6 October 1978, Volume 202, Number 4363 SCIENCE

Antibody Diversity

The structure of cloned immunoglobin genes suggests a mechanism for generating new sequences.

J. G. Seidman, Aya Leder, Marion Nau
Barbara Norman, Philip Leder

Summary. Three important aspects of immunoglobulin gene organization and structure have emerged from studies of cloned immunoglobulin kappa chain genes. (i) Multiple variable genes are encoded separately in the genome of both immunoglobulin-producing and uncommitted (embryonic) cells, thereby establishing the evolutionary base for generating immunoglobulin diversity. (ii) These genes exist as many small, closely related families (subgroups) that share close sequence homology largely within their own subgroup. (iii) Comparison of two cloned variable gene segments derived from a single subgroup reveals a feature of their structure that distinguishes them from fixed genes (that is, globin genes) and provides, through extensive surrounding sequence homology, a large target for intergenic recombination. This last observation suggests that a simple recombination mechanism may account for their genetic instability in both germ line and somatic cells.

The notion that antibody gene diversity arose through the amplification of a primitive gene family and the separate evolution of individual immunoglobulin genes was the simplest and originally the most popular framework in which to consider the genetic representation of antibody molecules (*1*). A considerable body of structural evidence, principally accumulated from inbred strains of mice, supported this hypothesis (*1a*). Nevertheless, considerations of genetic parsimony and common structural determinants quickly led to proposals of a group of models suggesting that the germ line contained few antibody genes and that further immunoglobulin gene diversification occurred during somatic differentiation (*2*, *2a*). That some unusual arrangement of immunoglobulin genes was likely became clear from the suggestion of Dreyer and Bennett that constant and variable regions were separately encoded in the genome (*3*). Their suggestion has been supported by the finding of only a small number of constant region genes (*4*, *5*) and, more recently, by evidence of the somatic rearrangement of variable and constant region genes (*6*, *7*). This suggestion, coupled with the fact that immunoglobulins are formed from two different polypeptide subunits, immediately led to an understanding of how a relatively small set of genes might produce millions of different antibody molecules without the use of the entire vertebrate genome for this purpose. Diversity could arise from the permutations and combinations possible between immunoglobulin heavy and light chains and from combining a large array of variable region genes or their products with a very small set of constant region genes (*3*). Nevertheless, the question of how diverse variable region genes arose remained.

As more detailed versions of the germ line hypothesis were advanced (*3*, *8*), these took into account the structural homologies retained by small sets or subgroups of variable sequences and suggested that these were the consequence of more recent evolutionary duplications in the germ line genome. In contrast, the more modern versions of the somatic mutation hypothesis held that these closely related sequences arose by somatic mutation from a common germ line precursor (*9*). These models now appear to be converging in that requirements of somatic models invoke very few changes in a moderately large array of germ line genes.

The Arguments

We propose two arguments concerning the origin of diversity of antibody genes. The first holds that the germ line genome is sufficiently rich in variable region genes so that it encodes a very large antibody repertoire. This large array of antibody genes does not rule out somatic mechanisms for generating additional diversity, but does tend to relegate such mechanisms to a secondary role in the final development of antibody diversity. Our second argument concerns a special feature of the structure of variable region genes that distinguishes them from evolutionarily fixed genes (such as the two β-globin genes). This feature, a close structural homology between genes that extends thousands of bases into their flanking sequences, allows us to account for their apparent genetic flexibility through recombination between nonallelic variable region genes, a mechanism that does not require special enzymatic mechanisms available only to antibody-producing cells. Such a recombinational mechanism closely resembles that proposed by Edelman and Gally (*2a*) to account for the possible somatic diversification of immunoglobulin genes.

The authors are all associated with the Laboratory of Molecular Genetics, National Institute of Child Health and Human Development, Bethesda, Maryland 20014.

0036-8075/78/1006-0011$01.75/0

Both arguments rest on evidence obtained from the mouse genome by means of recombinant DNA techniques to generate appropriate variable and constant region clones. By generating variable region probes directed against two different kappa light chain subgroups, two nonoverlapping sets of Eco R1 (*10*) fragments of BALB/c mouse DNA were identified, each consisting of five to ten fragments and each containing elements of variable region gene sequences. By determining the nucleotide sequence of two of these genes and portions of their flanking sequences (*11*) and by comparing their structures by means of heteroduplex mapping, we identified extensive regions of homology. Significantly, this close homology appears restricted to members of a subgroup set and does not extend to the entire repertoire of variable region genes. As this extensive stretch of homology is in sharp contrast to what we have observed by comparing two fixed β-globin genes (*12*), we suggest that this homology creates a large target for intragenic recombination. Such recombination could obviously occur in the germ line or somatic (and cultured) cells (or both) expanding and contracting various subgroups while constantly testing and creating new diversity. Detailed evidence and further features of these arguments are presented below.

