouse β -major globin RNA transcripts (Hofer and Darnell, 1981; Hofer *et al.*, 1982). It has been proposed that endonucleolytic cleavage of these precursors is followed by exonuclease trimming to within 15–20 bases of a consensus sequence AAUAAA (generally referred to as a poly(A) recognition sequence) (Proudfoot and Brownlee, 1976) found in most viral and eukaryotic mRNAs, with the exception of some yeast and histone mRNAs (Proudfoot, 1982). Subsequent polyadenylation and splicing of the primary RNA transcripts result in mature mRNAs. However, not every

AAUAAA sequence is recognized efficiently as an RNA termination/polyadenylation site as this sequence is present in the coding region and intervening sequences of several genes (McReynolds *et al.*, 1978; Perricaudet *et al.*, 1980; Tucker *et al.*, 1980). Thus, we will refer to the AAUAAA sequences that actually terminate mRNAs as 'end sites'.

mRNAs from a particular gene may stop at any of several different end sites. In these cases, the choice of end site affects the length of the 3'-untranslated region of the mRNA but does not alter the translated protein product. Examples include the mouse immunoglobulin membrane δ chain (Cheng, *et al.*, 1982) and other eukaryotic and viral genes (Setzer *et al.*, 1982; Tosi *et al.*, 1981; Unterman *et al.*, 1981; Parnes and Seidman, 1982; Hellig *et al.*, 1980; Fraser *et al.*, 1982; Benetzen and Hall, 1982).

However, some genes contain internal end sites. In these cases, both elongation and termination of the primary transcript, and recognition of a particular end site by the RNA endonuclease are important determinants in the expression of a particular mRNA species, and ultimately of a particular protein product. Such a system that has been partially characterized is the shift between membrane (m) and secreted (s) expression in murine heavy (H) chain constant (C) region immunoglobulin (Ig) genes. During the course of B cell differentiation, a B cell first expresses membrane IgM on the cell surface which acts as a receptor for antigen (Vitetta and Uhr, 1975). Later in differentiation, the B cell secretes IgM into the serum. Although both types of μ chains have an identical rearranged VDJ (variable region) segment, the carboxy-terminal amino acids of the membrane μ (μ_m) and secreted μ (μ_s) chains differ (Bergman and Haimovich, 1978; Vassalli *et al.*, 1979; Yuan *et al.*, 1979; Kehry *et al.*, 1980). The two μ chains are encoded by two different mRNA species derived from a single μ gene (Alt *et al.*, 1980; Rogers *et al.*, 1980; Early *et al.*, 1980; Singer *et al.*, 1980), and it was proposed that differential termination and polyadenylation of the primary RNA transcript determines which of the two μ mRNAs is expressed. Recently, similar models have been described for the expression of membrane and secreted δ (Cheng *et al.*, 1982), ϵ (Ishida *et al.*, 1982), γ_1 , γ_2b and γ_2a Ig H chains (Rogers *et al.*, 1981; Yamawaki-Kataoka *et al.*, 1982; Tyler *et al.*, 1982). Each Ig H chain gene (see Figure 9) contains at least two end sites. The promoter-proximal end site is located immediately 3' to a short gene segment (contiguous with the terminal C_H domain in all cases except δ) that encodes the hydrophilic carboxy terminus of the secreted H chain. Termination and polyadenylation of the primary transcript at this end site results in an RNA encoding a secreted H chain. If, however, termination and polyadenylation occur at an end site located several kilobases to the 3' side of the first end site, an mRNA is produced that contains sequences from two additional exons (M1 and M2) which encode the hydrophobic carboxy-terminal portion (see bottom of Figure 9) of the membrane H chain. RNA processing events remove the first end site from the membrane mRNA by splicing the C_H3 or C_H4 exon to the M1 exon at a conserved donor (5') site within the terminal C_H domain. Thus, it appears that in the case of Ig H chain

*To whom reprint requests should be sent.

gene expression, termination at alternative end sites determines the protein products expressed.

IgA is primarily a mucosal antibody and provides the major neutralizing activity in the gut. The constant region portion of the single copy α chain gene (Shimizu *et al.*, 1982) has been sequenced (Tucker *et al.*, 1981). We show here that a murine plasmacytoma and normal B cells express three major kinds of α mRNA. We have sequenced the germline DNA that spans the carboxy-terminal end of the α_m chain and have found that, unlike all of the other murine H chain genes that have been described, a single exon ($\alpha M1$) encodes both the carboxy-terminal amino acids of α_m and the 3'-untranslated region of the α_m mRNA. Multiple α mRNA expression from a single gene can be explained by a model of alternate precursor control which may be differentially regulated during B cell development.

Results

mRNAs from an IgA-producing plasmacytoma encode two α chains

The BALB/c plasmacytoma, XRPC 24, synthesizes and secretes IgA(κ) which has anti- β -(1 $\rightarrow$ 6)-D-galactan activity (Rudikoff *et al.*, 1973). Immunofluorescent analysis of XRPC 24 cell membrane Ig using anti-IgA revealed a low but

detectable level of cell surface staining (F.D.Finkelman and J.F.Mushinski, unpublished results), indicating that, in addition to secreting IgA, XRPC 24 also expresses IgA on the cell membrane (mIgA). To examine the RNAs coding for these forms of IgA, XRPC 24 poly(A)⁺ mRNA, before and after sucrose gradient fractionation, was translated *in vitro*, and the protein products were analyzed for the presence of α chains. Two major bands were specifically immunoprecipitated with the anti- α antisera (Figure 1). These were not present in the anti- κ control (Figure 1), nor in the normal rabbit serum or anti-bovine serum albumin antiserum controls (data not shown). The predominant species of ~58 000 daltons is similar in size to the α chains secreted by other murine plasmacytomas (Potter, 1972), and we assume that it is the unprocessed α_s chain. The larger species with an approximate mol. wt. of 62 000 daltons is not as abundant as the α_s species. Its appearance correlates with the presence of larger mRNAs in the *in vitro* translation system as it is clearly enriched in the translation products of RNA from fractions 16 (Figure 1) and 17 (data not shown). Since the membrane forms of other murine Ig H chains have been shown to be larger than the secreted forms (Bergman and Haimovich, 1978; Maki *et al.*, 1981; Word and Kuehl, 1981), we assume that this is the precursor for the α_m chain. *In vitro* translated α chains of similar sizes have also been reported to be coded by mRNAs from a B lymphoma, BFO.3 (Kikutani *et al.*, 1981).

Three major species of α chain mRNA are synthesized in plasmacytoma and normal splenic and Peyer's patch B cells

The observation that two α chain polypeptides are translated *in vitro* from XRPC 24 mRNA suggests that there are a minimum of two α chain mRNA species. XRPC 24 RNA from fractions of a sucrose gradient were, therefore, Northern blotted with a ³²P-labelled fragment that contains only C α 2 and C α 3 sequences. Three species of XRPC 24 total poly(A)⁺ RNA hybridize to the C α probe (Figure 2, lanes labelled A⁺). The smallest, most intensely hybridizing mRNA is ~2.0 kb long and is also seen in fractions 13–17. This corresponds to the size of mRNA appropriate to encode the α_s chain as predicted from the C α sequence (Tucker *et al.*, 1981). The two larger α mRNA species have molecular sizes of 2.5 kb and 3.4 kb. We postulate that one or both of these mRNAs encodes the α_m chain. The 2.5-kb α_m mRNA ($\alpha_{m2.5}$) is present primarily in fractions 15 and 16 while the 3.4-kb α_m mRNA ($\alpha_{m3.4}$) is present in fraction 17. We conclude that these α RNAs are cytoplasmic since all three species are present in microsomal preparations (data not shown). Larger RNAs in XRPC 24 poly(A)⁺ RNA that are also seen to hybridize to the C α probe are probably precursors or aggregates of the three cytoplasmic mRNA species. The synthesis of three cytoplasmic α mRNA species in other IgA producing tumors has also been reported (Kikutani *et al.*, 1981; G.Milburn and R.Lynch, personal communication). It is possible that both $\alpha_{m2.5}$ and $\alpha_{m3.4}$ mRNAs contain an α membrane exon, and that the two mRNA species differ in size because they contain different lengths of 3'-untranslated sequences which terminate near different end sites, analogous to the situation we reported for δ chains (Cheng *et al.*, 1982; Fitzmaurice *et al.*, 1982).

