19. P. B. Williams, R. T. Kubo, H. M. Grey, *J. Immunol.* **121**, 2435 (1978).
20. S. Tonegawa, personal communication.
21. K. B. Marcu, O. Valbuena, R. P. Perry, *Biochemistry* **17**, 1723 (1978).
22. R. L. Jilka and S. Peska, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5692 (1977).
23. P. H. Hamlyn, G. G. Brownlee, C. C. Cheng, M. J. Gait, C. Milstein, *Cell* **15**, 1067 (1978).
24. D. A. Konkel, S. M. Tilghman, P. Leder, *ibid.*. p. 1125.
25. P. H. Seeberg, J. Shine, J. A. Martial, J. D. Baxter, H. M. Goodman, *Nature (London)* **270**, 486 (1977).
26. A. Efstratiadis, F. C. Kaftos, T. Maniatis, *Cell* **10**, 571 (1977); C. A. Marotta, J. T. Wilson, B. G. Forget, S. M. Weissman, *J. Biol. Chem.* **252**, 5040 (1977).
27. C. H. Faust, I. Heim, J. Moore, *Biochemistry* **18**, 1106 (1977).
28. R. D. Marshall, *Annu. Rev. Biochem.* **41**, 673 (1972).
29. J. Rogers, P. Clarke, W. Salser, *Nucleic Acid Res.* **6**, 3305 (1979).
30. H. Sakano, J. H. Rogers, K. Hüppi, C. Brack, A. Traunecker, R. Maki, R. Wall, S. Tonegawa, *Nature (London)* **277**, 627 (1979).
31. B. K. Birshtein, R. Campbell, M. L. Greenberg, in preparation.
32. F. Sanger *et al.*, *Nature (London)* **265**, 687 (1977).

33. We thank Drs. B. Birshtein, J. Rodgers, P. Clarke, W. Salser, and S. Tonegawa for communicating results to us prior to publication; Dr. W. Fitch for his computer search analyses and insightful discussions; J. Schroeder for implementing additional computer search routines; and Dr. R. B. Perry for making pγ2b(11)^7 available. All experiments were carried out in accordance with the NIH Guidelines on Recombinant DNA Research. Supported by grant GM 21812 (F.R.B.), GM 06526 (P.W.T.), and GM 20069 (Dr. O. Smithies).

* Present address: Department of Biochemistry, University of Mississippi, School of Medicine, Jackson 39216.

† Visiting investigator. Permanent address: Department of Biochemistry, State University of New York, Stony Brook 11794.

13 August 1979; revised 10 October 1979

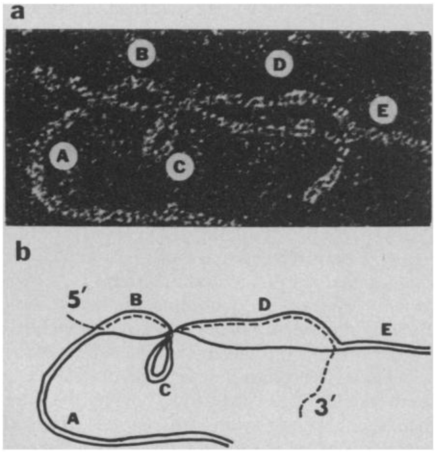
Sequence of the Cloned Gene for the Constant Region of Murine γ2b Immunoglobulin Heavy Chain

Abstract. *The complete nucleotide sequence of the γ2b constant region gene cloned from BALB/c liver DNA is reported. The sequence of approximately 1870 base pairs includes the 5′ flanking, 3′ untranslated, and 3′ flanking regions and three introns. The Cγ2b coding region is divided by these introns into four segments corresponding to the homology domains and hinge region of the protein. The introns separating the hinge from the C_H2 domain and the C_H2 from the C_H3 domain are small (106 and 119 base pairs). A larger intervening sequence of 314 base pairs separates the C_H1 and hinge regions. The stretch of DNA comprising this large intron plus the hinge shows a strong homology with the other C_H domains.*

The protein chains comprising immunoglobulins as discussed (*1*) are organized into domains, and this organization is reflected at the gene level. The variable (V) and constant (C) region domains of light chains are encoded at separate loci, and even in the committed lymphocyte they are separated by introns (*2, 3*). Two studies of DNA cloning of myeloma tumor DNA have appeared showing that the C_H domains of heavy chains are also separated by introns. Early *et al.* (*4*), using electron micrographic and restriction mapping techniques, identified two introns in the α constant region gene from the MOPC 603 myeloma. They concluded that within measurement errors, the introns were located at presumptive domain boundaries. Using mapping techniques plus partial DNA sequence analysis, Sakano *et al.* (*5*) studied a γ1 class constant region gene from the MOPC 21 myeloma in much greater detail. They reached the same conclusion and determined placement of introns between the domains of the Cγ1 gene. They also showed that the γ1 hinge region is encoded by a separate DNA segment.

