Proc. Natl. Acad. Sci. USA
Vol. 78, No. 12, pp. 7684–7688, December 1981
Immunology

Mouse IgA heavy chain gene sequence: Implications for evolution of immunoglobulin hinge exons

(Cα gene organization/RNA splicing/hinge evolution/α-chain sequence)

PHILIP W. TUCKER*, JERRY L. SLIGHTOM†, AND FREDERICK R. BLATTNER‡

*Department of Microbiology, University of Texas Southwestern Medical School, Dallas, Texas 75235; †Department of Chemistry, Southern Illinois University, Carbondale, Illinois 62901; and ‡Department of Genetics, University of Wisconsin, Madison, Wisconsin 53706

Communicated by Joseph L. Goldstein, August 5, 1981

ABSTRACT **The complete nucleotide sequence of the gene and mRNA coding for the constant (C) region of the secreted form of the BALB/c mouse IgA immunoglobulin α heavy (H) chain has been determined. As in other immunoglobulins, the three C region domains of the α protein, Cα1, Cα2, and Cα3 are coded in separate exons. However, the hinge region of Cα is not coded on a separate exon as it is in other hinge-containing immunoglobulins. Instead, the α hinge is coded as a 5′ extension of the Cα2 exon, and we suggest that it may have evolved by duplication leading to incorporation of an acceptor RNA splice site into the coding portion of the Cα2 exon. Extension of this concept could provide an explanation for duplications in the human α1 chain.**

IgA, the characteristic antibody of secretory fluids, occurs as a single class in mouse whereas two distinct nonallelic forms, IgA1 and IgA2, are found in human (1). Amino acid sequences have been published (2–4) for both human and mouse α chains and they show that the constant (C) regions of these proteins consist of three homology domains, Cα1, Cα2, and Cα3, with a proline-rich hinge between Cα1 and Cα2.

The hinge of IgA presents an interesting evolutionary case. The hinge serves an important function by introducing segmental flexibility for the antigen binding arms, thus allowing an optimal fit of the combining site to antigenic determinants at varying distances and angles on antigens (1). However, this exposed region also can be highly sensitive to proteases, which is a liability in the environment of the secretory fluids. Human IgA1 (3) has an extended hinge which has been shown (5) to be sensitive to streptococcal and gonococcal IgA proteases, whereas IgA2, with a shorter hinge, is resistant to these proteases.

The hinge portion of the human IgA1 molecule contains a perfect duplication of eight amino acids (indicated by the arrows above) which, in itself, is terminally redundant (arrows below):

Pro Ser Thr Pro Pro Thr Pro Ser Pro Ser Thr Pro Pro Thr Pro Ser.

The mouse IgA sequence is eight amino acids shorter in the hinge region and contains only a single copy of the duplicated region. Human IgA2 is 13 amino acids shorter than IgA1 and could have arisen by deletion of 13 amino acids from IgA1 or by deletion of 5 amino acids from an unduplicated precursor similar to the mouse genes. These findings at the protein level indicate that the IgA hinge is rapidly evolving, and the instability of this region in evolution can be interpreted as conferring an evolutionary advantage.

Analysis of the mouse Cα gene at the nucleotide level was initiated by Early *et al.* (6) who cloned the mRNA and tumor DNA from the α chain-secreting mouse myeloma MOPC 603. They concluded from results of electron micrographic and restriction mapping techniques that the gene has three domain-sized exons and, within measurement errors, introns were located at presumptive domain boundaries. A recent electron micrographic analysis by Ravetch *et al.* (7) of a cloned human Cα1 gene gave similar results. No separate hinge-encoding exon, such as in the IgD (8) and IgG classes (9–11), was revealed by either of these experiments. Thus, the location of the hinge-encoding segment has remained obscure.

We have now extended the analysis of the Cα region by determining the DNA sequence both from a different IgA myeloma mRNA, J558, and from a cloned gene derived from genomic DNA of tissue uncommitted to the production of immunoglobulin. The results show that the hinge has an unusual genetic structure.

MATERIALS AND METHODS

Plasmid pα(J558)[13] (Fig. 1) was constructed by cloning a double-stranded DNA complementary to α chain mRNA from plasmacytoma J558 into the *Eco*RI site of plasmid vector pMB9 according to the method of Maniatis *et al.* (13). Genomic clones Ch4A 142.4 and Ch4A 140.7 were isolated from our shotgun collection of partial *Eco*RI fragments of BALB/c liver DNA (14) by using the pα(J558)[13] probe.