We examined these mRNAs for the presence of V_H region sequences by probing XRPC 24 total poly(A)⁺ RNA or RNA from the fractions of a sucrose gradient (as above) with a ³²P-labelled XRPC 24 V_H probe (Figure 2). A pattern of

Fig. 1. *In vitro* translation of XRPC 24 poly(A)⁺ mRNA. 1 μ g of XRPC 24 total poly(A)⁺ mRNA or poly(A)⁺ mRNA from two fractions (15,16) of a sucrose gradient was translated in a rabbit reticulocyte lysate system containing [³⁵S]methionine. Translation products were specifically immunoprecipitated with antisera against mouse α or κ chains, and the complex was absorbed to Cowan strain staphylococci. The immunoprecipitates were reduced with β -mercaptoethanol and electrophoresed on a discontinuous polyacrylamide gel. ¹⁴C-Labelled mol. wt. markers (MM) (New England Nuclear) were included in the gel. A photo-fluorogram of the slab gel is shown. The minor bands are probably artefacts since they are observed only in fraction 15. However, as discussed by Fitzmaurice *et al.* (1982) for δ chains, the possibility exists that there are α gene products expressed transiently or at low levels.

hybridization identical to that observed with the $C\alpha$ probe is seen. This indicates that the α_s , $\alpha_{m2.5}$ and $\alpha_{m3.4}$ mRNAs from the XRPC 24 plasmacytoma encode complete α_s and α_m chains and suggests they are encoded by a single rearranged α gene. Consistent with this conclusion is the presence of a single rearranged *KpnI* fragment in XRPC 24 DNA that spans the V_H and $C\alpha$ gene in Southern blot analysis (data not shown).

We compared the sizes of α mRNAs in XRPC 24 tumor

cells to those of normal cells. Total RNA was isolated from non-immune BALB/c spleen and Peyer's patch cells, Northern blotted and probed with the $C\alpha$ probe (Figure 2). The same three sizes of α mRNA (2.0 kb, 2.5 kb, 3.4 kb) found in the tumor cells were also present in the normal tissues. Even though the Peyer's patches along the intestine and splenic tissue represent the classical sources for normal membrane IgA⁺ B cells, their α mRNA expression is dominated by the secreted form. This can be attributed to contaminating

Fig. 2. Northern blot analysis of XRPC 24 and normal cell α chain RNA species. 5 μ g of XRPC 24 total poly(A)⁺ RNA (lanes labelled A⁺), 1 μ g of poly(A)⁺ RNA from sucrose gradient fractions 13, 14, 15, 16 and 17, and 25–50 μ g of total RNA from spleen and Peyer's patch tissue of non-immune BALB/c mice were electrophoresed and blotted as described in Materials and methods and probed with ³²P-labelled nick-translated DNA fragments. The $C\alpha$ cDNA probe is the 660-bp *PvuII*-*PvuII* fragment from the α_s cDNA clone p α (J558)¹³, and the XRPC 24 V_H probe is the 258-bp restriction fragment from the V_H cDNA clone (Mushinski *et al.*, in preparation). Mol. wts. were determined by ethidium bromide staining of 18S and 28S rRNA, tobacco mosaic virus RNA and brome mosaic virus RNAs. The XRPC 24 2.0-kb α mRNA is clearer in the shorter exposure (first panel). The second panel is an overexposure of the first to emphasize the 2.5-kb and 3.4-kb mRNAs. We assume that the smears below the 2.0-kb mRNA are due to degradation.

Fig. 3. Partial restriction endonuclease map of the membrane exon of the murine immunoglobulin α heavy chain gene and DNA sequencing strategy. A partial restriction endonuclease map is shown for the 4.5-kb *EcoRI* restriction fragment in p α 4.5. The tall boxes represent coding sequences and the short boxes represent 3'-untranslated regions (open boxes for the $C\alpha$ region; closed boxes for the $\alpha M1$ region). The partially open box in the 3'-untranslated region of $\alpha M1$ (B) indicates the region that is intronic in a minor species of α_m mRNA. Solid lines indicate intervening sequences. End sites are identified by white dots in the boxes. The black dots and arrows below the map in B indicate the DNA sequencing strategy. Fragments were cleaved and labelled at the dots on their 5' ends for all of the fragments shown, except for the *TaqI* fragment which was labelled at its 3' end. The arrows indicate the direction and length of the nucleotide sequences determined by the Maxam and Gilbert method. All nucleotide sequences were determined using p α 4.5 DNA with the exception of the 259-bp sequence extending rightward (3') from the last *EcoRI* site (right-most arrow) which was determined using CH28-S107R DNA.

IgA-secreting plasma cells which are also present in these tissues in mature animals not reared in germ-free conditions. Note that the ratio of the 2.5-kb to 3.4-kb mRNAs is reversed in the normal *versus* tumor cells.

A genomic DNA clone contains the membrane exon of the murine α chain

The plasmid p α 4.5 was constructed as described in Materials and methods. A partial restriction endonuclease map of the 4.5-kbp genomic DNA insert in p α 4.5 is shown in Figure 3A. To map the exon(s) which encodes the carboxy terminus of the α_m chain, p α 4.5 plasmid DNA was digested with several restriction endonucleases, electrophoresed on agarose gels and blotted to nitrocellulose paper. Total poly(A)⁺ XRPC 24 mRNA was partially hydrolyzed, labelled with [γ -³²P]ATP and used to probe the blots. All of the DNA fragments known to contain α sequences hybridized as expected to the XRPC 24 mRNA. No DNA fragments between the 3' end of the α 3 exon (α S) and the *Xba*I site hybridized (data not shown) which suggests that the α membrane exon(s) was located 3' to the *Xba*I restriction site.

Approximately 1.9 kbp of p α 4.5 DNA which spans the region of positive hybridization was sequenced (Figures 3B and 4). The sequence of the 259 bases located 3' to the second *Eco*RI site was determined from CH28-S107R DNA. The DNA sequence revealed several noteworthy features. The first is the presence of the sequence 5'-TCTTACAG/A-3' at position 481 which matches the consensus acceptor (3') splice site [(A)_nN⁺AG/G] (Breathnach *et al.*, 1978; Mount, 1982). Splicing of the proposed donor splice sequence in the α 3

exon (5'-CGG/GTAAAC-3'; Tucker *et al.*, 1981) to this acceptor sequence gives an open reading frame for 65 codons. The translation product predicted from the DNA sequence (Figure 4) would contain a sequence of 25 hydrophilic amino acids (spacer), followed by a stretch of 26 hydrophobic amino acids (trans-membranal peptide), and a unique sequence of 14 hydrophilic amino acids (cytoplasmic peptide) (see Discussion). The nucleotide sequence encoding the trans-membranal peptide in α is highly homologous with the corresponding sequences for μ (69%), δ (63%), ϵ (43%) and the γ (67%) genes.

To prove formally that the α membrane exon begins at this acceptor, two fragments (A and B, Figure 5) which span this putative junction were used in S1 nuclease mapping studies. As expected, the predominant DNA fragments of 121 bases for A and 166 bases for B were protected.

A second striking feature of the sequence is that the consensus donor splice site (5'- $\bar{C}$ AG/GT $\bar{A}$ AGT-3') (Breathnach *et al.*, 1978; Mount, 1982) used to terminate the M1 exons in all of the murine membrane H chain gene sequences (Figure 9) is not present in the α M1 exon at position 641. Instead a sequence 5'-ACTGTTCTGA-3' (termed 'degenerate donor' in Figure 4) that does not conform to the consensus donor splice sequence is present. Splicing of an α M1 exon to an M2 exon (as shown below) does not occur here.

Lastly, three end sites (AATAAA) were identified at positions 1024 (end site 2), 1965 (end site 3) and 2015 (end site 4). Termination and polyadenylation at end site 2 is consistent with the production of the $\alpha_{m2.5}$ mRNA. Termination and polyadenylation at end site 3 or end site 4 is consistent with the production of the $\alpha_{m3.4}$ mRNA.

