Sequences of Mouse Immunoglobulin Light Chain Genes before and after Somatic Changes

Ora Bernard, Nobumichi Hozumi and
Susumu Tonegawa
Basel Institute for Immunology
Grenzacherstrasse 487
Postfach, CH-4005 Basel 5
Switzerland

Summary

We have determined the nucleotide sequences of the germ line gene as well as a corresponding somatically mutated and rearranged gene coding for a mouse immunoglobulin λ_I type light chain. These sequencing studies were carried out on three Eco RI-DNA fragments which had been cloned from BALB/c mouse embryos or a λ_I chain-secreting myeloma, H2020. The embryonic DNA clone Ig 99λ contains two protein-encoding segments, one for the majority of the hydrophobic leader (L) and the other for the rest of the leader and the variable (V) region of the λ_{II} chain (Cohn et al., 1974); these segments are separated by a 93 base pair (bp) intervening sequence (I-small). The coding of the V region ends with His at residue 97. The second embryonic DNA clone Ig 25λ includes a 39 bp DNA segment (J) coding for the rest of the conventionally defined V region (that is, up to residue 110), and also contains the sequence coding for the constant (C) region approximately 1250 untranslated bp (I-large) away from the J sequence. The J sequence is directly linked with the V-coding sequence in the myeloma DNA clone, Ig 303λ, which has the various DNA segments arranged in the following order: 5′ untranslated region, L, I-small, V linked with J, I-large, C, 3′ untranslated sequence. The Ig 303λ V DNA sequence codes for the V region synthesized by the H2020 myeloma and is different from the Ig 99λ V DNA sequence by only two bases. No silent base change was observed between the two DNA clones for the entire sequence spanning the 5′ untranslated regions and the V-coding segments. These results confirm the previously drawn conclusion that an active complete λ_I gene arises by somatic recombination that takes place at the ends of the V-coding DNA segment and the J sequence. No sequence homology was observed at or near the sites of the recombination.

Introduction

An immunoglobulin polypeptide chain is composed of two parts: a highly variable amino terminal region (V region) and a relatively constant carboxy terminal region (C region). The DNA segments coding for the two regions (V DNA and C DNA) are distant in embryonic cells and are rearranged by recombination that takes place in the precursors of lymphocytes (Hozumi and Tonegawa, 1976; Tonegawa et al., 1976; Brack et al., 1978; Lenhard-Schuller et al., 1978).

We have previously cloned, from both BALB/c mouse embryos and myeloma cells, various DNA fragments containing part or all of the λ or κ chain genes. Characterization of these DNA clones by R loop mapping, heteroduplex analysis and restriction enzyme mapping provided direct evidence for somatic recombination (Brack and Tonegawa, 1977; Tonegawa et al., 1977; Brack et al., 1978; Lenhard-Schuller et al., 1978). The same studies revealed a surprising feature of gene organization: somatic recombination does not generate a DNA stretch containing contiguously arranged V and C DNA segments. In a complete λ_I chain gene expressed in myelomas, the $V_{\lambda I}$ and $C_{\lambda I}$ DNA segments are 1.2 kb apart, while in an active complete κ chain gene, the V_κ and C_κ DNA segments are 2.9 kb apart (Brack and Tonegawa, 1978; Lenhard-Schuller et al., 1978). In embryonic DNA, a short DNA segment called J (~50 bp) lies 1.2 kb upstream (that is, toward the 5′ end of the coding strand) from the $C_{\lambda I}$ DNA. The J sequence is homologous to a sequence near the V-C junction of a λ_I mRNA (Brack et al., 1978). The recombination seems to take place at or near the 3′ end of the $V_{\lambda I}$ DNA and at or near the J sequence, generating a complete λ_I gene composed of various DNA segments in the following order: 5′ V, J, 1.2 kb intervening sequence, C 3′ (Brack et al., 1978). The precise locations of the recombination sites, however, can only be revealed by determination of the DNA sequences.

Mouse λ chains are classified into the two subtypes λ_I and λ_{II}. Amino acid sequences of the two subtype chains are different at ~10–12 positions in the V region and at about 25 positions in the C region (Weigert et al., 1970; Dugan et al., 1973). Total embryonic DNA gives two Eco RI fragments of 4.8 and 3.6 kb, which carry sequences homologous to the V part of a λ_I hybridization probe (Tonegawa et al., 1977; Brack et al., 1978). We have previously cloned the 4.8 kb fragment and identified its V_λ DNA as that of a λ_{II} type by DNA sequencing. These studies revealed a 93 bp intervening sequence separating a DNA segment coding for the majority of the hydrophobic leader residues from the V DNA segment (Tonegawa et al., 1978). We would like to know whether an equivalent intervening sequence is present in the embryonic $V_{\lambda I}$ DNA, and if so, whether this intervening sequence persists in the expressed complete λ_I gene in myeloma cells.

The BALB/c mouse genome carries no more

than a few copies of $V_{\lambda I}$ DNA segments per haploid genome (Tonegawa, 1976; Honjo et al., 1976). Since seven different (but highly homologous) $V_{\lambda I}$ amino acid sequences are known and more are very likely to exist (Weigert et al., 1970), we believe that a large proportion of the V region diversity is generated by a somatic process such as somatic mutation. Based on the frequent recurrence among myeloma λ_I chains, Cohn et al. (1974) postulated that the V region represented by MOPC 104E is the product of a putative germ line V DNA segment. Sequencing of an embryonic $V_{\lambda I}$ clone will test this proposal directly. Furthermore, comparison of the embryonic V DNA sequence with that of a somatically derived myeloma V DNA may provide us with some information relevant to the mechanism of the somatic change.

To study the problems raised above and to analyze the fine structure of an immunoglobulin gene and the flanking regions, we have determined the nucleotide sequences of the relevant regions of three λ_I DNA clones: an embryonic $V_{\lambda I}$ (Ig 99λ), an embryonic $C_{\lambda I}$ (Ig 25λ) and a myeloma $V_{\lambda I}$ plus $C_{\lambda I}$ clone (Ig 303λ).

Results

The basic characteristics of the three λ_I DNA clones are summarized in Table 1. It should be noted that the clone Ig 303λ was isolated from DNA prepared from H2020 myeloma, and contains an Eco RI-DNA fragment that is absent from embryonic DNA (Brack and Tonegawa, 1977).

Restriction Enzyme Mapping of the Cloned DNA

Detailed restriction enzyme maps were constructed to sequence the cloned DNA. Our previous R loop mapping by electron microscopy showed the locations of the regions homologous to H2020 λ_I mRNA (Brack et al., 1978). These locations are indicated in Figure 1. We have also previously determined the regions of homology among the cloned mouse

DNA fragments by heteroduplex mapping (Brack et al., 1978). Thus about one third of Ig 303λ on one end is homologous to one end of Ig 99λ, and about two thirds of Ig 303λ on the other end is homologous to one end of Ig 25λ. Ig 99λ and Ig 25λ exhibited no detectable homology. The results of these analyses were used as guides in map construction.

The mapping was performed in two steps. First, isolated mouse Eco RI DNA fragments were digested with restriction enzymes such as Bam HI, Xba I and Kpn I, which cleave DNA relatively infrequently (Figure 1A). Digestion products of Ig 303λ and one of the two embryonic clones were fractionated by electrophoresis on a single slab gel to identify DNA fragments of a common size. Digestion products were also analyzed by the Southern gel blotting technique (Southern, 1975) to identify the DNA fragments containing sequences homologous to H2020 λ_I mRNA. DNA fragments containing the ends of the Eco RI fragment were also identified by comparing the electrophoretic patterns generated by single (enzyme X) and double (enzyme X and Eco RI) digestion of the whole λgt_{WES}-Ig phage DNA. These procedures allowed unique ordering of most of the DNA fragments. When ambiguity existed, fragment order was determined or confirmed by double enzyme digestion (enzyme X and enzyme Y). Figure 1A summarizes the results obtained with this approach.