RESULTS

Mouse Cα mRNA and Genomic DNA Sequences. The complete sequence of the Cα gene is shown in Fig. 2. Data derived from the two cDNA and two genomic clones agreed in all areas where they overlapped except at the 3′ end of p603α1.

The sequence shows that the three gene segments coding for protein domains are separated at the DNA level by two introns 235 and 215 base pairs (bp) long. All RNA splice sites conform to the G·T/A·G (5) rule and show good homology with the consensus donor (5′) and acceptor (3′) sites (16, 17) present in other genes. In contrast to the Cδ (8), Cγ2b (9, 11), and Cγ1 (10) genes, the α hinge structure is not encoded on a separate exon but constitutes the 5′ end of the Cα2 exon.

A typical polyadenylylation [poly(A)] site (A-A-U-A-A-A) occurs only 29 bp beyond the structural gene termination codon in Cα. We found that the 3′ end of the cDNA clone p603a1, which is apparently full-length at the 3′ end (5), ended about 30 nucleotides downstream from the A-A-U-A-A-A as expected. Thus, we conclude that the Cα 3′ untranslated region is short compared with other heavy (H) chain mRNAs (10, 11, 15, 18, 19).

Mouse α Protein Sequence. In comparing the amino acid sequence translated from DNA with that determined for MOPC

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: C, constant; bp, base pair(s); H chain, heavy chain.

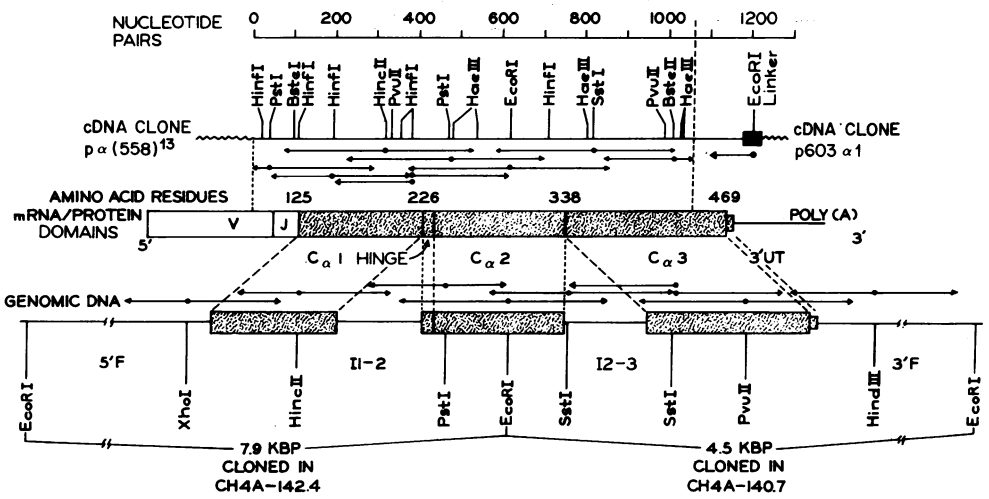

FIG. 1. Map of Cα gene. The restriction map of the cDNA insert of pα(3558)[11] cloned from J558 tumor mRNA is given in the upper portion. The insert begins in the variable (V) segment and ends in the third C region domain at the vertical dashed line. The 3′ end of α cDNA to the right of the dashed line was mapped and sequence-analyzed from cDNA clone p603α1. The bottom portion shows the region of the BALB/c genome cloned independently on two *Eco*RI restriction fragments, clone 142.4 (5′ end) and 140.7 (3′ end) (denoted by boxes). Flanking (F) regions and introns (I1-2 and I2-3) are designated by lines. In the center, the entire α mRNA is illustrated and domain boundaries are delineated. Dashed lines, corresponding positions in DNA and mRNA clones; dots and arrows, sequence analysis strategy—fragments were cleaved and labeled at the dots, and the corresponding arrows give the direction and length of the sequence obtained by the method of Maxam and Gilbert (12). KBP, kilobase pairs.