TACACGAGTCCTCTACCACTCCAGCCTTGGTGGTCTAACTTTGGCCCGAAATCATTGGCCTGTATCTGTTGAGGGTGTCTGATCGTTACACAACCC	100
TAGTACCCAAGCCAAGATTAGCCTAAGAACTGGGGTTTAGGGAGGTATTGCTAAATATCTAGAAAGGACTGAGACCTAAGGAGCCTTGGGTTCAAACCTCT	200
TCTCAGAGTTTGACAGAAACAACCTCTTACCATGACAGCGTCTCTAATTAAGGTTTCCCTAAGAGAGAGATGGCGTCTTGGGGCCAGTCTCAAGAAC	300
TGCTGGGATTCCAGTCCCTTACCTGGGACCCCTGCCTCCTCATTATCCCTGAACAAAGTCCCTTCCCGTGAATAAGGGACCCCTACCACCACTGTCCCCAA	400
AGCCCAAGGAGGCATAGAAGCAGAAACCTTAAGTCTGAGCCCCATCCCCATGGGTTTGATCTACGAAGTCATCTATGTCTTACAGAACGCTCAAGAGC	500
α MEMBRANE EXON: \ R Q E P	
CACTTTCTATGTGCTACTGGACAGTCACAAGACATCCTGGAGGAAGAGGCCCCAGGTGCCAGCCTGTGGCCACCAGTGTGACCTTCTCACCCTCTT	600
L S Y V L L D Q S Q D I L E E E A P G A S L W P T T V T F L T L F	
DEGENERATE DONOR ↓	
CCTACTGAGCTTGTCTACAGCACAGCACTCACTGTTACAACCTGTTGAGGCCCGTTTGGCAGCAAAGAGGTCCCCAGTACTGACCAGGAGCCAGCCAC	700
L L S L F Y S T A L T V T T V R G P F G S K E V P Q Y	
AGGTGGCAGTCATGGAGATGGTGGGTACAGGGTGGGGCAGGGGCACTCAGGATTTACTAGGCTCCTCCATATGTGGCCTTGCAACCTTTCACAGACAC	800
EXONIC/INTRONIC REGION \	
TAAGGAAGTGGACGTAATGATGGCGGCACTTCTGGGTTTAAACCCAGTCTTCCAGAAGCCAGGCTGGGACCTCAGCAAAAGGGAAGACATAACCCAACT	900
CTAGGATGCCCTGTGGCTGATCACTCTGTCTTGGCGGGGATGAAGGTTGGAGAAAGGCCCTCCAGTTTCAGAGCTAAGACCGACGTGACAACCTTCCAG	1000
CCCCTCAGACACTTCTCGAGC aataaa TTGATGCAAAACCACAACTCCTACTCTCAAAAACAAACATTAAACAAAGGATTGGGGAGGGGGTCAGGGG	1100
TAGTATGGTGGCGTTGGGACGGGTAACACAGGGTGCCCATTTGTCTAATGTCTGAGACATAACTTGAACATATGTGTAGCTGCAGCCAAAGATGAACAAGT	1200
GATGGTATTTGTGCTCTTTCAGACCCGATACCAAGTCAATTAAGCTTGATATCTGGACCTGATTTCTAAACATTGGCCTCTGTGAGCACCCCTTTGCAAA	1300
TACTGAACAGCAAAACCTGGTCTGGCTGTGAACCTTGGTCTGATCACTGAGCCCTATATTGGTAACCTGAACCGTGATCTCTGACCTCAGTCA	1400
TGGTCATGAGGCCCTGGTCTGGCCACTCATTCTAATAACTGGTCCCTAGTCTGAGCCCTGAACCTTGGTCCAGGTCACTGAATCCTAGTCATGATCAC	1500
TGGGTTTGATCCTCATTACTGAGCCTAGCTTGATCACCAGTACTGGTTTGTATCATAGCACTAAACCTGATCACTAGTCCCGTTCCTCATCACTGGGCC	1600
ATGATACTGGTCACTGGGTCCTGATCTGATCACTAAATCCTGTTTCTAAACAATGTGTAGTGAATGTATAGTGAAGCCTTTGTGTCTGGCTCTGGGTG	1700
AAATGTCTCAGCAGAGCCCTTGCTAGGTTTGGGTTAATCGGTTGGGGCTGAGAAATGTTTGGAGGCTGTTTGAACCTCAAAAGAAGAAATGTCTCCCTG	1800
ACAATCTGCACATTGGCAGCTGCGCAAAACCTTCATCCTAAATCTTAACCTCTGGCAAGCTTAGAATTACCTTTAAAAACGGCTAGCCACGGTTGAAAG	1900
GGACTGAGATGTCTGTGGGTGGGTGGAACCTTCCAGCTCCAAGAACTCTGTATAATGTTT aataaa GTAACCTGAAGCGAGCTAGCTGGGGTCAATC	2000
TTCTTTCAAGGAG aataaa GCCCTCCGCTCCTCCAGAAATGAAGGCTTAGCTCCTGGGTAGCTTCTCTCTACCGCGGCACCTACAACCAACTCAG	2100
CAGTCTAGGTTCTGTCAACAGA	

Fig. 4. Nucleotide sequence of the murine immunoglobulin α heavy chain membrane exon and flanking sequences. The nucleotide sequence was determined as described in Figure 3 from p α 4.5 and CH28-S107R DNA. (/) identifies 5' donor splice junctions, (\) identifies the 3' acceptor splice junction. The arrow at position 644 indicates a non-functional donor splice site (degenerate donor) used in all other Ig membrane exons. End sites (aataaa) are shown in bold type. The amino acid sequence predicted from the DNA sequence for the α membrane exon is displayed directly below the first base of the corresponding codon using the single letter amino acid code. A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

Both the 2.5-kb and 3.4-kb α heavy chain mRNA species contain the membrane exon

To confirm that the proposed α_m mRNA species indeed contain the membrane exon ($\alpha M1$) and are terminated at the designated end sites, XRPC 24 mRNA was Northern blotted with ^{32}P -labelled restriction fragments from $p\alpha 4.5$. Probe A (Figure 6), which contains the proposed membrane exon and end site 2, hybridized to both $\alpha_{m2.5}$ and $\alpha_{m3.4}$ mRNAs. Identical results were obtained when RNA from normal BALB/c spleen and Peyer's patch tissue were probed with this fragment (data not shown). Probe B (Figure 6), located 3' to the

Fig. 5. DNA-RNA hybridization and S1 nuclease protection experiments. Restriction endonuclease fragments from $p\alpha 4.5$ DNA (A–G) or CH28-S107R DNA (H) were labelled at their 5' ends with $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ or 3' ends with $[\alpha\text{-}^{32}\text{P}]\text{dCTP}$, digested with a second restriction enzyme (except for C) and isolated on 7.5% polyacrylamide gels. Sucrose gradient fractionated XRPC 24 poly(A)⁺ mRNA (16 and 17) and yeast tRNA (y) were denatured, hybridized with the appropriate end-labelled DNA probe and digested with S1 nuclease as detailed in Materials and methods. The S1 nuclease digests as well as untreated labelled DNA fragments (–) were separated on 7% (for A,B,E,F,G) or 20% (for C,D) polyacrylamide/8.3 M urea sequencing gels. Labelled *Hind*III fragments of pBR322 and DNA fragments sequenced by the Maxam and Gilbert method were always run on gels for size determination. Autoradiograms of the gels are shown above. When the RNA-DNA hybrids were short, band clusters were consistently observed (e.g., D). In these cases the fragment that corresponds to the size predicted from the DNA sequence is identified. Below the autoradiograms is a partial restriction endonuclease map. The solid box indicates the $\alpha M1$ coding region. The wavy line indicates the intervening sequence (see Figure 4 and text) deleted in a minor population of XRPC 24 α_m mRNAs. The dots represent end sites. The sizes of the intact labelled DNA fragments are aligned below their corresponding positions on the map. The solid bars indicate the 5' or 3' end-labelled (*) DNA fragments. The solid lines above indicate the major sizes of DNA protected by XRPC 24 α_m mRNA. The wavy lines indicate the sizes of the minor DNA fragments protected by the XRPC 24 α_m mRNA. (A) 5' end-labelled *Nco*I-*Alu*I; (B) 5' end-labelled *Nco*I-*Bgl*II; (C) 3' end-labelled *Dde*I-*Dde*I; (D) 3' end-labelled *Taq*I-*Pst*I; (E) 5' end-labelled *Bgl*II-*Stu*I; (F) 5' end-labelled *Rsa*I-*Hind*III; (G) 5' end-labelled *Hind*III-*Eco*RI; (H) 3' end-labelled *Eco*RI-*Bgl*II.

above fragment and between end sites 2 and 3, hybridized only to the $\alpha_{m3.4}$ mRNA. Similarly, a ^{32}P -labelled restriction fragment from CH28-S107R DNA, which contains the sequences 3' to the *Eco*RI site (including both end sites 3 and 4) hybridized only to the 3.4-kb mRNA (data not shown).

We concluded from these data that both the $\alpha_{m2.5}$ and $\alpha_{m3.4}$ mRNAs contain the $\alpha M1$ exon and are terminated near the end sites 2 and 3 or 4, respectively, as identified in Figures 3 and 4. Analyses of R-loops confirmed this conclusion. XRPC 24 poly(A)⁺ mRNA from fractions 15–17 was hybridized with DNA from $p\alpha 4.5$ (data not shown) and CH28-S107R (Figure 7). Three major species of R-loops were identified (Figure 7) which correspond to the 2.0-kb α_s mRNA (A), and the 2.5-kb (B) and 3.4-kb (C,D) α_m mRNA species.