We have now extended these findings to BALB/c mouse DNA from a tissue not committed to immunoglobulin production (liver). We assume that this DNA is identical in organization to that of the germ line. We have analyzed several constant region genes and report here the complete DNA sequence of one of them, γ2b. Our results show that, for this gene, introns occur precisely between domain and hinge encoding segments. Moreover, since the sequence is complete, we have also proved that no small introns within domains or the 3′ untranslated region have escaped notice. These results in conjunction with the messenger RNA (mRNA) sequence presented (*1*) allow determination of the splice sites at which mRNA presumably is processed to eliminate the introns.

A shotgun collection was constructed by inserting 10- to 20-kilobase-pair (kbp) *Eco* RI fragments from partially digested BALB/c liver DNA into the bacteriophage vector Charon 4A (*6*). The shotgun collection was screened on megaplates (*7*) with a mixture of ^{32}P-labeled plasmid C_H probes (*8*) which included pγ2b(11)^7. Several γ2b positive phages



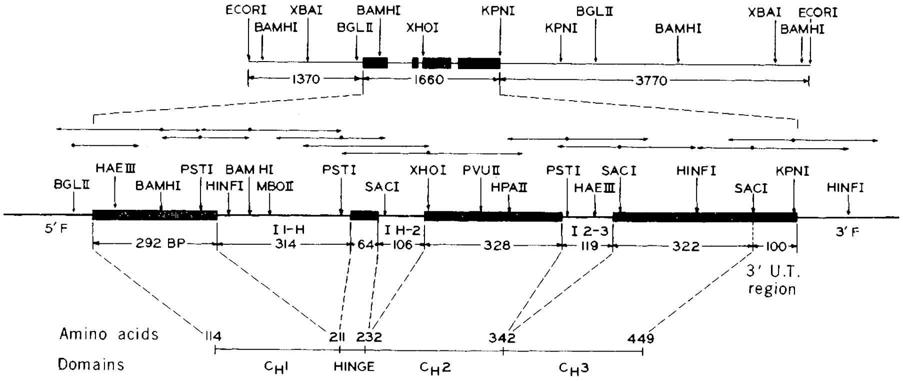


Fig. 1 (left). R-loop structure of Cγ2b genomic clone 144.11.γ2b hybridized to MPC 11 heavy chain mRNA. (a) Electron micrograph; *B* and *D*, correspond to the DNA-mRNA hybrid, *C* corresponds to double-stranded DNA. *A* leads to the right end of the Ch4A vector, *E* to the left end. Lengths given in text were measured in 23 molecules with double-stranded (pCo1E1) and single-stranded (G4) standards. (b) Interpretative drawing; MPC 11 heavy chain mRNA is represented by a dotted line, and single-stranded DNA by solid line. Fig. 2 (right). Endonuclease restriction map of the Cγ2b gene and flanking DNA regions. The cloned 6.8-kbp *Eco* RI fragment in the 5′ to 3′ orientation is shown in the upper panel. The position of the Cγ2b coding region (denoted by boxes) was initially determined by R-looping (legend to Fig. 1) and refined by restriction analysis and Southern hybridization techniques. We have no explanation other than chance for the unusual palindromic arrangement of restriction sites in this DNA. A fivefold magnification of the 1660 nucleotide transcribed Cγ2b gene and the adjoining flanking regions (5′F and 3′F) is shown in the same orientation below. The coding segments are interrupted by three introns: I1-H, IH-2, and I2-3. The dots and arrows above the restriction enzymes denote the sequencing strategy; fragments were cleaved and labeled at the dots, and the corresponding arrows give the direction and length of the sequence obtained. The C_H portion of the γ2b heavy chain protein is illustrated at the bottom of the figure. Dashed lines denote the corresponding positions in the protein and DNA. All nucleotide distances are in base pairs.

0036-8075/79/1214-1303$01.00/0 Copyright © 1979 AAAS

were purified and were completely digested with *Eco* RI and fractionated on agarose gels. Each of the six clones shared a single 6.8-kbp fragment that hybridized to the $p\gamma 2b(11)^7$ probe, and each contained an additional nonhybridizing band that varied in size for three of the candidates.

To confirm that the 6.8-kbp hybridizing fragment is the same size as that in genomic DNA we employed the hybridization technique of Southern (*9*) on a complete *Eco* RI digest of BALB/c mouse liver DNA. As was expected, a single 6.8-kbp fragment hybridized to the Cγ2b probe.