511 α chain by Robinson and Appella (2) we find eight differences (indicated in Fig. 2) between position 125 (V/C junction) and position 378. Because there is no evidence for IgA subgroups in mouse (1), we suggest that these are of a technical nature rather than true polymorphisms. However, at residue 379, midway through the Cα3 domain, a major discrepancy occurs. Robinson and Appella (2) reported a deletion of 36 amino acids at this point prior to restoration of homology with human Cα3 or other mouse COOH-terminal C_H domains. They point out that this gap would imply an intradomain disulfide loop of only 28 amino acids (including cysteine residues) compared to the normal 64 residues as derived from our germ-line DNA or J558 myeloma mRNA sequences. Because the proposed deletion in MOPC 511 is flanked with arginine residues, it is likely that a tryptic peptide was lost in this region.

Codon Usage. Codon usages for the Cα mRNA are given in Table 1. A possible significance of the increased usage of CCT for proline is discussed below.

Mechanism for a Shortened α Chain in MOPC 47A. Knowing the protein sequence and size of the mRNA, Robinson and Appella (2) predicted a mechanism to account for premature termination of the α chain of the MOPC 47A myeloma IgA. Our DNA sequence is in accord with their proposal that a single base deletion would lead to frameshift and termination of the protein at position 336. Moreover, the gene sequence shows that the mutation would have to be located only 7 bp from one of the points of RNA splicing. Thus, after splicing of the proposed MOPC 47A transcript, the only "codon" in the second exon would be a terminator. This may represent the ultimate in short coding segments.

Table 1. Codon usage for α H chain of mouse immunoglobulin

UUU	Phe(F) 2	UCU	Ser(S) 6	UAU	Tyr(Y) 0	UGU	Cys(C) 3
UUC	Phe(F) 7	UCC	Ser(S) 9	UAC	Tyr(Y) 8	UGC	Cys(C) 11
UUA	Leu(L) 2	UCA	Ser(S) 5	UAA	Ter(.) 0	UGA	Ter(.) 1
UUG	Leu(L) 4	UCG	Ser(S) 2	UAG	Ter(.) 0	UGG	Trp(W) 5
CUU	Leu(L) 1	CCU	Pro(P) 12	CAU	His(H) 3	CGU	Arg(R) 2
CUC	Leu(L) 4	CCC	Pro(P) 9	CAC	His(H) 3	CGC	Arg(R) 1
CUA	Leu(L) 3	CCA	Pro(P) 10	CAA	Gln(Q) 2	CGA	Arg(R) 2
CUG	Leu(L) 21	CCG	Pro(P) 3	CAG	Gln(Q) 9	CGG	Arg(R) 2
AUU	Ilu(I) 4	ACU	Thr(T) 5	AAU	Asn(N) 9	AGU	Ser(S) 4
AUC	Ilu(I) 6	ACC	Thr(T) 16	AAC	Asn(N) 6	AGC	Ser(S) 8
AUA	Ilu(I) 2	ACA	Thr(T) 7	AAA	Lys(K) 6	AGA	Arg(R) 2
AUG	Met(M) 5	ACG	Thr(T) 1	AAG	Lys(K) 7	AGG	Arg(R) 0
GUU	Val(V) 1	GCU	Ala(A) 8	GAU	Asp(D) 7	GGU	Gly(G) 4
GUC	Val(V) 7	GCC	Ala(A) 7	GAC	Asp(D) 4	GGC	Gly(G) 11
GUA	Val(V) 1	GCA	Ala(A) 2	GAA	Glu(E) 5	GGA	Gly(G) 7
GUG	Val(V) 18	GCG	Ala(A) 2	GAG	Glu(E) 18	GGG	Gly(G) 3

Total amino acids, 345; total molecular weight, 36,698. Distribution of bases: A = 22.4%; T = 21.6%; G = 26.2%; C = 29.8%; A+T, 44.1%; C+T, 51.4%.

This mechanism of premature termination is virtually identical to that of M 311, a variant of the MPC 11 cell line (20). This line produces a shorter γ2b chain via a deletion/frameshift mechanism within one base of the identical position relative to the intron between the C_H2 and C_H3 domains (I2-3) in the Cα gene. Comparison of DNA sequences of Cγ2b (9, 10) and Cα shows no particular homology either in the exon or the intron in this area. Clearly these events are not due to aberrant splicing *per se*, but the proximity of the splice site to these mutations may indicate genetic instability near intron–exon junctions.