In the second step, the DNA fragments containing sequences homologous to H2020 λ_I mRNA were labeled at 5' ends with γ-^{32}P–ATP using polynucleotide kinase and cleaved with restriction enzymes to yield two fragments labeled at only one of the 5' ends. These fragments were isolated and partially digested with enzymes such as Alu I, Hae III, Hinf I and others which cut DNA relatively frequently. The digestion products were fractionated by polyacrylamide gel electrophoresis and enzyme cleavage sites were deduced (Smith and Birnstiel, 1976). In most of the cases, cleavage

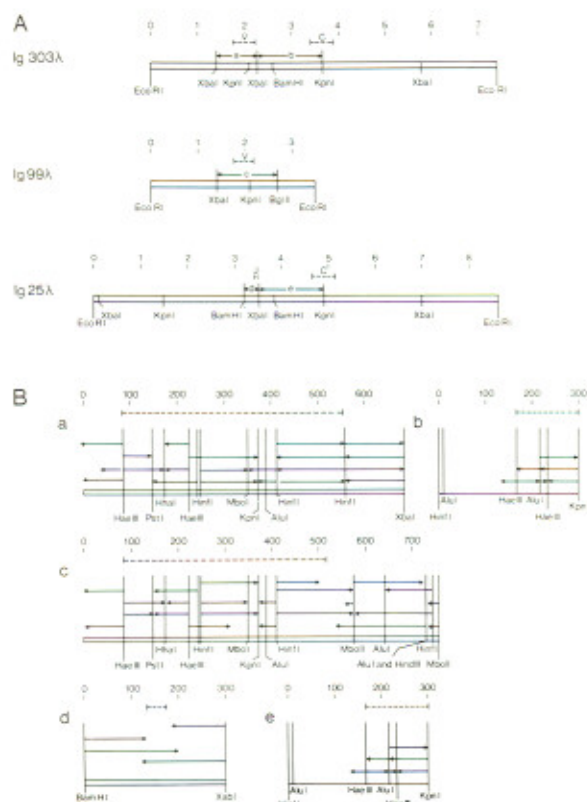| Clone	DNA Source	Size of Cloned DNA (kb)	Type of λ DNA Contained	References
Ig 99λ	Embryo 3.2–3.8 kb	3.5	$V_{\lambda I}$	Brack et al. (1978)
Ig 303λ	Myeloma H 2020 7.0–7.6 kb	7.4	$V_{\lambda I}$ + $C_{\lambda I}$	Brack and Tonegawa (1977)
Ig 25λ	Embryo 8.2–8.8 kb	8.6	$C_{\lambda I}$	Brack et al. (1978)
Ig 13λ	Embryo 4.5–5.2 kb	4.8	$V_{\lambda II}$	Tonegawa et al. (1977) Tonegawa et al. (1978)

Table 1. Characteristics of λ Light Chain DNA Clones

Figure 1. Restriction Enzyme Maps and Sequencing Strategies of the Three λ₁ DNA Clones

(A) Cleavage site maps of the cloned Eco RI-DNA fragments are shown together with the regions homologous to mRNA. The latter were determined by R loop mapping and Southern gel blotting analysis. (a) The fragment used to sequence the V DNA and the flanking regions of Ig 303λ. (b) The fragment used to sequence the 5' end region of the C DNA of Ig 303λ. (c) The fragment used to sequence the V DNA and the flanking regions of Ig 99λ. (d) The fragment used to sequence the J DNA and the flanking regions of Ig 25λ. (e) The fragment containing the sequence of the 5' end of the C DNA of Ig 25λ. Fragment sizes are given in kb.

(B) Detailed restriction enzyme maps of all or part of the DNA fragments (a–e) indicated in (A) are shown together with the regions sequenced. The dotted lines indicate regions homologous to mRNA. Fragment sizes are given in bp.

sites determined by the partial digestion were confirmed by complete digestion. Figure 1B summarizes the results obtained with this approach.

DNA Sequencing Strategy

For DNA sequencing of Ig 303λ, whole phage DNA was digested with Xba I and fractionated by 5% polyacrylamide gel electrophoresis (Figure 1A). Xba I digestion was most useful for our purposes because it does not digest λ phage arms and cleaves the inserted DNA at three positions. The 0.9 kb fragment obtained in this way contains sequences homologous to the V region of H2020 λ₁ mRNA (Figure 1A, fragment a). Fragment a was

further digested with the restriction enzymes indicated in Figure 1Ba, and then labeled at the 5' ends with γ-³²P-ATP using polynucleotide kinase. The labeled DNA was either cut with another restriction enzyme to give two fragments labeled only at one end, or it was subjected to strand separation. For sequencing the C DNA segment of Ig 303λ, whole λgt_WES-Ig 303λ DNA was digested with Xba I. The largest fragment (3.6 kb) was then digested with Kpn I to give fragment b (Figure 1A). This fragment was further digested with the restriction enzymes indicated in Figure 1Bb, and sequenced as described below.

Whole λgt_WES-Ig 99λ DNA was digested with Bgl II and Xba I, and the 1.4 kb DNA fragment containing the V DNA segment was isolated (Figure 1A, fragment c). This fragment was digested with the restriction enzymes indicated in Figure 1Bc and then labeled at the 5' ends for DNA sequencing.

Whole λgt_WES-Ig 25λ was digested with Xba I. A 3.5 kb DNA fragment which contains the J sequence and extends from 0.05–3.5 kb from the left Eco RI site was isolated by electrophoresis. This fragment was further digested with Bam HI and the 5' ends were labeled. A 300 base long fragment containing the J sequence was isolated (Figure 1A, fragment d), and the strands separated by electrophoresis on a 5% polyacrylamide gel were sequenced (Figure 1Bd). The C DNA segment of Ig 25λ is present in the 3.6 kb fragment of the Xba I digest. This fragment was further digested with Kpn I to give fragment e (Figure 1A). The sequencing strategy is outlined in Figure 1Be and was the same as for the corresponding region of Ig 303λ.

Using 40 cm long gels and the three polyacrylamide concentrations (12, 20 and 24%), we were able to read sequences up to about 200 nucleotides. Typical examples of the sequencing gels are shown in Figure 2. As shown in Figure 1B, gel sequences presented in this report were determined at least twice and, in the case of doubtful assignments, up to four times.

Nucleotide Sequences

DNA sequences around the V segments of Ig 303λ and Ig 99λ, around the J segment of Ig 25λ, and around the C segments of Ig 303λ and Ig 25λ are shown in Figure 3. The determined DNA sequences and the amino acid sequences derived from them are presented.

Ig 99λ

The DNA sequence of Ig 99λ corresponds to a Vλ₁ region and more specifically to the V region of the λ_II chains (Cohn et al., 1974). The coding beings with the first amino acid of the hydrophobic leader sequences, Met at position −19; continues to Ser at position −5 (or Gly at position −4; see below); is

interrupted by a 93 bp intervening sequence; resumes with Gly at position –4 (or Ala at position –3; see below); and ends abruptly with His at position 97. The leader amino acid sequence predicted by the determined DNA sequence is exactly that determined by protein chemistry (Burstein and Schechter, 1977b).

In the V region, the amino acid sequence predicted by the DNA sequence fits that determined by protein chemistry (Cesari and Weigert, 1973). Furthermore, the DNA sequence clarifies several ambiguous assignments made by protein chemistry due to difficulties in distinguishing between Glu and Gln or Asp and Asn. Thus according to the DNA sequences of both the Ig 99λ and Ig 303λ, Glu is found at positions 7, 16, 40, 83 and 85, while Gln is at positions 6, 39 and 81. Likewise, Asp is at positions 71 and 84, while Asn is at positions 33, 36, 54, 55 and 96. The above residue assignments are in partial agreement with those reported by Burstein and Schechter (1977a), while they are at variance with the sequence reported by Appella (1971) at positions 7, 40 and 71.