The clone 144.11.γ2b, which contains 6.8- and 7.2-kbp fragments, was selected for detailed analysis. R-loop molecules (*10*) were formed between highly purified γ2b H chain mRNA (*11*) and 144.11γ2b phage DNA; these were examined in the electron microscope (Fig. 1). Two R loops (B and D) were observed with lengths of 264 ± 18 (S.E.M.) and 564 ± 36, nucleotides respectively which roughly sum to the length of 1110 nucleotides determined for the Cγ2b region of the mRNA. These two loops are separated by approximately 249 ± 21 base pairs (bp) (C region) of double-stranded intervening DNA. The appearance of the R loops proves that 144.11.γ2b contains a C_H gene and demonstrates the presence of one large intervening sequence. But the sequence results discussed below demonstrate that there are in fact two additional small (~ 100 bp) introns which did not show up under the conditions used. This emphasizes the need for a complete sequence analysis before the complete intron structure can be known. In the R-loop structure, extensions of RNA corresponding to nonhybridizing 3′ poly(A) (polyadenylate) and 5′ variable region sequences were observed at both ends. The RNA tail closest to the large intron should correspond to the variable region end (right in Fig. 1). Failure of this tail to hybridize shows that no substantial segment of variable region DNA corresponding to the MPC 11 ideotype is present within 3.7 kbp of the C region. This is not surprising since the DNA from a nonantibody-producing tissue such as liver would not be expected to have undergone rearrangement to bring V regions near to a C region gene.