The α Hinge. Fig. 3 shows alignments of mouse and human amino acid sequences in the neighborhood of the hinge. The hinge is exactly eight amino acids shorter in mouse than in human IgA1 and can be aligned with it perfectly by simply omitting the eight amino acids duplicated in the human protein. Thus, an α chain of the human IgA1 type and length could be derived from mouse by unequal crossing over at the sister chromatid level between residues located eight amino acids apart. Also in Fig. 3 is the alignment of human IgA2 which requires deleting 5 amino acids relative to the mouse sequence or 13 amino acids relative to the human IgA1. Except for the hinge region, the rest of the three proteins can be aligned easily.

A remarkable fact revealed by DNA sequence is that the duplication in human IgA1 begins precisely at the intron–exon junction of IgA of mouse. Assuming this junction is the same in humans, we conclude that the hinge duplication IgA1 is due to a mutational event occurring at an exon border. It was also surprising to discover that the hinge of mouse IgA is encoded as a 5′ extension of the Cα2 exon rather than as a separate exon like other hinge-containing immunoglobulins. Inspection of the α hinge sequence revealed a sequence resembling an internal acceptor RNA splice site located at nucleotide position 629

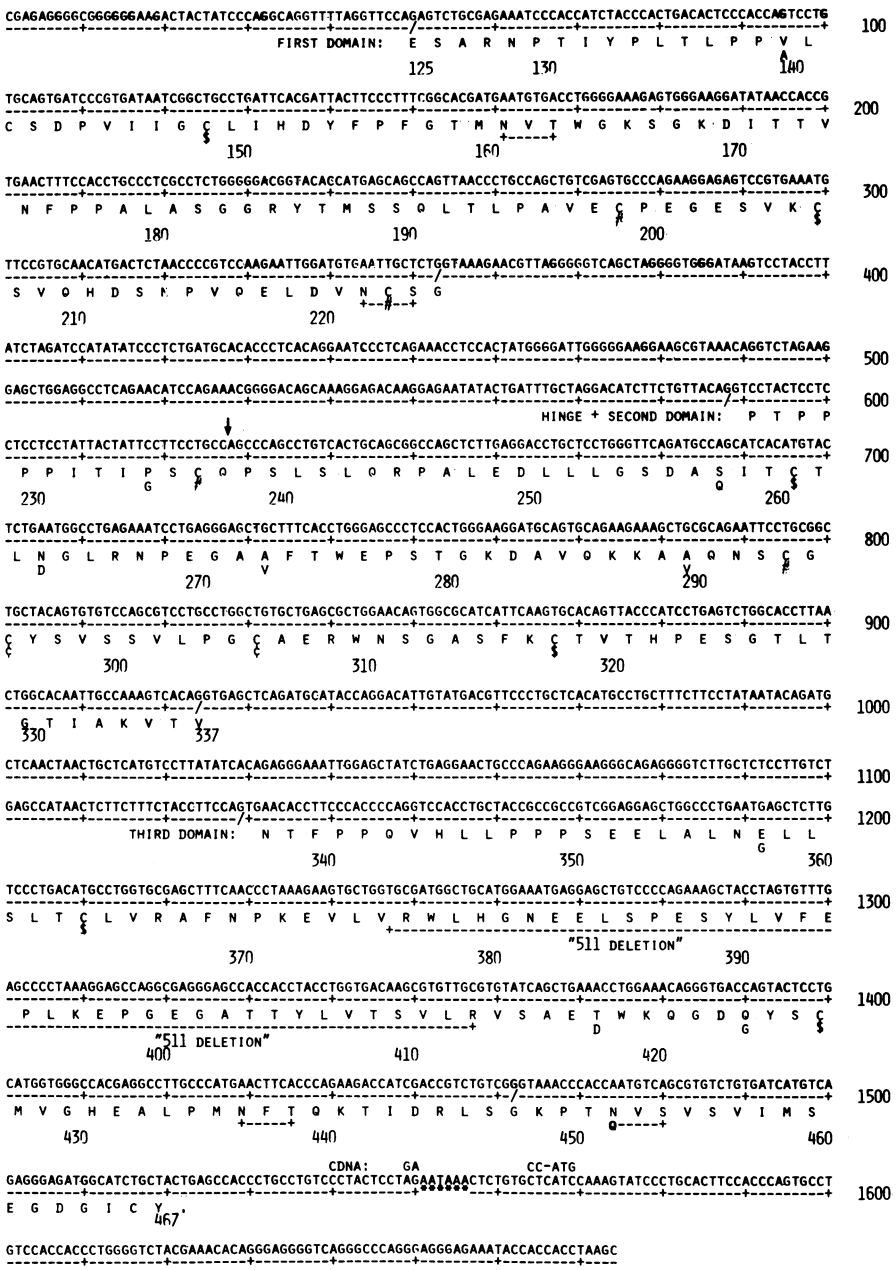

FIG. 2. DNA sequence of the mouse α H chain gene. The DNA sequence of the coding strand is shown. + marks every 10th nucleotide. Exact positions of intron–exon junctions are indicated with slashes and conform to the G·T/A·G rule (15). The slash at nucleotide position 1462 is the proposed splice point for secreted-membrane processing (see text). The characteristic poly(A) recognition sequence is denoted with asterisks. Amino acids are displayed directly below the first base of the corresponding codon and are numbered according to the MOPC 511 α chain protein sequence (2). A key to the single-letter amino acid code is in Table 1. Where different, amino acids from the MOPC 511 sequence (2) appear below. The residues included in the proposed MOPC 511 deletion (2) are underlined. \$, Invariant intradomain cysteine residues; #, other intradomain sulfhydryl-linked cysteine residues; ¢, sites of linkage between H chains; +----+, proposed *N*-linked sugar attachment sites; ↓, end of the hinge (arrow points to the A·G of a nonfunctional RNA splice site) (see text). Nucleotide substitutions indicated at positions 1549, 1550, 1565, and 1569 occur in the cDNA clone p603α1 and may be due to cloning artifacts.

which is 13 amino acids downstream from the beginning of the hinge exon (arrow in Fig. 2). Thus, there are two tandemly arranged potential RNA splice sites in this gene but only the first can actually be used to splice the precursor in order to produce mRNA. (If the second were used, the phasing of translation would be shifted, terminating the peptide.)