The exact boundaries of the 93 bp intervening sequence are ambiguous because the quadruplet CAGG followed by T or G is repeated at the opposite boundary regions. Thus coding may be interrupted either within the Ser codon TCA (position –5) or within or at the end of the Gly codon GGT (position –4). Note that the triplets GGT and GGG at the opposite ends of the intervening sequence are synonymous and both code for Gly.

Ig 303λ

As expected, the DNA sequence of the V coding segment corresponds to the V region of the H2020 λ chain (Weigert et al., 1970). As in Ig 99λ, coding beings with Met at position –19 and is interrupted by a 93 bp intervening sequence; unlike the case of Ig 99λ, the coding continues beyond His at position 97 until Gly at position 110, where it ends again abruptly. Almost all the DNA sequences of the two clones Ig 99λ and Ig 303λ are identical in the entire stretch extending from the 5′ end untranslated region, through the leader, through the intervening region and through the V coding region, to the His codon at position 97. Only two exceptions to this almost perfect sequence identity were found, and these were in the codons for residues 25 and 32.

The determined triplet codons of Ig 99λ are AGT for Ser at both positions, while those of Ig 303λ are ACT for Thr at position 25 and GGT for Gly at position 32 (Figure 2). In contrast, in the more than 160 bp that follow the His codon CAT at position 97, Ig 99λ and Ig 303λ have no sequence resemblance.

Approximately 1.2 kb away from the V DNA segment (Brack and Tonegawa, 1977), coding of the C region starts with Gly at position 110 and continues without interruption until Gly at position 156. The nucleotide sequence beyond this position is not dealt with in this paper. The precise boundaries of the 1.2 kb intervening sequence are again ambiguous because of the directly repeated Gly codons, GGT at the end of the expanded myeloma V DNA segment and GGC at the beginning of the C DNA segment. Due to the degeneracy of Leu codons, it is even possible that the coding interruption occurs between the second and third nucleotides of the codon CTA for Leu at position 109.

Ig 25λ

As indicated above, the Ig 303λ V DNA segment codes for amino acids until Gly at position 110, while coding by the Ig 99λ V DNA segment ends with His at position 97. The 39 base long sequence necessary for encoding the extra 13 residues at the end of the Ig 303 V region was discovered in the J segment of Ig 25λ. Coding in the J segment starts with Trp at position 98 and ends with Gly at position 110. The region preceding the J segment has no sequence resemblance with the corresponding regions of either Ig 303λ or Ig 99λ. In contrast, the sequence in the region following the J segment is identical with that of Ig 303λ, at least for the first 130 bp shown in Figure 3. The only exception was found at 48 bp away from the first nucleotide immediately following the triplet GGT for Gly at position 110 (indicated by an asterisk in Figure 3). We observed A in Ig 303λ and G in Ig 25λ. We do not know the exact reason for this single base difference. We speculate that it may be due to residual allele heterogeneity among individual BALB/c mice. Alternatively, the single base change might have occurred either during propagation of H2020 myeloma cells or during propagation of the DNA clones in E. coli.

Partial sequences of the Ig 25λ C DNA segment

Figure 2. Autoradiogram of Sequencing Gels Illustrating Minor Sequence Differences in the V Coding Part of Ig 99λ and Ig 303λ

(a) Fragment c (Figure 1A) isolated from Ig 99λ was also digested with Hinf I, labeled at its 5′ ends and separated on an 8% acrylamide gel. The largest fragment was digested with Mbo I and the larger of the two resulting fragments was sequenced. The sequence indicated in the figure codes for amino acids 19–34 (Figure 3).

(b) Fragment a (Figure 1A) isolated from Ig 303λ was digested with Hinf I and labeled at its 5′ ends. After fractionation on an 8% acrylamide gel, the second largest DNA fragment was isolated and digested into two fragments with Mbo I. The larger of the two fragments was sequenced from the Hinf I cleavage end (Figure 1Ba). The nucleotide sequence that appears in the autoradiogram and is indicated by letter labels codes for amino acids 15–33 (Figure 3).