Identification of Two Sets of Kappa Variable Region Genes

We have shown that a cloned nucleotide sequence corresponding to the messenger RNA (mRNA) for a variable region of a single immunoglobulin light chain (MOPC-149) anneals to at least six discrete Eco R1 fragments of BALB/c mouse DNA (*11*). We have cloned two of these genomic fragments, determined portions of their nucleotide sequence, and, thereby, shown that they contain two closely related kappa variable region gene sequences that extend 336 base pairs from a putative leader sequence to the codon corresponding to amino acid residue 97 (Fig. 1) (*11*). While these two cloned genes are identical in 313 of 336 base positions, such differences as occur tend to cluster in regions that determine complementarity (*11*). Neither fragment contains a constant region sequence or any other cross-hybridizing variable region sequence, although the smaller fragment (3 kilobases) is ten times longer than the variable region gene and the larger fragment (13 kilobases) is 30 times longer than the variable region gene sequence. The two gene sequences were compared to one another and to a portion of the sequence of the gene actually expressed in the MOPC-149 cells from which they were cloned (Fig. 1). [The sequence was determined from a cloned complementary DNA (cDNA) copy of MOPC-149

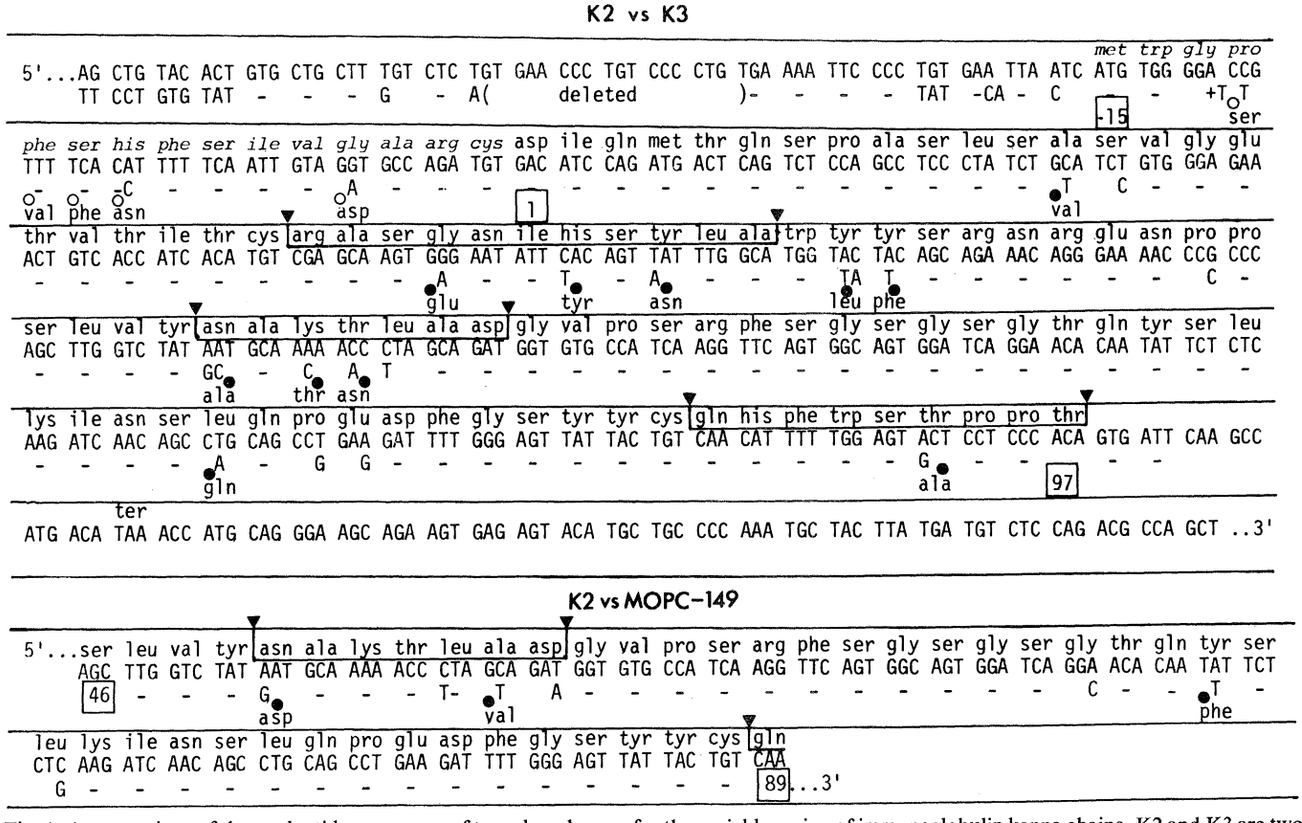


Fig. 1. A comparison of the nucleotide sequences of two cloned genes for the variable region of immunoglobulin kappa chains. K2 and K3 are two cloned nonallelic genes of the variable region of the kappa light chain encoded, respectively, in a 3- and a 13-kilobase (in length) Eco R1 fragment of mouse plasmacytoma DNA (MOPC-149) (*11*). The MOPC-149 is a partial sequence determined from a cDNA copy of MOPC-149 mRNA (*14*). The sequences are reproduced from Seidman *et al.* (*11*). One strand (5′ → 3′) of the K2 gene sequence is shown. Its putative leader sequence is indicated by italics. The amino acid residues are numbered in boxes where the first amino acid residue corresponds to the amino terminal residue of the light chain, MOPC-149 (*25*). The bottom two lines show a comparison of a portion of the K2 sequence to a corresponding portion of the MOPC-149 cDNA sequence. The enclosed amino acid residues correspond to the complementarity-determining segments of the light chain. The K3 sequences are shown as a dash where the sequence is identical to the K2 sequence, and differences in nucleotide sequence and amino acid sequence are indicated. Filled circles flag amino acid differences in the light chain sequence; open circles indicate amino acid differences in the putative leader sequence. Abbreviations for the amino acid residues are Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Met, methionine; Pro, proline; Phe, phenylalanine; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; and Val, valine. Abbreviations for codon bases are A, adenine; T, thymine; C, cytosine; and G, guanine.

mRNA (*11*)]. Again, the cloned genomic genes, K2 and K3, differ from the expressed sequence more intensely in the complementarity determining region.