We do not understand why the internal splice site-like sequence following the hinge is not used. It is similar in sequence to other acceptor splice sites that are functional—for example, the beginning of the second domain of the mouse Cμ gene (18, 19). Breathnach *et al.* (21) and Seif *et al.* (22) noted that splice sites in mRNA are always chosen so that introns begin with G-T and end with A-G. The consensus of intronic sequences preceding acceptor splice sites generally is pyrimidine-rich with the A-G almost always preceded by purine-C. Both the functional and nonfunctional α hinge splice sites follow these rules. Murray and Holliday (23), Rogers and Wall (16), and Lerner *et al.* (17) proposed that recognition of RNA splice sites involves base homology between sequences flanking donor and acceptor sites and the small nuclear RNA, U1. They suggested that U1 serves as a guide RNA for alignment of the ends of the intron prior to RNA splicing. Surprisingly, the nonfunctional internal α hinge splice site matches U1 RNA better than most known splice sites and better than the functional one in the Cα gene.

FIG. 3. Hinge region sequences of mouse and human Cα genes. The DNA sequence of the mouse Cα gene is shown at the top with most of the I1-2 intron deleted. Single-letter amino acid abbreviations are given in Table 1. The mouse protein sequence and the human (HU) IgA1 and IgA2 protein sequences aligned with it by overlapping the duplication (see text) in IgA1 with itself and by deleting five amino acids (denoted by parentheses) from IgA2 are shown. The duplication in human IgA1 begins precisely at the I1-2 intron border of the mouse sequence.

One reason for this good fit is that the G three nucleotides 5′ to the A-G (position 625 in Fig. 2) matches a C in U1 RNA. However, G does not occur at that position in any other H chain splice site sequenced to date. Conceivably, this G could interfere with acceptor splicing for some reason.

Factors other than sequence alone clearly must participate in determining where to splice. Kinniburgh and Ross (24) and Avvedimento *et al.* (25) have shown that, in globin and α-collagen genes, some splicing occurs to minor sites located within introns. There is no evidence for splicing to minor sites located within exons. A simple hypothesis could be that, all other factors being equal, the first site "in line" is the one that is preferentially used. One possible example is that recently described by Spritz *et al.* (26) for the mutation giving rise to the β^+-thalassemia human globin gene.

The sequence of DNA coding for the α hinge is unusual in itself. The coding strand is extremely rich in pyrimidines and the protein has 6 proline residues in a stretch of 10 amino acids. As shown in Table 2, all of these prolines use the CCU codon although this is a rather unusual codon in mRNA of H chains generally.