The lengths for the $\text{Ca}2$, Ca and $\text{Ca}3 + \alpha S$ exons determined from the R-loops of α_s mRNA (Table I) correspond very closely to the lengths expected from the DNA sequence of the Ca gene (Tucker *et al.*, 1981). Similarly, the lengths for the $\text{Ca}1$ and $\text{Ca}2$ exons, and for the spliced $\text{Ca}3$ exon in R-loops formed by the $\alpha_{m2.5}$ and $\alpha_{m3.4}$ mRNAs correspond to that predicted by the sequence. The measurements of the I1 and I2 introns are consistently small for all of the α mRNA species, probably due to the difficulties in measuring short lengths on R-loop preparations. An intron of ~2100 bp in

Fig. 6. Northern blot analysis of the XRPC 24 α_m mRNA species. 5 μg of XRPC 24 total poly(A)⁺ mRNA (lanes labelled A⁺) or 1 μg of poly(A)⁺ mRNA from sucrose gradient fractions 15, 16 and 17 were electrophoresed on a 1% agarose gel containing 5 mM CH_3HgOH , transferred to DPT paper and probed with ^{32}P -labelled nick-translated DNA fragments from $p\alpha 4.5$. **Probe A** (shown on map) is the ~1300-bp *Pst*I-*Pst*I fragment which contains the proposed coding region (open box) for the carboxy-terminal end of the α_m chains and end site 2 (dot). **Probe B** is the 625-bp *Hind*III-*Eco*RI fragment located 3' to **probe A**. Mol. wts. were determined as in Figure 2. A small amount of hybridization to the 2.0-kb α_s mRNA species is seen with both **probes A** and **B** in the A⁺ lanes (but not in lanes 15, 16, 17) due to minor contamination of the probes with Ca sequences.

Fig. 7. R-loop analysis of XRPC 24 α_s and α_m mRNA species. CH28-S107R DNA was hybridized at 7.5 $\mu\text{g}/\text{ml}$ with 2–3 mg/ml of XRPC 24 RNA from sucrose gradient fractions 15–17 as described in Materials and methods. Schematic drawings of the photographed R-loops are shown with solid lines indicating DNA and dotted lines indicating mRNA. I1 is the intron between the Ca1 and Ca2 exons, I2 is the intron between the Ca2 and Ca3 exons, I3 is the intron between the Ca3 and the α M1 exons and I4 indicates an intron within the M1 exon. V_H indicates the variable region of the mRNA. A indicates the poly(A) tail of the mRNA. (A) 2.0-kb α_s mRNA, (B) 2.5-kb α_m mRNA, (C) 3.4-kb α_m mRNA, (D) 3.3-kb α_m mRNA containing a small intron in the M1 exon.

length is present in the R-loops of both the 2.5- and 3.4-kb α_m mRNA species. This value is close to that predicted (~ 2350 bp) from restriction endonuclease mapping and DNA sequencing (Figures 3,4) for the intervening sequence between Ca3 and the α membrane exon. R-loops of the 2.5-kb α_m mRNA identify an exon that corresponds in length to that predicted from the DNA sequence assuming that the membrane exon begins at position 489 in Figure 4 at the consensus acceptor splice site and terminates near end site 2 (position 1024). Likewise, R-loops of the 3.4-kb α_m mRNA contain a longer membrane exon consistent in size with an exon beginning at the same acceptor splice site as above and terminating near end site 3 or 4 (positions 1965 and 2015, respectively).

A sequence in the α M1 exon is both exonic and intronic

In addition to R-loops containing the three major α mRNA species, a few R-loops (Figure 7D) of $\alpha_{m3.4}$ mRNA were identified that appeared to contain a small intron in the membrane exon (see below). This suggested that a small portion of the α membrane gene may be either retained in mature α_m mRNA or spliced during its maturation. To test if the degenerate donor sequence (Figure 4) is used in this minor population of α_m mRNAs, a fragment (C, Figure 5) was used in an S1 nuclease protection mapping experiment. Most of the protected DNA migrated at the size of the full length DNA probe. There was no evidence of the 38 base long fragment expected if the degenerate donor were used. Instead, a minor band of 97 bases was present. The 3' end of this fragment maps to a sequence (5'-CAG/GTGGCA-3') at position 699 (Figure 4) which is similar to the consensus donor splice sequence. S1 nuclease mapping studies using fragment D similarly identified a minor donor splice site at position 699. Two other potential donor splice sequences 5'-A₆GGTGGGG-3' are located just 16 and 28 bases 3' to this donor splice site, but no protected DNA fragments that map to these sites were detected. It should be noted that since this donor splice site is located 3' to the stop codon terminating the α_m protein, the translation products from both the minor α_m mRNA species (in which this sequence is intronic) and the major α_m mRNA species (in which the sequence is exonic) would be identical.

The 3' boundary of the intron in the M1 exon was mapped using fragments E and F (Figure 5). Most of the protected fragments for both probes using both fractions 16 and 17

Table 1. Length of introns and exons measured from R-loops of α_s and α_m mRNA molecules

Molecule	Ca1	I1	Ca2	I2	Ca3 $\pm$ α S	I3	α M1	N ^a
α_s	326 $\pm$ 37 ^b (303) ^c	183 $\pm$ 42 (236)	312 $\pm$ 71 (336)	178 $\pm$ 41 (205)	435 $\pm$ 67 (405)	— (—)	— (—)	8
$\alpha_{m2.5}$	353 $\pm$ 53 (303)	172 $\pm$ 42 (236)	351 $\pm$ 72 (336)	158 $\pm$ 49 (205)	371 $\pm$ 65 (333)	2108 $\pm$ 322 (~ 2350)	532 $\pm$ 85 (560)	18
$\alpha_{m3.4}$	338 $\pm$ 86 (303)	166 $\pm$ 42 (236)	325 $\pm$ 92 (336)	164 $\pm$ 55 (205)	314 $\pm$ 67 (333)	2061 $\pm$ 281 (~ 2350)	1432 $\pm$ 362 (1500)	14

^aN = number of molecules measured.

^bSizes are given in base pairs $\pm$ the standard deviation.

^cNumbers in parentheses are sizes in base pairs expected from the DNA sequence or from mapping data (for I3).

mRNA, but not yeast tRNA, were full length, again confirming that the majority of the α_m mRNA species contain no introns. Approximately one-fourth of the protected fragments, as determined by densitometric scans, maps to a consensus acceptor splice sequence (CCCTTCACAG/A) at position 787, just 94 bases 3' to the donor splice site. Several other potential acceptor splice sequences are present within the 94-bp intron, but there is no evidence that any of these are used. Since fragment F can be protected only by α_m mRNA terminated near end sites 3 or 4 (i.e., $\alpha_{m3.4}$ mRNA) but not by mRNAs terminated at end site 2 (i.e., $\alpha_{m2.5}$ mRNA), and since the minor splicing is seen to occur within this fragment, we conclude that $\alpha_{m3.4}$ mRNA can be processed to an ~ 3.3 -kb ($\alpha_{m3.3}$) mRNA species. No additional introns are present in the $\alpha_{m3.4}$ or $\alpha_{m3.3}$ mRNA since the entire fragment G is specifically protected by the poly(A)⁺ XRPC 24 mRNA from fractions 16 and 17, but not by yeast tRNA.

Both end site 3 and end site 4 are recognized for termination or polyadenylation of the 3.4-kb α_m mRNA

To examine whether the sequences identified as end sites 3 and 4 in Figure 4 are both used for the termination and/or polyadenylation of the $\alpha_{m3.4}$ mRNA, the following S1 nuclease protection experiment was performed. A fragment (H, Figure 5) of ~ 1600 bp was isolated from CH28-S107R that contains 259 bp of mouse genomic DNA (bases 1865–2124 in Figure 4) including both end sites 3 and 4 and part of the right arm of Charon 28 DNA. mRNA from fraction 17, but not yeast tRNA, protected two clusters of fragments from S1 nuclease digestion. The major cluster of

bands maps to ~ 20 bp 3' to end site 3, while approximately one-fourth of the protected fragments maps to ~ 20 bp 3' to end site 4. Since polyadenylation of RNAs generally occurs 15–20 bases 3' to the AAUAAA sequence (Proudfoot and Brownlee, 1976), we conclude that both end sites are recognized for the termination and polyadenylation of the 3.4-kb α_m mRNA, although end site 3 appears to be the preferred site.