The arrangement of cloned *Eco* RI fragments in 144.11.γ2b was determined by analysis of single and double digests with a variety of restriction endonucleases in combination with Southern hybridization with the $p\gamma 2b(11)^7$ probe. We found by mapping and by examining heteroduplexes between the Ch4A vector and 144.11.γ2b that the 7.2-kbp frag-

```
AATCTGGCACTGCTGCCCAGTCATTGTGAGCACGGTCTAGGCAAGCTAGACAGATCTCTCCAAATATCCGAGGTCACTGATAAGAAAAAAGCTCACACAT
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 100

                                             5' FLANKING

TCTCCTCTCTTGCAGCCAAAACAACACCCCCATCAGTCTATCCACTGGCCCCTGGGTGTGGAGATACAACTGGTTCCTCCGTGACCTCTGGGTGCCTGGT
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 200
              A  K  T  T  P  P  S  V  Y  P  L  A  P  G  C  G  D  T  T  G  S  S  V  T  S  G  C  L  V
                                120               CH1         130                           140

CAAGGGGTACTTCCCTGAGCCAGTGACTGTGACTTGGAACTCTGGATCCCTGTCCAGCAGTGTGCACACCTTCCCAGCTCTCCTGCAGTCTGGACTCTAC
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 300
 K  G  Y  F  P  E  P  V  T  V  T  W  N  S  G  S  L  S  S  S  V  H  T  F  P  A  L  L  Q  S  G  L  Y
                      150                     CH1 160                        170

ACTATGAGCAGCTCAGTGACTGTCCCCTCCAGCACCTGGCCAAGTCAGACCGTCACCTGCAGCGTTGCTCACCCAGCCAGCAGCACCACGGTGGACAAAA
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 400
T  M  S  S  S  V  T  V  P  S  S  T  W  P  S  Q  T  V  T  C  S  V  A  H  P  A  S  S  T  T  V  D  K  K
            180                           190 CH1                 200

AACTTGGTGAGAGGACATTCAGGGGAGGAGGGAATCACCAGAGTTGTAGGCAAAGGTATTAGCCTATCTAACCAGCCAGGCTGGGATCCATCACCAAGGA
----//---+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 500
  L
  210                                           I1-H

GGTGACCTTAGCCTAGGGAAGAGGGAGATACTGTCTCTGCCTCCCTCTTGGGAACATCTAGCTATGACCACCTACACTCAAGGACATGATCCTCTGGGAT
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 600

                                                I1-H

AGGTGTGCTTGTCATTTCCAGGATCATCCTGGACTAAGCCCATACCAGGGACAAACTTTCCTCTCTCTGGTTTGGTGCTTCTCTCCTTCAAAAACCAGTA
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 700

                                                I1-H

ACATCCAGCCTTCTCTCTGCAGAGCCCAGCGGGCCCATTTCAACAATCAACCCCTGTCCTCCATGCAAGGAGTGTCACAAATGCCCAGGTAAGTCACTAC
---------+---------+//-------+---------+---------+---------+---------+---------+----////-+---------+ 800
                      E  P  S  G  P  I  S  T  I  N  P  C  P  P  C  K  E  C  H  K  C  P
        I1-H                              HINGE 220                   230       IH-2

CAGAGCTCGGACTCCCAGGAGAATGGTAAGTGCTGTAAAAATCCCTGTAATGGAGGATAAGCCATGTACAAATCCATTTCCATCTCTCCTCATCAGCTCC
---------+---------+---------+---------+---------+---------+---------+---------+---------+--////---+ 900
                                                                                               A  P
                                                 IH-2

TAACCTCGAGGGTGGACCATCCGTCTTCATCTTCCCTCCAAATATCAAGGATGTACTCATGATCTCCCTGACACCCAAGGTCACGTGTGTGGTGGTGGAT
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 1000
 N  L  E  G  G  P  S  V  F  I  F  P  P  N  I  K  D  V  L  M  I  S  L  T  P  K  V  T  C  V  V  V  D
                240                           250 CH2                  260

GTGAGCGAGGATGACCCAGACGTCCAGATCAGCTGGTTTGTGAACAACGTGGAAGTACACACAGCTCAGACACAAACCCATAGAGAGGATTACAACAGTA
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 1100
V  S  E  D  D  P  D  V  Q  I  S  W  F  V  N  N  V  E  V  H  T  A  Q  T  Q  T  H  R  E  D  Y  N  S  T
      270                           280            CH2            290                           300

CTATCCGGGTGGTCAGCACCCTCCCCATCCAGCACCAGGACTGGATGAGTGGCAAGGAGTTCAAATGCAAGGTGAACAACAAAGACCTCCCATCACCCAT
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 1200
  I  R  V  V  S  T  L  P  I  Q  H  Q  D  W  M  S  G  K  E  F  K  C  K  V  N  N  K  D  L  P  S  P  I
                          310                  CH2   320                           330

CGAGAGAACCATCTCAAAAATTAAAGGTGGGACCTGCAGGACAACTGCATGGGGGCTGGGATGGGCATAAGAATAAATGTCTATGTGGACAGCCTTCCAC
---------+---------+---////--+---------+---------+---------+---------+---------+---------+---------+ 1300
 E  R  T  I  S  K  I  K
                340                              I2-3

TTCGGCCATGACCTCTATGTATGTTTCTAACCCCACAGGGCTAGTCAGAGCTCCACAAGTATACACTTTGCCGCCACCAGCAGAGCAGTTGTCCAGGAAA
---------+---------+---------+-----////+---------+---------+---------+---------+---------+---------+ 1400
                                     G  L  V  R  A  P  Q  V  Y  T  L  P  P  P  A  E  Q  L  S  R  K
                       I2-3                        CH3   350                           360

GATGTCAGTCTCACTTGCCTGGTCGTGGGCTTCAACCCTGGAGACATCAGTGTGGAGTGGACCAGCAATGGGCATACAGAGGAGAACTACAAGGACACCGC
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 1500
D  V  S  L  T  C  L  V  V  G  F  N  P  G  D  I  S  V  E  W  T  S  N  G  H  T  E  E  N  Y  K  D  T  A
                  370                           380 CH3                 390

CACCAGTTCTTGACTCTGACGGTTCTTACTTCATATATAGCAAGCTCAATATGAAAACAAGCAAGTGGGAGAAAACAGATTCCTTCTCATGCAACGTGAG
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 1600
  P  V  L  D  S  D  G  S  Y  F  I  Y  S  K  L  N  M  K  T  S  K  W  E  K  T  D  S  F  S  C  N  V  R
        400                           410        CH3            420                            430

ACACGAGGGTCTGAAAAATTACTACCTGAAGAAGACCATCTCCCGGTCTCCGGGTAAATGAGCTCAGCACCCACAAAGCTCTCAGGTCCTAAGAGACACT
---------+---------+---------+---------+---------+---------+---------+---------+---------+---------+ 1700
 H  E  G  L  K  N  Y  Y  L  K  K  T  I  S  R  S  P  G  K
                            440                   CH3  449'           3'UT

GGCACCCATATCCATGCATCCCTTGTATAAATAAAGCATCCAGCAAAGCTGGTACCATGTAAAACTGTCTTGGTTGTTTCCAAGGTATAGAGCATAGCTC
---------+---------+---------+---------+---------+---------+////-----+---------+---------+---------+ 1800

                         3'UT                                                3' FLANKING

ACGGGCTGATAGGTCTGGCCAGGGCCGGAGAACAGCCTTGTCTATAGGAAGAGAATGAGGTTTCTGCCTGCAT
---------+---------+---------+---------+---------+---------+---------+---

                         3' FLANKING
```

Fig. 3. Complete sequence of Cγ2b gene. The DNA sequence of the coding strand is shown on top with translated amino acids below. Nucleotide numbering is indicated at the right of a line; protein numbering is listed below the sequence. The symbol + appears below every tenth nucleotide. 5-Methylcytosine residues resulting from methylation by *Escherichia coli* occur at positions 646 and 790 and on the opposite strand at these positions less 2. These are in addition to those found in the cDNA plasmid [figure 1 in (*1*)]. Slashes indicate possible positions for splicing precursor RNA that would result in the correct mRNA sequence. Single letter abbreviations for amino acid residues are: 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; and Y, Tyr.