In both (a) and (b), the arrows indicate nucleotides that are different in Ig 99λ and Ig 303λ.

```
                                                                                                                    -1*
                                                                                                          MetAlaTrp
Ig  99λ  GCCCAGCCCAGCCCATACTAAGAGTTATATTATGTCTGTCTCACAGCCTGCTGCTGACCAATATTGAAAAGAATAGACCTGGTTTGTGAATTATGGCCTGG
         CGGGTCGGGTCGGGTATGATTCTCAATATAATACAGACAGAGTGTCGGACGACGACTGGTTATAACTTTTCTTATCTGGACCAAACACTTAATACCGGACC

Ig 303λ  GCCCAGCCCAGCCCATACTAAGAGTTATATTATGTCTGTCTCACAGCCTGCTGCTGACCAATATTGAAAAGAATAGACCTGGTTTGTGAATTATGGCCTGG
         CGGGTCGGGTCGGGTATGATTCTCAATATAATACAGACAGAGTGTCGGACGACGACTGGTTATAACTTTTCTTATCTGGACCAAACACTTAATACCGGACC
```

```
                                                                              -1
         IleSerLeuIleLeuSerLeuLeuAlaLeuSerSerGly
Ig  99λ  ATTTCACTTATACTCTCTCTCCTGGCTCTCAGCTCAGGTCAGCAGCCTTTCTACACTGCAGTGGGTATGCAACAATGCGCATCTTGTCTCTGATTTGCT
         TAAAGTGAATATGAGAGAGAGGACCGAGAGTCGAGTCCAGTCGTCGGAAAGATGTGACGTCACCCATACGTTGTTACGCGTAGAACAGAGACTAAACGA

Ig 303λ  ATTTCACTTATACTCTCTCTCCTGGCTCTCAGCTCAGGTCAGCAGCCTTTCTACACTGCAGTGGGTATGCAACAATGCGCATCTTGTCTCTGATTTGCT
         TAAAGTGAATATGAGAGAGAGGACCGAGAGTCGAGTCCAGTCGTCGGAAAGATGTGACGTCACCCATACGTTGTTACGCGTAGAACAGAGACTAAACGA
```

```
                                -*                 1
                                  GlyAlaIleSerGlnAlaValValThrGlnGluSerAlaLeuThrThrSerProGlyGluThrValThr                     10
Ig  99λ  ACTGATGACTGGATTTCTCATCTGTTTGCAGGGGCCATTTCCCAGGCTGTTGTGACTCAGGAATCTGCACTCACCACATCACCTGGTGAAACAGTCACA
         TGACTACTGACCTAAAGAGTAGACAAACGTCCCCGGTAAAGGGTCCGACAACACTGAGTCGTGAGTGGTGTAGTGGACCACTTTGTCAGTGT

Ig 303λ  ACTGATGACTGGATTTCTCATCTGTTTGCAGGGGCCATTTCCCAGGCTGTTGTGACTCAGGAATCTGCACTCACCACATCACCTGGTGAAACAGTCACA
         TGACTACTGACCTAAAGAGTAGACAAACGTCCCCGGTAAAGGGTCCGACAACACTGAGTCCTTAGACGTGAGTGGTGTAGTGGACCACTTTGTCAGTGT
```

```
         20                       30                                      40                                    50
         LeuThrCysArgSerSerThrGlyAlaValThrThrSerAsnTyrAlaAsnTrpValGlnGluLysProAspHisLeuPheThrGlyLeuIleGlyGly
Ig  99λ  CTCACTTGTCGCTCAAGTACTGGGGCTGTTACAACTAGTAACTATGCCAACTGGGTCCAAGAAAAACCAGATCATTTATTCACTGGTCTAATAGGTGGT
         GAGTGAACAGCGAGTTCATGACCCCGACAATGTTGATCATTGATACGGTTGACCCAGGTTCTTTTTGGTCTAGTAAATAAGTGACCAGATTATCCACCA
                      Thr              HV 1              Gly
Ig 303λ  CTCACTTGTCGCTCAACTACTGGGGCTGTTACAACTGGTAACTATGCCAACTGGGTCCAAGAAAAACCAGATCATTTATTCACTGGTCTAATAGGTGGT
         GAGTGAACAGCGAGTTGATGACCCCGACAATGTTGACCATTGATACGGTTGACCCAGGTTCTTTTTGGTCTAGTAAATAAGTGACCAGATTATCCACCA
                                          HV 1
```

```
                            60                      70                               80
         ThrAsnAsnArgAlaProGlyValProAlaArgPheSerGlySerLeuIleGlyAspLysAlaAlaLeuThrIleThrGlyAlaGlnThrGluAspGlu
Ig  99λ  ACCAACAACCGAGCTCCAGGTGTTCCTGCCAGATTCTCAGGCTCCCTGATTGGAGACAAGGCTGCCCTCACCATCACAGGGGCACAGACTGAGGATGAG
         TGGTTGTTGGCTCGAGGTCCACAAGGACGGTCTAAGAGTCCGAGGGACTAACCTCTGTTCCGACGGGAGTGGTAGTGTCCCCGTGTCTGACTCCTACTC
                                  HV 2

Ig  25λ      GATCCTGGGAAGAAGGATCTTTCAGTGATGTCACCACCTTCCAAGAATTACCAGGAGCTGCATACATCACAGATGCAACTTGAGAATAAAATG
             CTAGGACCCTTCTTCCTAGAAAGTCACTACAGTGGTGGAAGGTTCTTAATGGTCCTCGACGTATGTAGTGTCTACGTTGAACTCTTATTTTAC

Ig 303λ  ACCAACAACCGAGCTCCAGGTGTTCCTGCCAGATTCTCAGGCTCCCTGATTGGAGACAAGGCTGCCCTCACCATCACAGGGGCACAGACTGAGGATGAG
         TGGTTGTTGGCTCGAGGTCCACAAGGACGGTCTAAGAGTCCGAGGGACTAACCTCTGTTCCGACGGGAGTGGTAGTGTCCCCGTGTCTGACTCCTACTC
                                  HV 2
```

```
         90                       97
         AlaIleTyrPheCysAlaLeuTrpTyrSerAsnHis
Ig  99λ  GCAATATATTTCTGTGCTCTATGGTACAGCAACCATTTCCACAATGACATGTGTAGATGGGGAAGTAGATCAAGAACACTCTGGTACAGTCTCATAACT
         CGTTATATAAAGACACGAGATACCATGTCGTTGGTAAAGGTGTTACTGTACACATCTACCCCTTCATCTAGTTCTTGTGAGACCATGTCAGAGTATTGA
                              HV 3                    98                    110
                                                       TrpValPheGlyGlyGlyThrLysLeuThrValLeuGly
Ig  25λ  CATGCAAGGTTTTTGCATGAGTCTATATCACAGTGCTGGGTGTTCGGTGGAGGAACCAAACTGACTGTCCTAGGTGAGTCACTCCTTCCTCCTTTGTTA
         GTACGTTCCAAAAACGTACTCAGATATAGTGTCACGACCCACAAGCCACCTCCTTGGTTTGACTGACAGGATCCACTCAGTGAGGAAGGAGGAAACAAT

Ig 303λ  GCAATATATTTCTGTGCTCTATGGTACAGCAACCATTGGGTGTTCGGTGGAGGAACCAAACTGACTGTCCTAGGTGAGTCACTCCTTCCTCCTTTGTTA
         CGTTATATAAAGACACGAGATACCATGTCGTTGGTAACCCACAAGCCACCTCCTTGGTTTGACTGACAGGATCCACTCAGTGAGGAAGGAGGAAACAAT
                              HV 3
```

```
Ig  99λ  ACCACTTTCTTAACAGGTGGCTACATCTCCCTAGTCTGTTCTCTTTTACTATAGAGAAATTTATAAAAGCTGTTGTCTCAATCAATAAAAAGTTTTATT
         TGGTAGAAGAATTGTCCACCGATGTAGAGGGATCAGACAAGAGAAAATGTATATCTCTTTAAATATTTTCGACAACAGAGTTAGTTATTTTTCAAAATAA

Ig  25λ  TTGTTCTCTCCAAGACTTGAGGTGCTTTTTGTTGTATACTTTCCCTTTCTGTATTCTGCTTCATACCTATACTTCACACTAGGTAAAGAATTTCTTTCT
         AACAAGAGAGGTTCTGAACTCCACGAAAAACAACATATGAAAGGGAAAGACATAAGACGAAGTATGGATATGAAGTGTGATCCATTTCTTAAAGAAAGA

Ig 303λ  TTGTTCTCTCCAAGACTTGAGGTGCTTTTTGTTGTATACTTTCCCTTTCTGTATTCTGCTTCATACCTATACTTCACACTAGGTAAAGAATTTCTTTCT
         AACAAGAGAGGTTCTGAACTCCAAGAAAAACAACATATGAAAGGGAAAGACATAAGACCAAGTATGGATATGAAGTGTGATCCATTTCTTAAAGAAAGA
```

```
Ig  99λ  CAACAAATTGTATAATTATGCCTTGATGACAAGCTTTGTTTACCAACTTGGCACAACATAGAATCATTGAGAAGAGAACC
         GTTGTTTAACATATTAATACGGAACTACTGTTCGAAACAAATGGTTGAACCGTGTTGTATCTTAGTAACTCTTCTCTTGG
```

```
                               116                            120
                                GlyGlnProLysSerSerProSerValThrLeuPheProProSerSerGluGluLeuThr
Ig  25λ  TCTCTGA..//..TTTTGACCTTCTCTTACTTCATCCTGCGGCCAGCCCAAGTCTTCGCCATCAGTCACCCTGTTTCCACCTTCCTCTGAAGAGCTCACT
         AGAGACT..//..AAAACTGGAAGAGAATGAAGTAGGACGCCGGTCGGGTTCAGAAGCGGTAGTCAGTGGGACAAAGGTGGAAGGAGACTTCTCGAGTGA

Ig 303λ  TCTCTGA..//..TTTTGACCTTCTCTTACTTCATCCTGCGGCCAGCCCAAGTCTTCGCCATCAGTCACCCTGTTTCCACCTTCCTCTGAAGAGCTCACT
         AGAGACT..//..AAAACTGGAAGAGAATGAAGTAGGACGCCGGTCGGGTTCAGAAGCGGTAGTCAGTGGGACAAAGGTGGAAGGAGACTTCTCGAGTGA
```

```
         130                                     140                                  150
         GluAsnLysAlaThrLeuValCysThrIleThrAspPheTyrProGlyValValThrValAspTrpLysValAspGly
Ig  25λ  GAGAACAAGGCCACACTCGTGTGAACGATCACTGATTTCTACCCAGGTGTGGTGACAGTGGACTGGAAGGTAGATGGA
         CTCTTGTTCCGGTGTGACCACACTTGCTAGTGACTAAAGATGGGTCCACACCACTGTCACCTGACCTTCCATCTACCT

Ig 303λ  GAGAACAAGGCCACACTCGTGTGAACGATCACTGATTTCTACCCAGGTGTGGTGACAGTGGACTGGAAGGTAGATGGA
         CTCTTGTTCCGGTGTGAGCACACTTGCTAGTGACTAAAGATGGGTCCACACCACTGTCACCTGACCTTCCATCTACCT
```

and the region immediately preceding the C DNA segment were also determined and are shown in Figure 3. The sequences are identical with those of the corresponding regions of Ig 303λ.

Discussion

Precise Position and Characteristics of the Recombination Sites

Our previous heteroduplex and R loop mapping studies predicted that recombination takes place near the 3' end of the embryonic V DNA segment and near the J segment that lies 1.2 kb upstream (that is, toward the 5' side) from the embryonic C DNA segment (Brack et al., 1978). The present DNA sequencing studies confirmed this prediction and determined the recombination site to within two phosphodiester bonds. Thus in embryonic DNA, recombination takes place at the 3' side of the His codon CAT (position 97) on Ig 99λ and at the 5' side of the Trp codon TGG (position 98) on Ig 25λ. Alternatively, it takes place at the 3' side of T following the His codon and at the 5' side of G comprising the second nucleotide of the Trp codon. In either case, the recombination event attaches the embryonic V DNA segment directly to the J segment, generating an expanded V DNA segment observed in the myeloma clone.