Discussion

Multiple forms of α mRNA are expressed in a plasmacytoma and in normal B cells

A single murine α gene contains four poly(A) recognition/transcription termination sites, or end sites (AATAAA). Three are in the DNA sequence of Figure 4 and one is in the C α DNA sequence presented elsewhere (Tucker *et al.*, 1981). We have shown here that these sites are all used for expression of the three major forms of α mRNA and as many as four minor forms of α_m mRNA. Figure 8 shows the relationships of these multiple α mRNA species to the four end sites and the splicing patterns postulated to generate them from primary transcripts (preRNA). A 2.0-kb mRNA, α_s , which codes for the secreted α chain, accounts for $>90\%$ of the α mRNA present in XRPC 24 cells. This mRNA is terminated near end site 1 only 26 bp from the end of the C α coding sequence. The second and third most abundant species are the 2.5-kb and 3.4-kb α_m mRNAs ($\alpha_{m2.5}$ and $\alpha_{m3.4}$) which account for $\sim 6\%$ and 2% , respectively, of the α mRNA in the XRPC 24 plasmacytoma cells. The $\alpha_{m2.5}$ mRNA is terminated at end site 2 (base 1024 in the DNA sequence in

Fig. 8. Transcription and splicing patterns for murine α chain mRNA precursors. The top line gives a map of the α transcriptional unit beginning at the putative promoter (►). Open boxes indicate the coding regions for the leader (L), recombined variable, diversity and joining (V or VDJ), C $\alpha 1$, C $\alpha 2$, C $\alpha 3$ and α secreted (S) exons. The solid boxes indicate the α membrane (M1) coding region. Solid lines indicate intervening sequences and 3'-untranslated regions. Dotted lines indicate the regions of RNA which are not found in mature mRNA and are deleted by RNA splicing from the primary RNA transcript (pre RNA). AAAA indicates poly(A) tails. The (AAAA) indicates that it is not clear whether the small intron in the 3'-untranslated region of the M1 exon can be spliced from $\alpha_{m2.5}$ preRNA terminated near end site 2 to produce a 2.4-kb α_m mRNA. This intron is known to be spliced from a minor population of $\alpha_{m3.4}$ preRNA terminated near end sites 3 or 4 producing a 3.3-kb α_m mRNA.

Figure 4). The $\alpha_{m3.4}$ mRNA appears to be a family of mRNAs, the majority of which are terminated at end site 3 (position 1965); the remainder are terminated at end site 4 (position 2015). Another minor population of α_m mRNA expressed in XRPC 24 cells, $\alpha_{m3.3}$, that is indistinguishable from the $\alpha_{m3.4}$ mRNA by Northern blotting, appears to be derived from $\alpha_{m3.4}$ by removal of a 94 base intron from the 3'-untranslated region. Two forms of 3.3-kb α_m mRNA may exist if both end sites 3 and 4 are used in these molecules. Since the $\alpha_{m2.5}$ mRNA also contains the sequence for the 94 base intron, a seventh possible species of α mRNA would result if it were occasionally spliced to create an $\alpha_{m2.4}$ mRNA. We cannot verify if this minor species, $\alpha_{m2.4}$, is also expressed in XRPC 24 cells since S1 nuclease protection experiments could not be performed on isolated $\alpha_{m2.5}$ mRNA nor were R-loops of $\alpha_{m2.5}$ mRNA seen with a definitive intron loop at the expected position.

Since mIgA⁺ cells are sparse in spleen and Peyer's patch tissues, our analyses of their α mRNA expression is limited. Normal B cells express at least the same three major mRNAs (α_s , $\alpha_{m2.5}$ and $\alpha_{m3.4}$), suggesting that termination in the tumor and normal B cells occurs at the same end sites. However, the altered ratio of the $\alpha_{m2.4}$ and $\alpha_{m3.5}$ species in the normal B cells relative to that in XRPC 24 (Figure 2) suggests that the regulation of this termination event may be differentially modulated (cf. below).

Alternative RNA precursors may generate multiple α mRNAs

All Ig H chain mRNAs within a single clone of B cells appear to initiate at the same point, thus the expression of multiple mRNAs from the same DNA template appears to be controlled at transcription termination. This parallels events in adenovirus late transcription (reviewed in Fraser *et al.*, 1982) where the choice of 3' ends determines which mRNA is expressed from a complex transcriptional unit. In this context, it is important to distinguish true alternative RNA splicing from alternative RNA precursor control. Control at the level of RNA splicing requires differential use of optional donor and/or acceptor splice sites by changes directly affecting the splicing machinery. In alternative precursor control, the splicing machinery *per se* remains constant while the splicing pattern of each precursor is dictated only by the primary or secondary structure of the precursor itself. For the expression of multiple α mRNAs we make a case for this latter mechanism.

The simplest case for the alternative precursor model is the shift between the secreted mRNA (α_s) and the two major membrane mRNAs ($\alpha_{m2.5}$ or $\alpha_{m3.4}$). Two separate classes of RNA precursors (pre- α_s versus pre- α_m) terminate at different 3' end sites (end 1 for α_s versus end 2, 3 or 4 for α_m) to generate two separate classes of α mRNAs. The longer (α_m) primary transcript splices together the C_H3 and M1 exons. Since the shorter (α_s) primary transcript lacks the M1 exon (and the requisite M1 acceptor site for RNA splicing), it is impossible to splice out the S region (Figure 8) from this precursor; the secreted terminus is coded as an extension of the C α 3 domain. This pattern is analogous to that described for co-expression of μ_s and μ_m chains (Alt *et al.*, 1980; Rogers *et al.*, 1980; Early *et al.*, 1980; Singer *et al.*, 1980) which is often referred to as alternative splicing when, in fact, no alternation in the specificity of splicing is required. Recent data (see Introduction) suggest that primary transcripts may frequently extend well beyond the 3' end of the mature message. Assuming the same is possible for Ig genes, the end sites in

Figure 8 should be viewed as secondary recognition sequences for either a specific RNA endonuclease, poly(A) polymerase, or both. But even if an extended α_s preRNA of up to 2 kbp longer than drawn in the diagram were synthesized, it would still lack the necessary membrane sequences required for splicing out the S exon.

The more complex case of alternative precursor control is the differential expression of as many as six different α_m mRNAs. The model in Figure 8 proposes that the $\alpha_{m2.5}$ (end site 2) and $\alpha_{m3.4}$ (end site 3 or 4) mRNA species arise from distinct primary transcripts. Conversely, an extended precursor could give rise to three pre- α_m RNAs if alternative cleavage occurs at end sites 2, 3 or 4. There is still no logical requirement for true alternative splicing; however, RNA splicing must account for some of the minor α_m mRNAs (e.g., $\alpha_{m3.3}$ in Figure 8).

Multiple α membrane messages encode a single α membrane protein

The V_H data in Figure 2 suggests that no significant heterogeneity resides in the 5' end of these transcripts. If the length differences between the multiple forms of α_m mRNA result only from alternative termination and splicing within the 3'-untranslated region, all of the α_m mRNAs should encode an identical protein. Cell-free translation of total poly(A)⁺ XRPC 24 mRNA and sucrose fractions enriched for the 2.4-kb and 3.5-kb α_m mRNA species (Figure 1, fraction 16) produced, in addition to the α_s protein, only one larger immunoprecipitable polypeptide of a size that correlates with that predicted for α_m from the DNA sequence.

Precedents exist for 3' end heterogeneity generating multiple mRNAs that code for the same protein. Alternative termination of the mouse Ig δ chain gene at its two most distal end sites (Figure 9) produces either a 2.1-kb or 2.9-kb δ_m message (Cheng *et al.*, 1982; Fitzmaurice *et al.*, 1982). Like the $\alpha_{m3.3}$ mRNA, a few of the δ_m 2.9-kb mRNAs are also spliced within the 3'-untranslated region to generate even greater end heterogeneity.

Control of multiple termination in B cell differentiation

The function of 3' structural polymorphism that does not affect the coding potential of an mRNA is unclear. Possibilities include increased transport of smaller forms from nucleus to cytoplasm (Setzer *et al.*, 1982), increased resistance of longer forms to degradation during embryogenesis (Davidson *et al.*, 1982), and increased translational efficiency of longer forms.