The fact that the hinge is coded for by the pyrimidine-rich sequence preceding the nonfunctional acceptor splice site leads to the suggestion that the α hinge may have evolved from an acceptor splice site. To determine whether the unusual codon usage found in the hinge region could be explained by this hypothesis, we examined the translation pattern in three phases of a sample of DNA sequences drawn from intronic regions preceding immunoglobulin acceptor splice sites. As shown in Table 2, proline is one of the most frequent amino acids to be translated in the sample, and CCU is the predominant codon.

DISCUSSION

A specific proposal for the evolution of the α hinge is illustrated in Fig. 4. The key structure, depicted in Fig. 4*C*, consists of two splice sites in a row, the first actually serving a splicing function and the second forming the first part of the exon coding for the hinge sequence and the rest of the domain. Two ways are presented by which this key structure could have evolved from a simple primordial exon with a single splice site. The first way (labeled "fast crossover") is by unequal crossover between two

Table 2. Proline codon usage

	Usage, %				
	CCU	CCC	CCA	CCG	Total codons, no.
Ig H chains less α hinge	20.8	33.1	38.5	7.7	130
α hinge	100	0	0	0	6
Other hinges	22.2	44.4	33.3	0	9
Pre-acceptor (25)	41.4	22.4	36.2	0	58
Post-donor (25)	22.0	39.0	36.6	2.4	41

Ig H chains less α hinge includes the exons for Cμ, Cδ, Cγ1, and Cγ2b and all of the Cα gene except for the hinge. The six α hinge prolines are all coded for by CCU. Other hinges include γ2b, δ, and γ1. Pre-acceptor sequences were analyzed as follows. Stretches of 25 nucleotides on the 5′ side of each acceptor RNA splice site used in these genes were aligned at the consensus sequence CAG which defines the position of the acceptor RNA splice point. A point-by-point plot of pyrimidine content versus sequence position showed a gradient of pyrimidine richness from 95% pyrimidine near the splice point to about 1:1 equal pyrimidine/purine ratio 50 nucleotides away. The 25 nucleotides immediately preceding the splice point averaged about 70% pyrimidine. Proline codon usage was tabulated by translation of these pre-acceptor sequences in all three phases. A similar analysis was done on the sequences following donor splice sites (post-donor sequences).

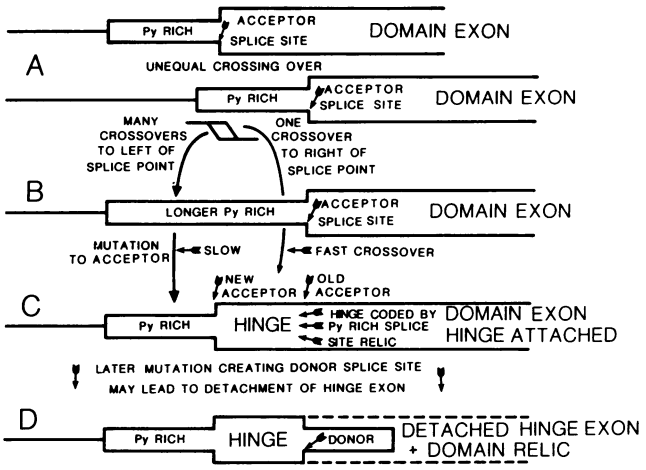

FIG. 4. Proposal for evolution of the hinge. Introns are shown as lines, pyrimidine (Py)-rich sequences 5′ of acceptor splice sites are narrow boxes, and exons are wide boxes. The RNA splice point is denoted by the widening of the narrow boxes to the wide boxes. (*A*) Misaligned crossover which can give rise either to the longer Py-rich region of *B* or to the duplicate splice site of *C*. *B* can mutate to *C* by gradual accumulation of point mutants. (*C*) The culmination of events in which a hinge has been created from a Py-rich splice site sequence. (*D*) The hypothesis that the hinge exon could become detached to form a separate exon. "Relic" refers to an exonic sequence which became intronic or to an intronic sequence which became exonic.

sister chromatids to produce an exact duplication of the splice site in phase. If the first splice site in line were used, a proline-rich hinge segment would appear in the protein as a result of this single genetic event. An additional event producing a triplication in the DNA sequence could incidentally give rise to a duplication of the hinge polypeptide (e.g., human IgA1).