A striking feature of the sequences around the recombination sites of the two embryonic clones is the lack of even a short homology (except for a single base homology of T). In fact, had there been a longer homology, it would not have been possible to define the recombination sites with the accuracy described above. The lack of sequence homology in the V-J recombination sites contrasts with the presence of 15 bp core homology at the *att* sites for the site-specific recombination of phage λ (Landy and Ross, 1977). In this case, although the biological significance of the common core remains unclear, its existence might be taken as confirmation of the suggestion that the core homology serves as the recognition signal for the

site-specific recombination (Shulman and Gottesman, 1973). Lack of a core homology at the V-J recombination sites rules out any mechanism in whcih a sequence homology has a key role in the site determination. In this respect, the V-J recombination bears some resemblance to the recombinations accompanying mobilization of the transposable elements in procaryotic systems (for reviews, see Cohen, 1976; Starlinger and Saedler, 1976; Bukhari, Shapiro and Adhya, 1977). There is increasing evidence suggesting that at least a major part of the site recognition in the procaryotic *rec*-independent recombinations is not carried out by sequence homology, as might have been predicted by the Campbell model (Campbell, 1962; L. Johnsrud, M. P. Calos and J. H. Miller, personal communication).

In procaryotic systems, the degree of site specificity on the recipient genomes varies widely from one case to another, with phage Mμ and phage λ exhibiting the least and the most restricted specificity. In the case of recombination involving the immunoglobulin DNA segments, one would expect that the specificity of the site recognition is high so that efficient generation of a properly rearranged, complete immunoglobulin gene is assured in lymphocyte precursors. Our recent studies suggest, however, that the V or the J-C segment occasionally interacts with DNA segments elsewhere in the genome, generating apparently "aberrant" recombinants (Lenhard-Schuller et al., 1978). This may be an indication that the site recognition system used by V-J recombination has specificity less stringent than its involvement in the generation of an active complete gene might imply.

Does V-J recombination always take place in as narrowly defined regions as the present studies suggest? The apparent precision of the V-J recombination sites should be reconsidered in the light of the unique nature of immunoglobulin polypeptide chains. As is discussed below, we believe that the entire population of the BALB/c V_λ DNA segments arises by somatic alteration (perhaps by

Figure 3. Nucleotide Sequences of the Three λ₁ DNA Clones

Determined nucleotide sequences of Ig 99λ, Ig 303λ and Ig 25λ, aligned in phase with the protein-encoding sequences, are shown together with the amino acid sequences predicted from the triplet codons. Nucleotide sequences are underlined where a pair of clones exhibit an identical sequence. Where sequence differences occur, they are indicated by an asterisk[*]. The hypervariable regions HV1, HV2 and HV3 are underlined with thick lines. Dotted lines with a double slash indicate part of the 1250 bp intervening sequences for which no nucleotide sequence data are reported in this paper. In 99λ, coding begins with the first amino acid of the hydrophobic leader (Met at position −19), is interrupted at Ser (−5)/Gly(−4) by a 93 bp intervening sequence and resumes with Gly(−4)/Ala(−3). The V region coded in Ig 99λ is that of the λ₀ chain, but coding ends prematurely with His at position 97. A 39 bp DNA segment (J sequence) coding for the rest of the conventionally defined V region — that is, a peptide starting with Trp (98) and ending with Gly (110) — was found in Ig 25λ. The J sequence is followed by an approximately 1250 bp intervening sequence that is followed in turn by the C region coding sequence beginning with Gly (110)/Gln(111). Except for the two nucleotides in HV1, the nucleotide sequence of Ig 303λ is identical to that of Ig 99λ in the entire DNA segment extending from the 5' untranslated region to the His (97) codon, after which it switches to the sequence of Ig 25λ except in the case of the one base pair indicated. The V region encoded in Ig 303λ is that of the H2020 λ chain. Short sequence repeats at the ends of the intervening sequences as well as the CTG repeats toward the 5' ends are indicated by underlining between the two complementary DNA strands (see the text).

mutations) from a single germ line $V_{\lambda I}$ DNA segment. As indicated in Figure 3, the V-J recombination site is immediately adjacent to one of the three regions (designated HV3) where the putative somatic alterations are frequent (Cohn et al., 1974). It thus remains possible that at least part of the somatic sequence change in HV3 is brought about by somatic recombinations that take place somewhere within the HV3 coding region and in a region to the left of the J segment. Such recombinations will generate a series of HV3-coding DNA composed partly of germ line V DNA sequence and partly of the sequence preceding the J segment. Although the available amino acid sequences of the $V_{\lambda I}$ regions do not support this notion, it nevertheless remains a possibility in view of the paucity of the data (Weigert et al., 1970).

Somatic Generation of V_λ DNA Diversity

Weigert et al. (1970) have analyzed the amino acid sequences of λ_1 chains isolated from eighteen independently isolated BALB/c mouse plasmacytomas. Twelve of the eighteen λ_1 chains have the same amino acid sequence, called λ_0 (Cohn et al., 1974). The amino acid sequences of the other six λ_1 chains differ from the λ_0 type by one to three residues, and all the differences occur in the three regions termed hypervariable. Previous hybridization studies indicated that the entire population of $V_{\lambda I}$ regions represented by BALB/c mouse plasmacytomas arise from no more than a few copies of germ line V DNA segments (Tonegawa, 1976; Honjo et al., 1976). There must be a somatic mechanism by which the coding sequence of the germ line V_λ DNA is diversified in clones of immunoglobulin-producing lymphocytes. Our subsequent restriction enzyme mapping of whole cell DNA confirmed this notion and strongly suggested that there is only a single copy of $V_{\lambda I}$ DNA per haploid genome (Brack et al., 1978; M. Hirama and S. Tonegawa, unpublished observations). Since it is highly improbable that the somatic mechanism will generate identical V regions at a frequency as high as one out of three cases, Cohn et al. (1974) postulated that the V region of the λ_0 type chain is a product of the germ line V DNA. The present DNA sequence of embryonic V_λ DNA of Ig 99λ confirmed this postulate.

Plasmacytoma H2020 synthesizes a λ chain whose amino acid sequence is different from the λ_0 type chain sequence at positions 25 and 32 (Weigert et al., 1970). Since Ig 303λ was cloned from an H2020 DNA fraction that had been enriched for the somatically rearranged active λ gene, we expected its V DNA segment to code for the H2020 V region. As shown in Figure 3, this expectation was fulfilled. The DNA sequences of Ig 99λ and Ig 303λ show

that in each of the two positions where the λ_0 and the H2020 V regions are different, the involved triplet codons are related by a single base difference. The DNA sequences of the two clones are identical except for these two base positions until the His codon at position 97. We observed neither silent base changes in the protein-encoding regions, nor any difference in the 5' noncoding regions or in the 93 nucleotide long intervening sequences (see below for additional discussion of intervening sequences). These results are compatible with the notion that lymphocytes have a special enzymatic mechanism that preferentially introduces mutations to the hypervariable regions (Brenner and Milstein, 1966). The results, however, are also compatible with the hypothesis that mutations occur indiscriminately along the entire V DNA segment at a relatively low rate, and only those V DNAs that have acquired missense mutations in hypervariable regions are selected at the cellular level. In this case, our results show that the rate of mutation should be sufficiently low that the appearance of neutral base changes throughout the V DNA segment is infrequent.