A convincing physiological rationale can be made, however, when functionally different proteins are produced, as in the shift between membrane and secreted Ig H chains during B cell differentiation. B cells terminally differentiated to secretion (plasma cells or plasmacytomas) express mRNA terminated at the most promoter-proximal end site. Conversely, in less differentiated cells (resting or memory B cells), a greater percentage of mRNA molecules (e.g., membrane forms) are terminated at more promoter-distal end sites (Figure 8). Blattner and Tucker (1983) proposed that the RNA endonuclease which cleaves the preRNA at end sites plays the key role in the selection of alternative 3' ends. In cells with high endonuclease activity, cleavage occurs predominantly at the first end site encountered by the endonuclease on the nascent RNA chain. Downstream end sites are then cleaved at a rate proportional to their distance from the promoter. Our results support this model. In the XRPC

24 cells, the expression of the different α mRNAs is inversely proportional to the distance of the end sites from the promoter (i.e., $\alpha_s > \alpha_{m2.4} > \alpha_{m3.4}$). However, the expression of $\alpha_{m2.5}$ (end 2) and $\alpha_{m3.4}$ (end 3 or 4) is reversed in spleen and Peyer's patches, consistent with the idea that in less differentiated cells the endonuclease activity is lower.

The AAUAAA sequence alone is not always a sufficient or necessary signal for polyadenylation/termination. This hexanucleotide is present in the coding regions and introns of several genes and is missing from at least three of seven DHFR mRNAs (Setzer *et al.*, 1982). However, comparisons of the DNA sequences flanking the end sites for α -amylase (Tosi *et al.*, 1981), DHFR (Setzer *et al.*, 1982) and δ_m (Cheng *et al.*, 1982) and α_m Igs reveal no features which distinguish major and minor end sites. So it is not yet obvious from DNA sequences alone what other factors might be involved in the regulation of multiple termination.

Structure and function of the α membrane carboxy terminus

The amino acid sequence predicted from the nucleotide sequence of the α M1 exon (Figure 4) is compared in Figure 9 (bottom) to the membrane carboxy-terminal amino acid sequences of the murine μ , δ , γ_1 , γ_{2b} , γ_{2a} and ϵ chains. These se-

quences have been arbitrarily separated into three structural and functional regions based on the similarity of these peptides to other membrane proteins.

The 'spacer' peptide, encoded by the M1 exon in all H chain genes, connects the terminal constant region domain with the 'trans-membranal' peptide and is thought to reside external to the B cell membrane. This hydrophilic sequence is 25 residues long for α_m . With the exception of a stretch of three glutamic acid residues shared by the α_m , μ_m and δ_m chains, there is very little amino acid sequence homology between α and the other membrane chains in this region. However, it is interesting that all of the spacers have a net charge of at least -6 suggesting that all Ig membrane chains require some type of polar interaction to stabilize their binding. The unique nature of the spacer region makes this portion of the molecule an attractive target for isotype-specific regulatory T cells. That is, B cells expressing membrane IgA may be specifically triggered only by T cell factors which recognize the spacer region of the α_m chain. If true, although secreted IgA is present in far higher concentrations in serum than membrane IgA, it could not compete for a T cell factor since it lacks the α_m spacer.

Also encoded by the M1 exon for all H chains is a trans-

Fig. 9. Comparison of murine immunoglobulin heavy chain carboxy-terminal exons and the corresponding membrane amino acid sequences. The organization of the exons encoding the carboxy-terminal ends of the murine Ig H chains for α , μ , δ , γ_1 , γ_{2b} , γ_{2a} , γ_3 , ϵ are shown. Solid boxes indicate coding regions. The large open box (δX) indicates an exon in δ determined by DNA sequencing and Northern blot analysis that may not be used to encode a δ protein. The open boxes indicate 3'-untranslated regions. The dots within the open boxes indicate end sites (AATAAA or AATTAAA). The lengths of the M2 exons for the γ and ϵ genes are not known but have been estimated from the differences in size of the secreted *versus* membrane mRNA species. The report of two cytoplasmic ϵ_m RNAs suggests two end sites are present in the ϵ M2 exon. The sequences of the amino acids encoded by the M1 and M2 exons of the murine Ig H chain genes as deduced from their nucleotide sequence are shown below the map using the single letter amino acid code (see Figure 4). Gaps (-) are introduced to maximize homology. Amino acids in italics type are encoded by the C_H3 or C_H4 exons. Amino acids in bold type are encoded by the M1 exons. Amino acids in light type are encoded by the M2 exons. Division of the peptides into spacer, trans-membranal and cytoplasmic segments is based on structural and functional considerations detailed in the text. Note that the entire α membrane terminus is encoded by a single exon. References: μ (Rogers *et al.*, 1980; Early *et al.*, 1980); δ (Cheng *et al.*, 1982); γ (Rogers *et al.*, 1981; Tyler *et al.*, 1982; Yamawaki-Kataoka *et al.*, 1982; J. Wels and F. Blattner, personal communication); ϵ (Ishida *et al.*, 1982).

membrane peptide with a hydrophobic stretch of 26 amino acids that presumably span the lipid bilayer. This peptide is conserved (9/26 identical) in all murine (Figure 9) and human (Rabbitts *et al.*, 1981) Ig membrane H chains analyzed. The α_m membrane protein is most homologous in this region with μ_m (62%). Other than their extremely hydrophobic character, the membrane Ig trans-membranal peptides share no significant sequence homology with the trans-membranal peptides of other Ig-like molecules such as the major histocompatibility (MHC) class I (Steinmetz *et al.*, 1981; Moore *et al.*, 1982; Evans *et al.*, 1982) and class II (Korman *et al.*, 1982; McNicholas *et al.*, 1982) molecules. These different receptors may utilize their sequence variation to satisfy different inner membrane contacts.

The carboxy-terminal most amino acids are generally hydrophilic and make up the 'cytoplasmic' peptide. This sequence may protrude into the cytoplasm of the B cell and may communicate external signals to intracellular structures (Rogers *et al.*, 1980). Although the cytoplasmic peptides differ in length, sequence and overall charge among the different classes of Ig, a valine is conserved in all of the H chains, and four other amino acids are conserved between the α and the γ chains. We speculate that these conserved residues may be important for intracellular interactions, whereas the differences in amino acid sequence engender an isotype specificity to the signal transmitted. The first and third amino acids of the cytoplasmic peptide for μ , δ , ϵ and γ chains are lysines which may serve to anchor the trans-membranal segment in the membrane (Engelman and Steitz, 1981). The first lysine is encoded by the M1 exon, while the remaining amino acids are encoded by the M2 exon. However, in the α_m chain, only one basic residue (arginine) is conserved, and the entire cytoplasmic peptide is encoded by the M1 exon.

Evolution of the α membrane exon

Is it significant from an evolutionary point of view that α does not split its trans-membranal and cytoplasmic segments into two exons? Rogers *et al.* (1980) postulated that the primordial Ig gene consisted of three exons: a signal peptide, a V-like extracellular domain for ligand recognition, and a trans-membranal sequence. Igs and MHC antigens have probably descended from this same ancestral gene (Steinmetz *et al.*, 1981). Although the α trans-membranal sequence is homologous to other Ig membrane genes (Figure 9) but not to MHC genes, α_m and MHC genes have structural similarities, i.e., the trans-membranal and cytoplasmic peptide segments are fused onto a single exon (Steinmetz *et al.*, 1981; Moore *et al.*, 1982; Evans *et al.*, 1982; Larhammar *et al.*, 1982; Korman *et al.*, 1982; McNicholas *et al.*, 1982).

Ig and MHC genes may have evolved separately from a common ancestral gene that was duplicated. The unique α_m gene structure could have arisen from either duplicated gene and would represent a relic of the past history of this event. We predict that the coding region for the cytoplasmic peptide of α evolved either from the M2 exon of the other Ig genes, or from the intervening sequence directly contiguous to the M1 exon. Nucleotide homology searches have failed to differentiate between these two possibilities. The presence of a GT dinucleotide at the degenerate donor site in the α_m gene (Figure 4) is consistent with the creation of a donor splice site in the primordial gene or elimination of the site in the α gene. According to the GT-AG rule (Breathnach *et al.*, 1978) a functional splice at this site would occur between codons, i.e., in exact register with the translation phase. This type of phas-

ing for Ig genes is unique to the trans-membranal-cytoplasmic splice. In contrast, all other splices within the Ig and MHC genes occur one base into a codon. The selective forces that engineer such a disparate gene structure for such a seemingly similar function will only be understood when the biochemistry of the receptor molecules themselves is understood.