Exact duplicate structures of this sort, however, would be unstable in that a second unequal exchange could easily occur by homologous recombination that would eliminate the hinge or reduplicate to generate additional copies of the hinge. To increase the stability of the structure, the duplication should be rather inexact. The pathway labeled "slow" in Fig. 4 uses intermediate *B* in which a slow process of repeated crossovers has led to evolution of a longer-than-normal pyrimidine-rich area. (Variation of the length of the pyrimidine-rich area can be noted in the immunoglobulin splice sites.) The development of a lengthened pyrimidine-rich region could also result from "slippage" during DNA synthesis as proposed by Streisinger *et al.* (27). From intermediate *B*, one or a few single base changes are all that would be needed to generate the second splice site.

The third aspect of our model (Fig. 4*D*) is a speculation showing how a hinge might subsequently become detached from a domain to form a separate exon, as is the case of γ and δ genes. This could occur by mutation creating a donor splice site within the coding region. The resulting splice pattern would lead to a skipping of the COOH-terminal portion of the domain exon, leaving it to remain as part of the intron. We would expect this domain homology to persist as a relic in the intron for some evolutionary time, but eventually the sequence would drift by mutation beyond recognition or be deleted. During the decay of such a relic, some parts might be reincorporated elsewhere.

There is some evidence that this mechanism may have been operative in other genes. We previously presented evidence for a domain relic in the large intron of the γ2b gene (see figure 5 of ref. 9). On examining that intervening sequence in light of our new hypothesis, we noted that the beginning of the hinge is located precisely at the COOH terminus of the relic in the intron. Thus, the hinge of γ2b could have evolved from the relic's putative 3′ donor splice site by a mechanism analogous

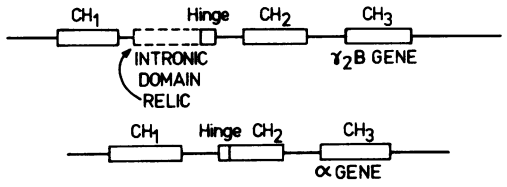

FIG. 5. Hinge placement in γ2b and α genes of mouse. The γ2b gene hinge is coded on a separate exon within a large intron. Within the large intron is a region, denoted by the dashed box, with a strong homology to immunoglobulin domains noted by Tucker *et al.* (9). The α gene hinge is coded on a single exon within the Cα2 domain. The production of a domain relic as a consequence of evolution to a detached hinge exon is discussed in the text.

to that in Fig. 4 but at the other end of the exon. Fig. 5 illustrates the arrangement of the γ2b relic relative to the γ2b hinge.

A second possible example of this type of mechanism is the mouse δ immunoglobulin gene (7). In this case, the second C region domain is missing from the protein so that the hinge is joined directly to the third domain. In the DNA a large intervening sequence is located between the hinge and the third domain but the sequence analysis of this intron is not yet completed so we do not know whether it contains a domain relic. However, the overall structure is consistent with the hypothesis that a splice site mutation leading to domain skipping occurred at some point in the history of this gene.

A third example of a gene whose coding sequence may have evolved from a splice site is the α-collagen gene of chicken. This gene has at least 51 intervening sequences, most of which are in the part coding for the helical portion of the protein consisting of (approximately) a repeating tripeptide with consensus Gly-Pro-Pro (28, 29). These amino acids are encoded, on the average, by GGU, CCU, CCU. As shown in Table 2, these codons are among the most prevalent in our sample of pre-acceptor and post-donor splice sites. It seems very difficult, in fact, to code for Gly-Pro-Pro without engendering sequences similar to splice sites. Our supposition is that this gene may have evolved almost in its entirety from splice sites, although the reverse could be true as well. Because most of the exons are the same length (54 bp) it would be reasonable to assume that duplication of exons as proposed by Ohkubo *et al.* (28) also played a major role, at least in the later stages.

The discovery that eukaryotic genes are punctuated by intervening sequences is a profound and unexpected finding whose meaning is still under discussion. Gilbert (30) has suggested that intervening sequences can speed evolution by allowing novel proteins to be constructed from segments of existing ones. Our results indicate that, in addition, intervening sequences themselves can contribute to the generation of new protein segments as a result of mutations that create, eliminate, and shift RNA splice sites. It remains to be seen whether this evolution of intervening sequence borders into exons and the reverse is a general mechanism, or whether it has operated only in a few special cases. We think it likely that intervening sequence borders will turn out to be preferential sites for mutations, especially those leading to incremental changes in proteins. The rapid evolution of the hinge in the human IgA case is a clear example. In fact, we suggest that the number of intervening sequences within a gene as well as their placement (i.e., in regions critical to protein folding versus between domains) may be factors that control their rate of evolution.