Multiple Intervening Sequences in Both Embryo and Plasmacytoma λ Chain Genes

The sequence of Ig 99λ DNA revealed a 93 bp long intervening sequence toward the amino terminal end of the embryonic $V_{\lambda I}$ DNA segment. The position and length of this intervening sequence are identical with those found in the embryonic $V_{\lambda II}$ DNA segment (Tonegawa et al., 1978), although the two sequences are different at four nucleotide positions. We term the two intervening sequences $I_{\lambda I}$-small and $I_{\lambda II}$-small. $I_{\lambda I}$-small is also present in Ig 303λ, indicating that somatic rearrangement at the DNA level does not eliminate the noncoding sequence. The 1.2 kb intervening sequence present in Ig 25λ (termed $I_{\lambda I}$-large) also remains in the active rearranged complete λ_1 chain gene represented by Ig 303λ. Sequencing of the C DNA segment did not reveal any additional intervening sequence in the region extending from position 111 to position 155. Our preliminary sequence data for the rest of the C DNA segment indicate that no additional coding interruption occurs in this part of the molecule (O. B. and S. T., unpublished data). Thus it seems that a complete λ_1 gene has two intervening sequences, $I_{\lambda I}$-small and $I_{\lambda I}$-large, in the translated portion of the gene. From the size of the λ mRNA (Tonegawa, 1976) and from the known anatomy of the κ mRNA (Milstein et al., 1974), we expect that there are about 150 and 200 base long untranslated regions at the 5' end and the 3' end of a λ mRNA molecule, respectively. R loop mapping of both Ig 99λ and Ig 303λ did not give any hint of

additional intervening sequences in these untranslated regions (Brack et al., 1978). The presence of short (<100 base) intervening sequences, however, would not be revealed by this procedure. In this respect, it is interesting to note that in both Ig 99λ and Ig 303λ, three repeats of CTG are present about 50 nucleotides upstream from the translation initiation codon AUG for Met at position −19 (Figure 3). The same nonanucleotide was present at the equivalent position in the embryonic $V_{\lambda II}$ clone, Ig 13λ (Tonegawa et al., 1978). A cloned rabbit β–globin DNA fragment (A. Efstratiadis, T. Maniatis and E. Lacy, personal communication) and a cloned hen ovalbumin DNA fragment (Breathnach et al., 1978) also carry a similar repeat of CTG about 60 and 35 nucleotides upstream from the AUG initiation codon, respectively. In the case of the ovalbumin DNA clone, the CTG repeat is present within the 5' end-most intervening sequence and near its junction with the adjacent DNA sequence which is expressed in the mature ovalbumin mRNA. It is tempting to speculate that the CTG repeats have a common role in all these gene systems.

Intervening sequences are thought to be eliminated on the RNA level by a splicing mechanism (Klessig, 1977; Berget, Moore and Sharp, 1977). While this remains to be demonstrated in the case of the mouse λ_I chain gene, evidence in support of this notion is accumulation for other gene systems (Knapp et al., 1978; O'Farrell et al., 1978; Tilghman et al., 1978). What are the recognition signals for the putative RNA splicing enzymes? Probable candidates are hairpin structures and/or base sequences at or near the boundaries of protein-encoding and intervening sequences. We have previously reported that a hairpin structure of considerable stability can be formed around the $V_{\lambda II}$-small intervening sequences (Tonegawa et al., 1978). While a similar hairpin structure can also be formed around the $V_{\lambda I}$-small intervening sequences, nucleotide sequences around the larger intervening sequences do not allow the formation of a hairpin structure of analogous stability. As we have previously pointed out (Tonegawa et al., 1978), however, we should keep in mind the possibility that a sequence lying elsewhere in the RNA transcript may hold the ends of adjacent protein-encoding sequences in juxtaposition and facilitate RNA splicing.

Do the nucleotide sequences at the junction of protein-encoding and intervening sequences exhibit unusual features? The sequences of the λ_I and λ_{II} DNA at the splice points are listed in Figure 4. We notice the following sequence features. First, short nucleotide sequences are directly repeated at the ends of a given intervening sequence. Thus a

Figure 4. Sequence Features at the Boundaries of Protein-Encoding and Intervening Sequences

Three intervening sequences of the mouse λ_I (present paper) and λ_{II} (Tonegawa et al., 1978) genes are compared. Short sequences repeated at the opposite ends of an intervening sequence are underlined. Short sequences that are common among the three intervening sequences and are present at or near either their 5' or 3' ends are indicated by dots over the nucleotides. Brackets indicate the maximal and minimal boundaries of the intervening sequences.

tetranucleotide CGAA is repeated at opposite ends of both $V_{\lambda I}$-small and $V_{\lambda II}$-small intervening sequences, while a dinucleotide GG is repeated at the ends of the $V_{\lambda I}$ large intervening sequence. It is partly for this reason that unique identification of the exact splice points is impossible (see above). Second, an oligonucleotide AGGTXAG appears at the 5' end of the three intervening sequences, while no common sequence other than a dinucleotide GG is present at the 3' ends. These sequence features apply to other genes carrying intervening sequences, such as mouse and rabbit β–globin gene (J. Van den Berg, A. Van Ooyen, N. Mantei, A. Schambäck, G. Grosveld, R. A. Flavell and C. Weissmann, personal communication), hen ovalbumin genes (Breathnach et al., 1978), and SV40 early and late genes (Reddy et al., 1978; Ghosh et al., 1979). In all cases, 1–4 nucleotides are repeated at the opposite ends of a given intervening sequence. In most cases, either AGGTXAG or sequences related to it appear at the 5' end. We believe that these sequences constitute at least part of the recognition signals for the splicing enzymes.

The presence of short direct repeats at the boundaries of all intervening sequences may be a requirement for the enzymatic reaction leading to RNA splicing. This is reminiscent of an analogous situation at the ends of insertion elements and transposons. In all cases investigated so far, these procaryotic DNA elements are flanked by a direct repeat of five or nine bases at the site of insertion (Grindley, 1978; Calos, Johnsrud and Miller, 1978; Rosenberg et al., 1978; L. Johnsrud, M. P. Calos and J. H. Miller, personal communication). It is possible that the short repeats at the boundaries of intervening sequences are evolutionary relics of their initial insertion events, which might have occurred by a mechanism similar to transposition of the procaryotic insertion elements. If this is the case, however, the origin of intervening sequences

is not in accord with a proposal made previously. We hypothesized that the intervening sequences are there not because they were inserted into a preexisting gene, but because genes are often created in evolution as a result of the emergence of a pair of RNA splice points at two separate DNA sites within a preexisting transcription unit (Brack et al., 1978; Tonegawa et al., 1978).

Comparison of Nucleotide Sequences of Ig 99λ and Ig 13λ

Mouse λ chains are classified into the two subtypes λ_I and λ_{II}. Amino acid sequences of the two subtype chains are different at ~10–12 positions in the V region and ~25 positions in the C region (Weigert et al., 1970; Dugan et al., 1973). Total embryonic DNA gives two Eco RI fragments of 4.8 and 3.6 kb carrying sequences homologous to the V part of a λ_I hybridization probe (Tonegawa et al., 1977; Brack et al., 1978). We have previously cloned the 4.8 kb fragment and identified its V_λ DNA as that of a λ_{II} type by DNA sequencing (Tonegawa et al., 1978). Present identification of Ig 99λ (contains the 3.6 kb fragment) as a clone carrying a $V_{\lambda I}$ DNA sequence confirmed the notion that the two λ subtypes are encoded by separate germ line V DNA segments. Although a $C_{\lambda II}$ DNA segment has not been identified (it does not appear in a Southern gel blot assayed with a λ_I hybridization probe, presumably because of the lack of sufficient sequence homology between the two C DNA segments), it is almost certain that such a DNA segment exists as a separate entity other than the $C_{\lambda I}$ DNA segment.