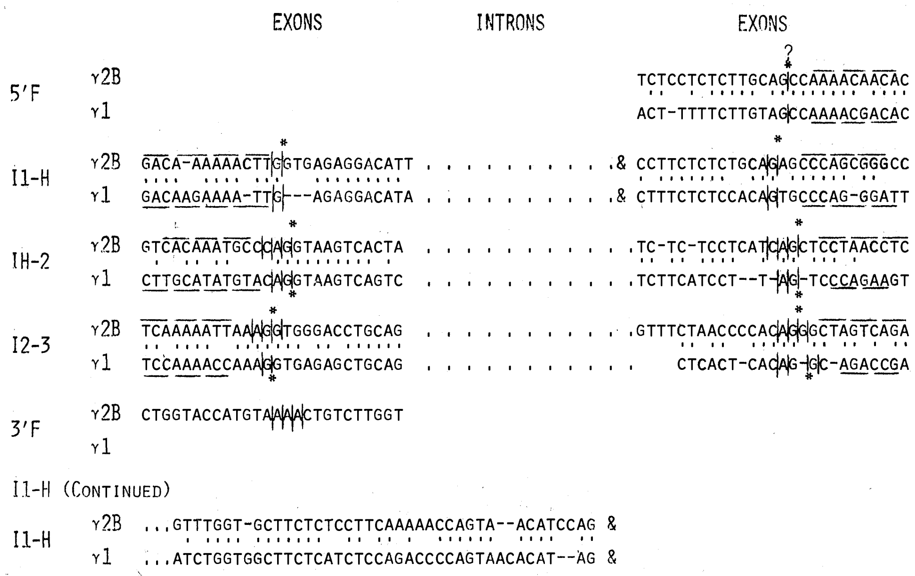

EXONS
INTRONS
EXONS
?
5′F
γ2B
TCTCCTCTCTTGCAGCCAAAACAACAC
γ1
ACT-TTTTCTTGTAGCCAAAACGACAC
I1-H
γ2B
GACA-AAAAACTTGGTGAGAGGACATT
&
CCTTCTCTCTGCAGAGCCCAGCGGGCC
γ1
GACAAGAAAA-TTG---AGAGGACATA
&
CTTTCTCTCCACAGTGCCCAG-GGATT
IH-2
γ2B
GTCACAAATGCCCAGGTAAGTCACTA
TC-TC-TCCTCATCAGCTCCTAACCTC
γ1
CTTGCATATGTACAGGTAAGTCAGTC
TCTTCATCCT--T-AG-TCCCAGAAGT
I2-3
γ2B
TCAAAAATTAAAGGTGGGACCTGCAG
GTTTCTAACCCCACAGGGCTAGTCAGA
γ1
TCCAAAACCAAAGGTGAGAGCTGCAG
CTCACT-CACAG-GC-AGACCGA
3′F
γ2B
CTGGTACCATGTAAAACTGTCTTGGT
γ1
I1-H (CONTINUED)
I1-H
γ2B
...GTTTGGT-GCTTCTCTCCTTCAAAAACCAGTA--ACATCCAG &
γ1
...ATCTGGTGGCTTCTCATCTCCAGACCCCAGTAACACAT--AG &

Fig. 4. Intervening sequence junctions in mouse C_H genes. Sequences spanning the splice sites of the Cγ2b and Cγ1 genes were aligned for maximal homology by introducing gaps. Alternative exon-intron boundaries are represented by the vertical lines. The asterisks denote splices that follow the GT-AG rule without altering the protein sequence. The assignment of the 3′ boundary of the 5′ flanking region (the V-C junction) (*) is tentative. Coding triplets are shown by horizontal lines. The extensive homology seen at the 3′ end of the I1-H intron is continued at the bottom of the figure with breakpoint shown by ampersand (&).

ment is not mouse DNA but is one of the internal *Eco* RI fragments from the Charon 4A vector (*6*). The 6.8-kbp mouse fragment is located to the right of the recloned vector fragment in the clone. According to our nomenclature (*12*), the orientation of the recloned vector fragment is n and that of the mouse fragment is u. The orientation of the mouse fragment derived from restriction mapping agrees with the orientation derived from the asymmetric placement of the R loop in the clone assuming the large intervening sequence is at the 5′ end of the gene.

A detailed restriction map (Fig. 2) was constructed for the transcribed and flanking regions of the C_H gene cloned in 144.11.γ2b. Comparison of the lengths of a number of restriction fragments with those mapped for the Cγ2b complementary DNA (cDNA) indicated the approximate locations and sizes of three introns.

The complete sequence of the Cγ2b gene, obtained by the method of Maxam and Gilbert (*13*), is presented in Fig. 3. The sequencing runs used to derive it are indicated by arrows in Fig. 2. The introns and intron-exon junctions were sequenced on both strands except for 20 bp near the center of the largest intron. Most of these regions were sequenced more than once on each strand. We estimate that the overall accuracy of these sequences is better than 90 percent. Of the coding regions, the left half of C_H1 was the only part not included in the cDNA plasmid sequence (*1*), but for confirmation, the entire coding region was sequenced, most of it on both strands (Fig. 2). No differences were found between mRNA and gene.