Materials and methods

Recombinant DNA and probes

The plasmid p α 4.5 was constructed by inserting the ~4.5 kbp *Eco*RI fragment from the genomic clone CH4A-140.7 (Tucker *et al.*, 1981) into the *Eco*RI site of pBR322 (Figure 3). The plasmid p α (J558)¹³ containing α cDNA sequences has been described (Tucker *et al.*, 1981). The CH28-S107 recombinant clone (Kirsch *et al.*, 1981) was generously provided by I.Kirsch and P.Leder; it contains, in addition to other sequences, several hundred base pairs 3' to the end of the above *Eco*RI fragment. Restriction enzyme fragments from the clones were isolated by electrophoresis in polyacrylamide gels (Maniatis *et al.*, 1975), ³²P-labelled by nick translation according to the procedure of Rigby *et al.* (1977) to 1–10 × 10⁷ c.p.m./ μ g of DNA and used as probes for Northern blots. The V_H cDNA clone was prepared (Mushinski *et al.*, in preparation) by the method of Maniatis *et al.* (1976) from sucrose gradient fraction 15 of XRPC 24 poly(A)⁺ RNA (see below) and identified by DNA sequence analysis compared with the amino acid sequence of XRPC 24 α chain (Rao *et al.*, 1979). A 258-bp restriction fragment containing sequences encoding the three framework regions and the first and second hypervariable regions was prepared from this cDNA clone insert and labelled by nick translation for use as a hybridization probe.

Isolation of mRNA and Northern blotting

Total RNA was prepared by 4 M guanidinium thiocyanate extraction of XRPC 24 plasmacytoma tissue which had been frozen in liquid nitrogen and pulverized in a mortar and pestle chilled in liquid nitrogen (Chirgwin *et al.*, 1979). Poly(A)⁺ RNA was fractionated by centrifugation on a 5–40% sucrose gradient. The α chain mRNAs of XRPC 24 were studied by denaturation of 1–5 μ g of each RNA sample in 5 mM methyl mercuric hydroxide, transfer to diazotized paper (Seed, 1982) and hybridization (Wahl *et al.*, 1979) with ³²P-labelled nick-translated DNA probes. After hybridization, the paper was washed 3 times at 20°C or 65°C for 5 min each wash in 3 × SSC, 0.1% SDS and once for 60 min in 0.1 × SSC, 0.1% SDS at 42°C or 65°C, and then autoradiographed using Kodak XAR-5 film and intensifying screens. Sizes of hybridizing bands of RNA were determined by reference to ethidium bromide-stained 18S and 28S rRNAs which are assumed to be 2.0 and 4.7 kb, respectively, to tobacco mosaic virus RNA (6.4 kb; Caspar, 1963) and to brome mosaic virus RNAs (0.8, 2.3, 3.1 and 3.4 kb; Lane and Kaesberg, 1971).

Total RNA was isolated from spleen and Peyer's patch cells of 2–3 month old non-immunized BALB/c mice using a modification of the guanidinium thiocyanate extraction procedure of Chirgwin *et al.* (1979). The RNA was denatured by heating for 2 min at 90°C in 50% formamide, 6% formaldehyde, 20 mM Na-phosphate, electrophoresed on a 2% agarose-6% formaldehyde gel (Rave *et al.*, 1979) and transferred to a nitrocellulose filter in 20 × SSC. The filter was then treated as described above.

5' End-labelling of mRNA

Approximately 0.5–1 μ g of XRPC 24 poly(A)⁺ mRNA was incubated in 10 μ l 0.1 M Na borate (pH 9.3) at 65°C, 45 min. The mRNA hydrolysate was 5' end-labelled according to the procedure of Biessman *et al.* (1979). The labelled mRNA was isolated on a 10 ml G-100 Sephadex (Sigma) column in 0.2 M NaCl, 0.1% SDS, 10 mM Tris (pH 7.5), 1 mM EDTA.

Southern blotting

DNA from the plasmid p α 4.5 was digested with various restriction endonucleases (New England Biolabs), electrophoresed on 0.8% or 1.2% agarose (Seakem) gels and transferred to nitrocellulose paper (Schleicher and Schuell) according to the procedure of Southern (1975). The paper was prehybridized, and hybridized with ³²P-labelled XRPC 24 mRNA according to the procedure of Wahl *et al.* (1979). After hybridization, the paper was washed 3 times at 20°C in 3 × SSC, 0.1% SDS and once for 60 min in 0.1 × SSC, 0.1% SDS at 42°C, and then subjected to autoradiography.

DNA sequencing

DNA from p α 4.5 or CH28-S107R was digested with restriction endonucleases, labelled at 5' ends with [γ -³²P]ATP (ICN Biochemical) and polynucleotide kinase (Boehringer-Mannheim), or 3' labelled with [α -³²P]dCTP (New England Nuclear) and the Klenow fragment of DNA polymerase I (New England Biolabs) (Maxam and Gilbert, 1980), sequenced according to the

procedure of Maxam and Gilbert (1980), and electrophoresed on 40 cm long gels of 8% or 20% polyacrylamide/8.3 M urea and on 120 cm long gels of 7% polyacrylamide/8.3 M urea that were 0.3 mm thick (Sanger *et al.*, 1977). Gels were autoradiographed as above.

S1 nuclease protection experiments

The hybridizations and S1 nuclease digestions were performed essentially as described by Sharp *et al.* (1980). Restriction endonuclease digested p α 4.5 plasmid DNA was 5' or 3' end-labelled as above. Less than 50 ng of labelled DNA ($\sim 5 \times 10^3$ c.p.m.) were mixed with 2–4 μ g of poly(A)⁺ XRPC 24 mRNA from fractions 16 or 17 of the sucrose gradient (above), or yeast tRNA, and dried under a vacuum. The DNA-RNA mixtures were resuspended in 8 μ l deionized formamide (Eastman Chemical Co.), heated to 90°C to denature the DNA, and 2 μ l of 2 M NaCl, 0.2 M PIPES (pH 6.4), 5 mM EDTA were added. Hybridization proceeded overnight at 45°C. To terminate the hybridization reaction, 90 μ l of 0.25 M NaCl, 0.03 M potassium acetate (pH 4.5), 1 mM zinc sulfate were added, and the single-stranded nucleic acids were digested with 500 units S1 nuclease (Miles) for 10 min at 37°C. 5 μ g carrier yeast tRNA were added and the samples were ethanol precipitated. The S1 digests were denatured by heating to 90°C in 90% formamide and electrophoresed on a 7% or 20% polyacrylamide/8.3 M urea gel and autoradiographed.

R-loop analysis

DNA from the bacteriophage clone CH28-S107R was hybridized with RNA from XRPC 24 sucrose gradient fractions 15–17. Each hybridization reaction was carried out for 4 h at 52°C in sealed capillaries containing 20 μ l of solution containing 70% deionized formamide, 0.5 M NaCl, 1 mM EDTA, 0.1 M Tricine (pH 8), 7.5 μ g/ml DNA and 2–3 mg/ml RNA. The solution was then chilled to 4°C, spread on a hypophase of 50% formamide, picked up on parlodion-coated grids, stained with uranyl acetate and shadowed with platinum/palladium and carbon on a rotating stage. The spreads were examined with a Zeiss electron microscope, EM109, and measured with a Numonics digitizer interfaced to a PDP-11 computer. The internal standard was pBR326.

In vitro translation

1 μ g of XRPC 24 total poly(A)⁺ RNA and individual fractions of the sucrose gradient (above) was translated *in vitro* using an mRNA-dependent (staphylococcal nuclease treated) rabbit reticulocyte lysate (Pelham and Jackson, 1976) containing L-[³⁵S]methionine (Amersham, 1022 Ci/mmol). The translation products were incubated with rabbit antisera specific for mouse α chains or mouse κ chains and then precipitated with Cowan strain staphylococci (Kessler, 1975) and analyzed by SDS-polyacrylamide gel electrophoresis and fluorography (Bonner and Laskey, 1974).

Acknowledgements

We thank Dr. Stuart Rudikoff and James Owens for their collaboration in generating the V_H probe, Mrs. Margery Sullivan for spreading and photographing the R-loops, Dr. Fred Finkelman for performing the immunofluorescence analysis, Drs. I. Kirsch and P. Leder for generously providing the CH28-S107 recombinant clone and Ms. Madeline Allan for her skilled assistance in the preparation of the manuscript. This work was supported in part by U.S. NIH grants AI 18016 and 5T32 CA 09082. C.J.W. is a Damon Runyon-Walter Winchell fellow.