Figure 5 illustrates the difference in the DNA sequences of Ig 99λ and Ig 13λ around the V DNA segments. Comparison was made for ~350 base long protein-encoding regions, the 5' end noncoding region, the 93 base long small intervening

sequences and the 3' end noncoding regions. The cumulative length of the three noncoding regions amounts to about 380 bp. As shown in Figure 5, the basic architecture of the two V DNA segments and the flanking regions is very similar in the two genes. A minor difference was observed at the end of the V-coding DNA segment. As mentioned above, Ig 99λ codes for the $V_{\lambda I}$ region up to position 97, while we have previously reported that Ig 13λ codes for the $V_{\lambda II}$ region up to position 98 (Tonegawa et al., 1978). We do not yet know whether the difference in the size of the V-coding region between the two DNA clones reflects a true difference in the distribution of coding sequences between V and J DNA segments in the two λ chain genes. An alternative possibility is that in the λ_{II} gene, two Phe codons (which are not necessarily identical) appear at both the end of the V DNA segment and the beginning of the putative $J_{\lambda II}$ segment (J segment for λ_{II} chain).

In the coding regions, we observed differences in 12 nucleotides. Ten were missense changes, causing assignment of different amino acids at the corresponding positions of the two subtype V regions. The other two differences were silent changes in the third base of triplet codons. In no case was more than one base difference observed within a corresponding pair of codons. In the noncoding regions, seventeen differences were observed within the approximately 400 bp long stretches that have been sequenced. Of these, three were in the 5' noncoding region (85 bp), four were in the small intervening sequences (93 bp) and ten were in the 3' noncoding region (200 bp). In the 3' noncoding regions, six differences were due to three base long deletions or insertions occurring at two different sites. In this respect, it is interesting to note that the 3' noncoding regions are rich in a short stretch of identical nucleotides,
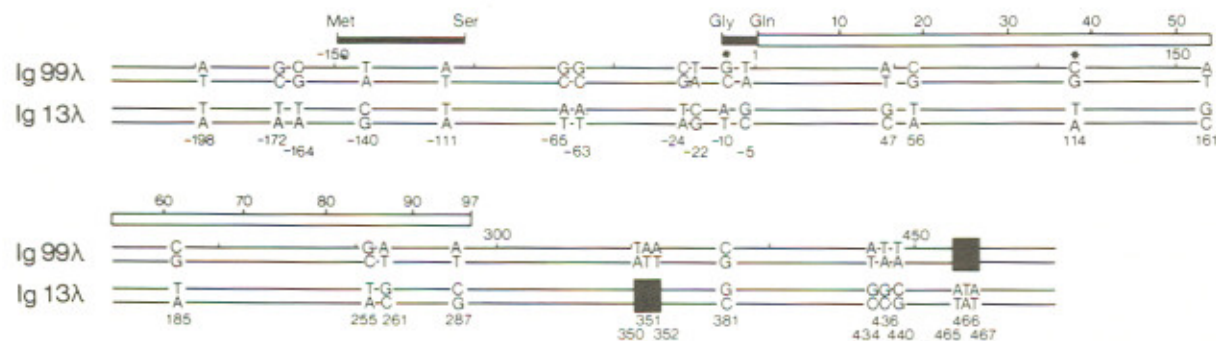

Figure 5. Schematic Illustration Comparing Nucleotide Sequences of Ig 99λ and Ig 13λ

Numbers indicate either nucleotide positions or amino acid positions. For nucleotide positions, number 1 is given to the first nucleotide of the triplet codon corresponding to the cyclized glutamine found at the amino terminus of mature light chains. Where the two sequences are different, the nucleotides involved are indicated together with their positions. Asterisks (*) indicate neutral base changes in the protein-encoding regions. The leader-encoding part is indicated with a black bar, while the mature V region-encoding part is indicated by open bars. Filled boxes indicate deletions or insertions.

particularly of T and A. Except for the 3' end noncoding regions (where the density of base differences may be slightly higher than in other regions), the variation density is no higher in the noncoding regions than in the protein-encoding regions. The noncoding regions in question are considered to be transcribed into pre mRNA. The functions of the noncoding transcribed region may require a rather rigorous secondary structure, and this may be the primary reason why the nonprotein-encoding regions appear to be as conserved as the protein-encoding regions.

Experimental Procedures

Materials

$\gamma-^{32}$P-ATP (spec. act. 3000–5000 Ci/mmole) was purchased from The Radiochemical Centre (Amersham, England). T4 polynucleotide kinase was purified from T4-infected cells by the modified method described by Richardson (1965). Bacterial alkaline phosphatase was obtained from Worthington Biochemicals, and was used after removing the ammonium sulfate by centrifugation and dissolving the enzyme in 10 mM Tris (pH 7.4). Restriction endonucleases Eco RI (Greene, Betlach and Boyer, 1974), Bam HI (Wilson and Young, 1975) and Bgl II (Bickel, Pirrotta and Imber, 1977) were prepared according to published methods. Hinf I was prepared according to unpublished methods of R. J. Roberts (personal communication). All other restriction enzymes were purchased from New England Biolabs.

Phage λgt$_{WES}$ Ig 25λ, 99λ and 303λ were grown in an EK2 host DP50 (SuII$^+$, SuIII$^+$) (Leder, Tiemeier and Enquist, 1977) in a P3 facility in accordance with the NIH Guidelines issued in June 1976.

Preparation of Eco RI Inserts

Phage DNA was digested with Eco RI and the fragments were separated as previously described (Hozumi and Tonegawa, 1976). DNA was eluted from the gel by passing the agarose gel slice through a 22 gauge needle and centrifuging the crushed gel in a Beckman SW41 rotor at 35,000 rpm for 30 min. The DNA was precipitated from the supernatant by the addition of 2 vol of ethanol and dissolved in 10 mM Tris (pH 7.4).

Preparative Gel Electrophoresis

DNA fragments for preparation of restriction enzyme maps and for sequencing were obtained from digestion of 303λ and 25λ DNA with Xba I and Eco RI. 2 mg of phage DNA were digested for 16 hr at 37°C with 100 U of Xba I in 0.15 M NaCl, 6 mM Tris–HCl (pH 7.9), 6 mM MgCl$_2$ and 6 mM 2-mercaptoethanol. For Eco RI digestion, the concentration of Tris buffer was adjusted to 0.1 M with 2 M Tris (pH 7.2) and the DNA was incubated with 1000 U of Eco RI for 1 hr. DNA was then concentrated twice with equal volumes of butanol-2 (Stafford and Bieber, 1975), dialyzed against 10 mM Tris (pH 7.4) for 4 hr and loaded onto a 5% acrylamide gel (slot size 10 × 0.5 cm). The mouse DNA insert fragments were extracted as described (Maxam and Gilbert, 1977), except that the eluate was passed through a glass-wool layer in a 5 ml syringe before ethanol precipitation. Ig 99λ DNA was digested with Eco RI, Xba I and Bgl II to yield three mouse DNA fragments (Figure 1) which were separated and eluted as described above.

5' Terminal Labeling of Restriction Endonuclease Fragments

DNA fragments (10 μg) were incubated with 1 U of alkaline phosphatase for 1 hr in 50 mM Tris–HCl (pH 9), 10 mM MgCl$_2$, 6 mM dithiothreitol, and then extracted 3 times with phenol saturated with water. Residual phenol was removed by ether extrac-

tion followed by two ethanol precipitations. The DNA fragments were incubated at 37°C for 1 hr in 100 μl of 50 mM Tris–HCl (pH 7.5), 10 mM MgCl$_2$, 10 mM β-mercaptoethanol, 0.085 mM spermidine, 0.5 mCi $\gamma-^{32}$P-ATP and several units of T4 polynucleotide kinase. The reaction was stopped by adjusting the ammonium acetate concentration to 2 M. The majority of the free $\gamma-^{32}$P-ATP was removed by two successive ethanol precipitations. The fragments were separated on an 8% acrylamide gel and eluted as indicated above.

DNA Sequencing

DNA fragments labeled at their 5' end were cut with another restriction enzyme or the strands were separated as described by Maxam and Gilbert (1977). DNA sequencing was also performed according to the procedure described by Maxam and Gilbert (1977). For sequencing long fragments (more than 150 nucleotides), the DNA samples were loaded onto 0.5 × 300 × 400 mm, 8, 20 and 24% acrylamide gels.