The nucleotide sequence shows that for γ2b the domains and hinge are separated at the DNA level by introns. These can now be delineated as follows: C_H1, amino acid residues 114 to 210; hinge, residues 211 to 232; C_H2, residues 233 to 342; and C_H3, residues 343 to 449. There are no additional introns within the DNA coding segments, 3′ untranslated region, or at the junction between C_H3 and the 3′ untranslated region.

Since the numbering or lettering of immunoglobulin introns and exons from the left or variable region end is problematical, especially since the germ line clone does not include a variable region, we have adopted a nomenclature in which exons are designated by coding function and introns are named by context; that is, C_H1 refers to the DNA coding for the C_H1 domain, H refers to the DNA that codes for the hinge domain and I1-H refers to the intron between them. Regions surrounding the transcribed Cγ2b DNA are denoted as 5′ flanking (5′F) and 3′ flanking (3′F) sequence.

The functions of intervening sequences, if any, are unknown. They are present in the primary RNA transcript, then later excised, presumably by site-specific ribonuclease attack, and the adjacent coding regions are respliced to give mature, contiguous mRNA's. Nucleotide sequences at the intron-exon junctions are obvious candidates for this processing specificity.

We present a comparison (Fig. 4) between the Cγ2b intron border sequences and the corresponding Cγ1 sequences determined by Sakano *et al.* (*5*). For illustration the sequences have been aligned for maximal homology. The sequences at the ends of introns are strikingly similar, with pyrimidine-rich bases at the 3′ end and purine-rich areas at the 5′ end. Most of the homology regions do not extend very far into the introns (to the extent that this can be seen from the data available for Cγ1). However, the 3′ ends of the I1-H introns are extremely similar over 50 bp, sharing 80 percent homology with three gaps (as shown at the bottom of Fig. 4). The 5′ flanking region of the γ2b gene is also very homologous with that of Cγ1 and is pyrimidine-rich. We anticipate that this sequence will participate in the splicing of V region DNA sequences located to the 5′ side of the gene.

Breathnach *et al.* (*14*) have proposed that splice points may always be chosen so that the intron begins (5′ side) with GT- and ends with AG-. This rule appears to hold for Cγ2b as well as for all cases of globin (*15*, *16*), light chain immunoglobin (*3*, *17*), SV40 (*18*), and ovalbumin (*14*, *19*) for which both DNA and either RNA sequences or definitive protein sequences are available (22 cases). Sakano *et al.* (*5*) have noted a violation, however, for the I1-H intron of Cγ1.

As mentioned by Tucker *et al.* (*1*), there is no protein or direct mRNA sequence data at the V-C junction of γ2b. Thus, an unequivocal assignment of the beginning of the C_H1 coding region cannot be made. But the point at which γ2b protein sequence diverges from the corresponding amino acid sequence of γ1 is at residue 114, and there is a pyrimidine-rich stretch followed by an AG dinucleotide that could be used for splicing these according to the GT-AG rule. There are no other AG dinucleotides within 20 nucleotides in either direction. We have therefore used this as the tentative assignment for the position of the joint.

The 3′ flanking region does not have any introns in the portion coding for the 3′ untranslated RNA. In the germ line DNA, a TGT followed by four A residues occurs at the point at which poly(A) presumably is added to mRNA (*1*). The only other gene sequenced beyond the 3′ noncoding region is mouse β globin (*15*), which has TGC followed by two A's. There is no homology between the area to the right of this putative poly(A) addition site and the left ends of the other introns. This may suggest that splicing enzymes are not involved in defining the right end of the message.

We have used a computer program, designed by Dr. Walter Fitch, to investigate homologies within the γ2b C_H gene. This program can compare two DNA sequences in all possible registers to identify alignments in which significant homology exists. In rare cases, these can then be assembled into larger homology units with the introduction of gaps. We were quite surprised to find that, when the sequence of the I1-H intron plus the hinge was compared with the rest of the gene, there was a high degree of homology with each of the full-sized C_H domains, and in the case of C_H1 the homology extended to include some of the 5′ flanking DNA. In fact the degree of homology was comparable to that of the domains with each other. No other alignments with this degree of match were found.

Figure 5 presents an illustration of the alignment of C_H1 versus C_H3 compared with the alignment of the 5′F plus C_H1 compared to I1-H plus hinge sequences. The homologies observed in all comparisons are as follows. For 5′F plus C_H1 versus intron plus hinge, 38 percent homology can be achieved by inserting five gaps that include 33 nucleotides; for C_H2, the homology is 37 percent with five gaps and 37 nucleotides; and for C_H3, the homology is 39 percent with three gaps and 31 nucleotides. These values compare with 36, 42, and 40 percent for homology, respectively, between C_H domains themselves (*1*).