References

Alt, F.W., Bothwell, A.L.M., Knapp, M., Siden, E., Mather, E., Koshland, M. and Baltimore, D. (1980) *Cell*, **20**, 293–302.
 Bennetzen, J.L. and Hall, B.D. (1982) *J. Biol. Chem.*, **257**, 3018–3025.
 Bergman, Y. and Haimovich, J. (1978) *Eur. J. Immunol.*, **8**, 876–880.
 Biessmann, H., Craig, E.A. and McCarthy, B.J. (1979) *Nucleic Acids Res.*, **7**, 981–996.
 Blattner, F.R. and Tucker, P.W. (1983) *Nature*, in press.
 Bonner, W.M. and Laskey, R.A. (1974) *Eur. J. Biochem.*, **46**, 83–88.
 Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 4853–4857.
 Caspar, D.L.D. (1963) *Adv. Protein Chem.*, **18**, 37–121.
 Cheng, H.-L., Blattner, F.R., Fitzmaurice, L., Mushinski, J.F. and Tucker, P.W. (1982) *Nature*, **296**, 410–415.
 Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry (Wash.)*, **18**, 5294–5299.
 Darnell, J.E. (1982) *Nature*, **297**, 365–371.
 Davidson, E.H., Thomas, T.L., Scheller, R.H. and Britten, R.J. (1982) in Dover, G.A. and Flavell, R.B. (eds.), *Genome Evolution*, Academic Press, NY, pp. 177–191.
 Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L.

(1980) *Cell*, **20**, 313–319.
 Engelman, D.M. and Steitz, T.A. (1981) *Cell*, **23**, 411–422.
 Evans, G.A., Margulies, D.H., Camerini-Otero, R.D., Ozato, K. and Seidman, J.G. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1994–1998.
 Fitzmaurice, L., Owens, J., Blattner, F.R., Cheng, H.-L., Tucker, P.W. and Mushinski, J.F. (1982) *Nature*, **296**, 459–462.
 Ford, J. and Hsu, M.-T. (1978) *J. Virol.*, **28**, 795–801.
 Fraser, N.W., Baker, C.C., Moore, M.A. and Ziff, E.B. (1982) *J. Mol. Biol.*, **155**, 207–233.
 Hellig, R., Perrin, F., Gannon, F., Mandel, J.L. and Chambon, P. (1980) *Cell*, **20**, 625–637.
 Hofer, E. and Darnell, J.E. (1981) *Cell*, **23**, 585–593.
 Hofer, E., Hofer-Warbinek, R. and Darnell, J.E. (1982) *Cell*, **29**, 887–893.
 Ishida, N., Ueda, S., Hayashida, H., Miyata, T. and Honjo, T. (1982) *EMBO J.*, **1**, 1117–1123.
 Kehry, M., Ewald, S., Douglas, R., Sibley, C., Raschke, W., Fambrough, D. and Hood, L. (1980) *Cell*, **21**, 393–406.
 Kessler, S.W. (1975) *J. Immunol.*, **115**, 1617–1624.
 Kikutani, H., Sitia, R., Good, R.A. and Stavnezer, J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 6436–6440.
 Kirsch, I.R., Ravetch, J.V., Kwan, S.-P., Max, E.E., Ney, R.L. and Leder, P. (1981) *Nature*, **293**, 585–587.
 Korman, A.J., Auffray, C., Shamboeck, A. and Strominger, J.L. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6013–6017.
 Lai, C.-J., Dhar, R. and Khoury, G. (1978) *Cell*, **14**, 971–982.
 Lane, L.C. and Kaesberg, P. (1971) *Nature New Biol.*, **232**, 40–43.
 Larhammer, D., Gustafsson, K., Claesson, L., Bill, P., Wiman, K., Schenning, L., Sundelin, J., Widmark, E., Peterson, P.A. and Rask, L. (1982) *Cell*, **30**, 153–161.
 Maki, R., Roeder, W., Traunecker, A., Sidman, C., Wabl, M., Raschke, W. and Toneyawa, S. (1981) *Cell*, **24**, 353–365.
 Maniatis, T., Jeffrey, A. and Kleid, D.G. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 1184–1188.
 Maniatis, T., Kee, S.G., Efstratiadis, A. and Kafatos, F.C. (1976) *Cell*, **8**, 163–182.
 Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.*, **65**, 499–560.
 McNicholas, J., Steinmetz, M., Hunkapiller, T., Jones, P. and Hood, L. (1982) *Science (Wash.)*, **218**, 1229–1232.
 McReynolds, L., O'Malley, B.W., Nisbet, A.D., Fothergill, J.E., Givol, D., Fields, S., Robertson, M. and Brownlee, G.G. (1978) *Nature*, **273**, 723–728.
 Moore, K.W., Sher, B.T., Sun, Y.H., Eakle, K.A. and Hood, L. (1982) *Science (Wash.)*, **215**, 679–682.
 Mount, S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
 Nevins, J.R. and Darnell, J.E. (1978) *Cell*, **15**, 1477–1493.
 Nevins, J.R., Blanchard, J.-M. and Darnell, J.E. (1980) *J. Mol. Biol.*, **144**, 377–386.
 Parnes, J.R. and Seidman, J.G. (1982) *Cell*, **29**, 661–669.
 Pelham, H.R.B. and Jackson, R.J. (1976) *Eur. J. Biochem.*, **67**, 247–256.
 Perricaudet, M., le Moulec, J.-M. and Tiollais, P. (1980) *Nature*, **288**, 174–176.
 Potter, M. (1972) *Physiol. Rev.*, **52**, 631–719.
 Proudfoot, N. (1982) *Nature*, **298**, 516–517.
 Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature*, **263**, 211–214.
 Rabbitts, T.H., Foster, A. and Milstein, C.P. (1981) *Nucleic Acids Res.*, **9**, 4509–4524.
 Rao, D.N., Rudikoff, S., Krutzsch, H. and Potter, M. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 2890–2894.
 Rave, N., Crkvenjakov, R. and Boedtker, H. (1979) *Nucleic Acids Res.*, **6**, 3559–3567.
 Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.*, **113**, 237–251.
 Rogers, J., Early, P., Carter, C., Calame, K., Bond, M., Hood, L. and Wall, R. (1980) *Cell*, **20**, 303–312.
 Rogers, J., Choi, E., Souza, L., Carter, C., Word, C., Kuehl, M., Eisenberg, D. and Wall, R. (1981) *Cell*, **26**, 19–27.
 Rudikoff, S., Mushinski, E.B., Potter, M., Glaudemans, C.P.J. and Jolley, M.E. (1973) *J. Exp. Med.*, **138**, 1095–1106.
 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
 Seed, B. (1982) *Nucleic Acids Res.*, **10**, 1799–1810.
 Setzer, D.R., McGrogan, M. and Schimke, R.T. (1982) *J. Biol. Chem.*, **257**, 5143–5147.
 Sharp, P.A., Berk, A.J., Berget, S.M. (1980) *Methods Enzymol.*, **65**, 750–768.
 Shimizu, A., Takahashi, N., Yaoita, Y. and Honjo, T. (1982) *Cell*, **28**, 499–506.
 Singer, P.A., Singer, H.H. and Williamson, A.R. (1980) *Nature*, **285**, 294–300.
 Southern, E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
 Steinmetz, M., Frelinger, J.G., Fisher, D., Hunkapiller, T., Pereira, D., Weissman, S.M., Uehara, H., Nathenson, S. and Hood, L. (1981) *Cell*, **24**, 125–134.

- Tosi,M., Young,R.A., Hagenbüchle,O. and Schiber,U. (1981) *Nucleic Acids Res.*, **9**, 2313-2323.
- Tucker,P.W., Liu,C.-P., Mushinski,J.F. and Blattner,F.R. (1980) *Science (Wash.)*, **209**, 1353-1360.
- Tucker,P.W., Slightom,J.L. and Blattner,F.R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7684-7688.
- Tyler,B.M., Cowman,A.F., Gerondakis,S.D., Adams,J.M. and Bernard,O. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2008-2012.
- Unterman,R.D., Lynch,K.R., Nakhasi,H.L., Dolan,K.P., Hamilton,J.W., Cohn,D.V. and Feigelson,P. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 3478-3482.
- Vassalli,P., Tedghi,R., Lisowska-Bernstein,B., Tartakoff,A. and Jaton,J.C. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5515-5519.
- Vitetta,E.S. and Uhr,J.W. (1975) *Science (Wash.)*, **189**, 964-969.
- Wahl,G.M., Stern,M. and Stark,G.R. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 3683-3687.
- Word,C.J. and Kuehl,W.M. (1981) *Mol. Immunol.*, **18**, 311-322.
- Yamawaki-Kataoka,Y., Nakai,S., Miyata,T. and Honjo,T. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2623-2627.
- Yuan,D., Uhr,J.W., Knapp,M.R., Slavin,S., Strober,S. and Vitetta,E.S. (1979) in Cooper,M., Mosier,D., Scher,I. and Vitetta,E. (eds.), *B-Lymphocytes in the Immune Response*, Elsevier/North-Holland, NY, pp. 23-31.