References

Appella, E. (1971). Amino acid sequences of two mouse immunoglobulin lambda chains. Proc. Nat. Acad. Sci. USA *68*, 590–594.

Berget, S. M., Moore, C. and Sharp, P. A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc. Nat. Acad. Sci. USA *74*, 3171–3175.

Bickle, T. A., Pirrotta, V. and Imber, R. (1977). A simple, general procedure for purifying restriction endonucleases. Nucl. Acids Res. *4*, 2561–2572.

Brack, C. and Tonegawa, S. (1977). Variable and constant parts of the immunoglobulin light chain gene of a mouse myeloma cell are 1250 nontranslated bases apart. Proc. Nat. Acad. Sci. USA *74*, 5652–5656.

Brack, C., Hirama, M., Lenhard-Schuller, R. and Tonegawa, S. (1978). A complete immunoglobulin gene is created by somatic recombination. Cell *15*, 1–14.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978). The ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-nitron boundaries. Proc. Nat. Acad. Sci. USA, in press.

Brenner, S. and Milstein, C. (1966). Origin of antibody variation. Nature *211*, 242–243.

Bukhari, A. I., Shapiro, J. A. and Adhya, S., eds. (1977). DNA Insertion Elements, Plasmids and Episomes (New York: Cold Spring Harbor Laboratory).

Burstein, Y. and Schechter, I. (1977a). Glutamine as a precursor to N-terminal pyrolid-2-one–5–carboxylic acid in mouse immunoglobulin λ-type light chains. Biochem. J. *165*, 347–354.

Burstein, Y. and Schechter, I. (1977b). Amino acid sequence of the NH$_2$-terminal extra piece segments of the precursors of mouse immunoglobulin λ_1-type and κ-type light chains. Proc. Nat. Acad. Sci. USA *74*, 716–720.

Calos, M. P., Johnsrud, L. and Miller, J. H. (1978). DNA sequence at the integration sites of the insertion element IS1. Cell *13*, 411–418.

Campbell, A. M. (1962). Episomes. Adv. Genet. 11, 101–145.

Cesari, I. M. and Weigert, M. (1973). Mouse lambda-chain sequences. Proc. Nat. Acad. Sci. USA 70, 2112–2116.

Cohen, S. N. (1976). Transposable genetic elements and plasmid evolutions. Nature 263, 731–738.

Cohn, M., Blomberg, B., Geckler, W., Raschke, W., Riblet, R. and Weigert, M. (1974). First order considerations in analyzing the generator of diversity. I. In The Immune System, Genes, Receptors and Signals, E. E. Sercarz, A. R. Williams and C. F. Fox, eds. (New York: Academic Press), pp. 89–117.

Dugan, E. S., Bradshaw, R. A., Simms, E. S. and Eisen, H. N. (1973). Amino acid sequence of the light chain of a mouse myeloma protein (MOPC-315). Biochemistry 12, 5400–5416.

Ghosh, P. K., Reddy, V. B., Swinscoe, J., Lebowitz, P. and Weissman, S. U. (1979). The heterogeneity and 5′ terminal structures of the late RNAs of simian virus 40. J. Mol. Biol., in press.

Greene, P. J., Betlach, M. C. and Boyer, H. W. (1974). The EcoRI restriction endonuclease. In Methods in Molecular Biology, 1, R. B. Wickner, ed. (New York: Marcel Dekker), pp. 87–111.

Grindley, N. D. F. (1978). IS1 insertion generates duplication of a nine base pair sequence at its target site. Cell 13, 419–426.

Honjo, T., Packman, S., Swan, D. and Leder, P. (1976). Quantitation of constant and variable region genes for mouse immunoglobulin λ chains. Biochemistry 15, 2780–2785.

Hozumi, N. and Tonegawa, S. (1976). Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc. Nat. Acad. Sci. USA 73, 3628–3632.

Klessig, D. F. (1977). Two adenovirus mRNAs have a common 5′ terminal leader sequence encoded at least 10 kb upstream from their main coding regions. Cell 12, 9–21.

Knapp, G., Beckmann, J. S., Johnson, P. F., Fuhrman, S. A. and Abelson, J. (1978). Transcription and processing of intervening sequences in yeast tRNA genes. Cell 14, 221–236.

Landy, A. and Ross, W. (1977). Viral integration and excision: structure of the lambda att sites. Science 197, 1147–1160.

Leder, P., Tiemeier, D. and Enquist, L. (1977). EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the λgt$_{WES}$ system. Science 196, 175–177.

Lenhard-Schuller, R., Hohn, B., Brack, C., Hirama, M. and Tonegawa, S. (1978). DNA clones containing mouse immunoglobulin κ chain genes isolated by in vitro packaging into phage λ coats. Proc. Nat. Acad. Sci. USA 75, in press.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA 74, 560–564.

Milstein, C., Brownlee, G. G., Cartwright, E. M., Jarvis, J. M. and Proudfoot, N. J. (1974). Sequence analysis of immunoglobulin light chain messenger RNA. Nature 252, 354–359.

O'Farrell, P. Z., Cordell, B., Valenzuela, P., Rutter, W. J. and Goodman, H. M. (1978). Structure and processing of yeast precursor tRNAs containing intervening sequences. Nature 274, 438–445.

Reddy, V. B., Thimmappaya, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. and Weissman, S. M. (1978). The genome of simian virus 40. Science 200, 494–502.

Richardson, C. C. (1965). Phosphorylation of nucleic acid by an enzyme from T4 bacteriophage-infected Escherichia coli. Proc. Nat. Acad. Sci. USA 54, 158–165.

Rosenberg, M., Court, D., Wulff, D. L., Shimatake, H. and Brady, C. (1978). The relationship between function and DNA sequence in an intercistronic regulatory region in phage λ. Nature 272, 414–423.

Shulman, M. J. and Gottesman, M. E. (1973). Attachment site mutants of bacteriophage lambda. J. Mol. Biol. 81, 461–482.

Smith, H. O. and Birnstiel, M. L. (1976). A simple method for DNA restriction site mapping. Nucl. Acids Res. 3, 2387–2398.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503–517.

Stafford, D. W. and Bieber, D. (1975). Concentration of DNA solutions by extraction with 2-butanol. Biochim. Biophys. Acta 378, 18–21.

Starlinger, P. and Saedler, H. (1976). IS-elements in microorganisms. Curr. Topics Microbiol. Immunol. 75, 111–152.

Tilghman, S. M., Curtis, P. J., Tiemeier, D. C., Leder, P. and Weismann, C. (1978). The intervening sequence of a mouse β-globin gene is transcribed within the 15S β-globin mRNA precursor. Proc. Nat. Acad. Sci. USA 75, 1309–1313.

Tonegawa, S. (1976). Reiteration frequency of immunoglobulin light chain genes: further evidence for somatic generation of antibody diversity. Proc. Nat. Acad. Sci. USA 73, 203–207.

Tonegawa, S., Hozumi, N., Matthyssens, G. and Schuller, R. (1976). Somatic changes in the content and context of immunoglobulin genes. Cold Spring Harbor Symp. Quant. Biol. 41, 877–889.

Tonegawa, S., Brack, C., Hozumi, N. and Schuller, R. (1977). Cloning of an immunoglobulin variable region gene from mouse embryo. Proc. Nat. Acad. Sci. USA 74, 3518–3522.

Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. and Gilbert, W. (1978). Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. Proc. Nat. Acad. Sci. USA 75, 1485–1489.

Weigert, M. G., Cesari, I. M., Yonkovich, S. J. and Cohn, M. (1970). Variability in the lambda light chain sequences of mouse antibody. Nature 228, 1045–1047.

Wilson, G. A. and Young, F. E. (1975). Isolation of a sequence-specific endonuclease (Bam I) from Bacillus amyloliquefaciens H. J. Mol. Biol. 97, 123–125.

Note Added in Proof

Nobumichi Hozumi's present address is: The Ontario Cancer Institute, 500 Sherbourne Street, Toronto, Ontario, Canada.