The evolution of the hinge has been a puzzle for some time. The more primitive chains such as μ and ϵ have four full-sized domains instead of three plus a hinge. Putnam (*20*) proposed that the hinge evolved by deletion from the second domain of a primitive μ-like chain. We are led to propose instead that the evolution proceeded by a shift of a splice site followed by mutational divergence. Mutations of the full-sized primordial domain may have led to the elimination of the acceptor component of an RNA processing site at the beginning of the second C_H domain and its replacement by an acceptor site with the domain itself. This would convert the left end of the gene for the second domain into a "pseudogene" within the intron. This mechanism would account for the intron homology as an evolutionary remnant of the first part of this primordial domain.

```
                          A)     5'F + CH1   VS    I1-H + HINGE

22   5'F                                                                                  5'F:CH1
CATTGTGAGCACGGTCTAGGCAAGCTAGACAGATCTCTCCAAATATCCGAGGTCACTGATAAGAAAAAAGCTCACACATTCTCCTCTCTTGCAGCCAAAA
   TG GAG AC  TC  GG A G   GA   A C           C  AGGT    G   A   AA     CA  C      C  TC    AG
TGGTGAGAGGACATTCAGGGGAGGAGGGAATCACCAGAGTTGTAGGCAAAGGTATTAGCCTATCTAACCAGCCAGGCTGGGATCCATCACCAAGGAGGTG
405   I1-H                                                                                 I1-H

CAACACCCCAT          CAGTCTATCCACTGGCCCCTGGGTGTGGAGATACAACTGGTTCCTCCGTGACCTCTGGGTGC      CTGGTCAAGGG
       CC A             T     C CTG C CC    TG G A AT  A CT    CC CC   AC    GG          CTGG   AGG
ACCTTAGCCTAGGGAAGAGGGAGATACTGTCTCTGCCTCCCTCTTGGGAACATCTAGCTATGACCACCTACACTCAAGGACATGATCCTCTGGGATAGGT

GTACTTCCCTGAGCCAGTGACTGTGACTTGGAACTCTGGATCCCTGTCCAGCAGTGTGCACACCTTCCCAGCTCTCCTGCAGTCTGGACT   CTACACT
GT CTT  C    CCAG   C              CTGGA       C   C   G  CA AC TTCC   CTCT  T   GT           T  A
GTGCTTGTCATTTCCAGGATCATC           CTGGACTAAGCCCATACCAGGGACAAACTTTCCTCTCTCTGGTTTGGTGCTTCTCTCCTTCAAAA

                            CH1                                                        I1-H  393
ATGAGCAGCTCAGTGACTGTCC    CCTCCAGCACCTGGCCAAGTCAGACCGTCACCTGCAGCGTTGCTCACCCAGCCAGCAGCACCACGGTG
A  AG A C     G CT   C    C      CA C GGCC A T   AC  TCA C  C G   T C   C   G   G   CA
ACCAGTAACATCCAGCCTTCTCTCTGCAGAGCCCAGCGGGCCCATTTCAACAATCAACCCCTGTCCTCCATGCAAGGAGTGTCACAAATGCCCA
                        I1-H:HINGE                                                     HINGE 885

 139 MATCHES   5 GAPS   394 NUCLEOTIDES   366 COMPARISONS   38.0 PERCENT MATCH

                          B               CH1  VS  CH3

115   CH1
GCCAAAACAACACCCCCATCAGTCTATCCACTGGCCCCTGGGTGTGGAGATACAACTGGTTCCTCCGTGACCTCTGGGTGCCTGGTCAAGGGGTACTTCC
G    A   A A C CCA  AGT TA  C  TG C CC       G   A     C  G       GT A        TGCCTGGTC  GGG T C  CC
GGGCTAGTCAGAGCTCCACAAGTATACACTTTGCCGCCACCAGCAGAGCAGTTGTCCAGGAAAGATGTCAGTCTCACTTGCCTGGTCGTGGGCTTCAACC
1335  CH3

CTGAGCCAGTGACTGTGACTTGGAACTCTGGA      TCCCTGTCCAGCAGTGTGCACACCTTCCCAGCTCTCCTGCAGTCTGGACTCTACACTATGAG
CTG      T A TGTG   TGGA C             C   G   A C     G ACACC   CCAG TCT          GG    TAC   AT
CTGGAGACATCAGTGTGGAGTGGACCAGCAATGGGCATACAGAGGAGAACTACAAGGACACCGCACCAGTTCTTGACTCTGACGGTTCTTACTTCATATA

                                                                                         CH1     405
CAGCTCAGTGACTGTCCCCTCCAGCACCTGGCCAAGTCAG   ACCGTCACCTGCAGCGTTGCTCACCCAGCCAGCAGCACCACGGTGGACAAAAACTT
 AGC    T A T T     C AGCA  TGG   A         CC TC C TGCA CGT    CAC   G     A  A           AA AA
TAGCAAGCTCAATATGAAAACAAGCAAGTGGGAGAAAACAGATTCCTTCTCATGCAACGTGAGACACGAGGGTCTGAAAAATTACTACCTGAAGAAGACC
                                                                                         CH3    1656

 118 MATCHES   2 GAPS   300 NUCLEOTIDES   293 COMPARISONS   40.3 PERCENT MATCH
```