We thank Ken Marcu for mRNA isolation and cDNA construction of pα (J558), Phil Early and Lee Hood for providing p603α1, Fred Mushinski for helpful discussions, Robin Robinson and Hwei-Ling Cheng for technical assistance, and Gerry Ann Cheek for typing the manuscript. This work was supported by Grants AI18016 (P.W.T.) and GM21812 (F.R.B.) from the National Institutes of Health.

1. Putnam, F. W. (1977) in *The Plasma Proteins*, ed. Putnam, F. W. (Academic, New York), 2nd Ed., Vol. 3, pp. 1–153.
2. Robinson, E. A. & Appella, E. I. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 4909–4913.
3. Putnam, F. W., Liu, Y.-S. & Low, T. L. K. (1979) *J. Biol. Chem.* **254**, 2865–2874.
4. Torano, A. & Putnam, F. W. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 966–969.
5. Platt, A. G., Gilbert, J. V., Artenstein, M. S. & Capra, J. D. (1975) *Science* **190**, 1103–1105.
6. Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N. & Hood, L. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4240–4244.
7. Ravetch, J. V., Kirsch, I. R. & Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6734–6738.
8. Tucker, P. W., Liu, C.-P., Mushinski, J. F. & Blattner, F. R. (1980) *Science* **209**, 1353–1360.
9. Tucker, P. W., Marcu, K. B., Newell, N., Richards, J. E. & Blattner, F. R. (1979) *Science* **206**, 1303–1306.
10. Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kawakami, T., Takahashi, N. & Mano, Y. (1979) *Cell* **18**, 559–568.
11. Yamawaki-Kataoka, Y., Kataoka, T., Takahashi, N., Obata, M. & Honjo, T. (1980) *Nature (London)* **283**, 786–789.
12. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
13. Maniatis, T., Kee, S. G., Efstratiadis, A. & Kafatos, F. C. (1976) *Cell* **8**, 163–182.
14. Blattner, F. R., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Richards, J. E., Slightom, J. E., Tucker, P. W. & Smithies, O. (1978) *Science* **202**, 1279–1283.
15. Tucker, P. W., Marcu, R. N., Slighton, J. & Blattner, F. R. (1979) *Science* **206**, 1299–1302.
16. Rogers, J. & Wall, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1877–1879.
17. Lerner, R. M., Boyle, J. A., Mount, M. S., Wolin, L. S. & Steitz, J. A. (1980) *Nature (London)* **283**, 220–224.
18. Kawakami, T., Takahashi, N. & Honjo, T. (1981) *Nucleic Acids Res.* **8**, 3933–3945.
19. Goldberg, G., Vanin, E. & Blattner, F. R. (1981) *Gene* **15**, 33–42.
20. Kenter, A. L. & Birshtein, B. R. (1979) *Science* **206**, 1307–1309.
21. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4853–4857.
22. Seif, I., Khoury, G. & Dhar, R. (1979) *Nucleic Acids Res.* **6**, 3387–3398.
23. Murray, V. & Holliday, R. (1979) *FEBS Lett.* **106**, 5–7.
24. Kinniburgh, A. J. & Ross, J. (1979) *Cell* **17**, 915–921.
25. Avvedimento, V. R., Vogeli, G., Yamada, Y., Maizel, J. V., Pastan, I. & de Crombrugghe, B. (1980) *Cell* **21**, 689–696.
26. Spritz, R. A., Jagadeeswaran, D., Choudary, P. V., Biro, P. A., Eldger, J. T., DeReil, J. K., Manley, J. L., Gefter, M. L., Forget, B. G. & Weissman, S. M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2455–2459.
27. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. & Inouge, M. (1966) *Cold Spring Harbor Symp. Quant. Biol.* **31**, 77–84.
28. Ohkubo, K., Vogeli, G., Mudryj, M., Avvedimento, V. E., Sullivan, M., Pastan, I. & de Crombrugghe, B. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7059–7063.
29. Ohkubo, H., Avvedimento, E., Yamada, Y., Vogeli, G., Sobel, M., Merlino, G., Mudryj, M., Pastan, I. & de Crombrugghe, B. (1981) *ICN/UCLA Symposia on Molecular and Cellular Biology*, Developmental Biology Using Purified Genes, ed. Brown, D. (Academic, New York), Vol. 24, pp. 1–100.
30. Gilbert, W. (1979) *Nature (London)* **271**, 501–503.