Fig. 5. Demonstration of homology between intron and exon. (A) The DNA sequence of the portion of the γ2b constant region gene from nucleotide residues 22 to 393 including 5′ flanking sequence and C_H1 domain is presented on the top line. Beneath it the sequence from residues 405 to 885, including the Il-H intron and hinge, has been aligned for maximum homology by introducing five small gaps. Where a match occurs the matching base is printed between. Percentage match is computed by treating gaps as single mismatches (see preceding report, Table 1). (B) For comparison, a similar alignment of the two most closely related exons, C_H1 and C_H3, is presented. C_H1 is shown on top and C_H3 on bottom.

PHILIP W. TUCKER*
KENNETH B. MARCU†
NANETTE NEWELL, JULIA RICHARDS
FREDERICK R. BLATTNER
Genetics Laboratory, University of Wisconsin, Madison 53206

References and Notes

1. P. W. Tucker, K. B. Marcu, J. L. Slightom, F. R. Blattner, *Science* **206**, 1299 (1979).
2. C. Brack, M. Hirama, S. Lenhard-Schuller, S. Tonegawa, *Cell* **15**, 1 (1978); J. G. Seidman, A. Leder, M. Nau, B. Norman, P. Leder, *Science* **202**, 11 (1978).
3. S. Tonegawa, A. M. Maxam, R. Tizard, O. Bernard, W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3518 (1978).
4. P. W. Early, M. W. Davis, D. B. Kaback, N. Davidson, L. Hood, *ibid.* **76**, 857 (1979).
5. H. Sakano, J. H. Rodgers, K. Huppi, C. Brack, A. Traunecker, R. Maki, R. Wall, S. Tonegawa, *Nature (London)* **277**, 627 (1979).
6. F. R. Blattner *et al.*, *Science* **196**, 161 (1977).
7. F. R. Blattner *et al.*, *ibid.* **202**, 1279 (1978).
8. P. Tucker, unpublished observations.
9. E. M. Southern, *J. Mol. Biol.* **98**, 503 (1975).
10. M. Thomas, R. L. White, R. W. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 2294 (1976).
11. K. B. Marcu, O. Valbuena, R. P. Perry, *Biochemistry* **17**, 1723 (1978).
12. O. Smithies, A. E. Blechl, K. Denniston-Thompson, N. Newell, J. E. Richards, J. L. Slightom, P. W. Tucker, F. R. Blattner, *Science* **202**, 1284 (1978).
13. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
14. R. Breathnach, C. Benoist, K. O'Hare, F. Gannon, P. Chambon, *ibid.* **75**, 4853 (1978).
15. D. A. Konkel, S. M. Tilghman, P. Leder, *Cell* **15**, 1125 (1978).
16. O. Smithies, J. Slightom, P. Tucker, F. R. Blattner, in *11th Annual Miami Winter Symposium*, in press; R. M. Lawn, E. F. Fritsch, R. C. Parker, G. Blake, T. Maniatis, *Cell* **15**, 1157 (1978); B. G. Forget, C. Covallesco, J. K. deRiel, R. A. Spritz, P. V. Choudary, J. T. Wilson, L. B. Wilson, V. B. Reddy, S. M. Weissman, *Proceedings of the 1979 ICN-UCLA Symposia*, in press; J. van den Berg *et al.*, *Nature (London)* **276**, 37 (1978).
17. O. Bernard, N. Hozumi, S. Tonegawa, *Cell* **15**, 1133 (1978).
18. V. P. Reddy, B. Thimmappaya, R. Dhar, K. N. Subramanian, B. S. Zain, J. Pan, P. K. Ghosh, M. L. Celman, S. M. Weissman, *Science* **200**, 494 (1978).
19. F. Gannon *et al.*, *Nature (London)* **278**, 428 (1979).
20. F. W. Putnam, in *The Plasma Proteins*, F. W. Putnam, Ed. (Academic Press, New York, 1977), vol. 3, pp. 1-53.
21. We thank Dr. Walter Fitch for computer search analyses and Dr. J. L. Slightom for help in shotgun screening. All experiments were done in accordance with NIH Guidelines on Recombinant DNA research. Supported by NIH grant GM 21812 (F.R.B.), GM 06526 (P.W.T.), and GM 20069 (O. Smithies). This is paper number 2382 from the Laboratory of Genetics, University of Wisconsin, Madison.

* Present address: Department of Biochemistry, University of Mississippi School of Medicine, Jackson 39216

† Permanent address: Department of Biochemistry, State University of New York, Stony Brook 11794.

13 August 1979; revised 10 October 1979