These cloned genes are so closely related to one another and to the sequence expressed in the MOPC-149 plasmacytoma that one would expect that the remaining fragments also encode other closely related genes. Since this pattern is observed both in embryonic and plasmacytoma DNA (see below), these results suggest to us the possibility that each subgroup consists of a closely related set of genes that are encoded in both germ line and somatic chromosomal DNA.

Strong experimental support for this possibility would come from generating and using several variable region probes corresponding to different subgroups of the mouse kappa light chains (*13*). Accordingly, in addition to MOPC-149 (*14*), we have cloned and characterized a reverse transcript of the mRNA that directs the synthesis of the light chain produced by the plasmacytoma MOPC-41. The MOPC-149 gene sequence has been subcloned and restriction fragments of it have been purified so as to obtain hybridization probes specifically corresponding to the nucleotide sequences encoding the variable or the constant regions or both. The extent of these and other probes is indicated diagrammatically in Fig. 2

With these specific hybridization probes, the distribution and identity of Eco R1 fragments of MOPC-149 plasmacytoma DNA that carry specific variable and constant region sequences were identified by a two-dimensional chromatographic electrophoretic (fingerprinting) technique (*11*) to resolve genomic fragments. One main constant region fragment and at least six different fragments that contain elements of variable region genes can be identified. The constant and variable regions occur on different Eco R1 fragments, suggesting that since no Eco R1 site occurs between constant and variable region sequences in the MOPC-149 mRNA sequence (*14*), these segments are separated from one another by intervening sequences in the mammalian genome. Evidence on this point has been presented for both the mouse kappa and lambda gene systems (*5, 7*).

The variable region sequences identified in Fig. 3 represent the MOPC-149 set (or subgroup) of variable region genes. The question then arises as to whether a probe directed against another light chain variable region sequence will cross-hybridize to a different or to an overlapping set of variable region genes. Such an experiment, involving a cloned MOPC-41 variable plus constant region sequence (codon position 48 through the 3′ untranslated sequence), is shown and compared to the MOPC-149 set in Fig. 4. Nine different bands can be clearly identified, but only the fragment encoding the constant region sequence is detected by both probes. Thus, the MOPC-41 probe is detecting a new and different set of variable gene sequences. If the relationship between these variable regions is like that between the two cloned MOPC-149 variable region sequences, then the MOPC-41 subgroup consists of at least seven or eight discrete variable region genes.

Arrangement of Genes in Embryonic and Plasmacytoma DNA

That analogous variable region genes also occur in the uncommitted, nonlymphocyte genome is shown by identifying Eco R1 fragments in embryonic DNA (Fig. 5). Obviously both embryonic and committed cells encode multiple variable region sequences. Again, with the use of probes directed against both MOPC-149 and MOPC-41 cloned cDNA sequences, it can be seen that the Eco R1 restriction pattern of variable and constant sequences is not significantly different between embryonic and plasmacytoma MOPC-149 DNA (compare the top two panels of Fig. 5 to the corre-

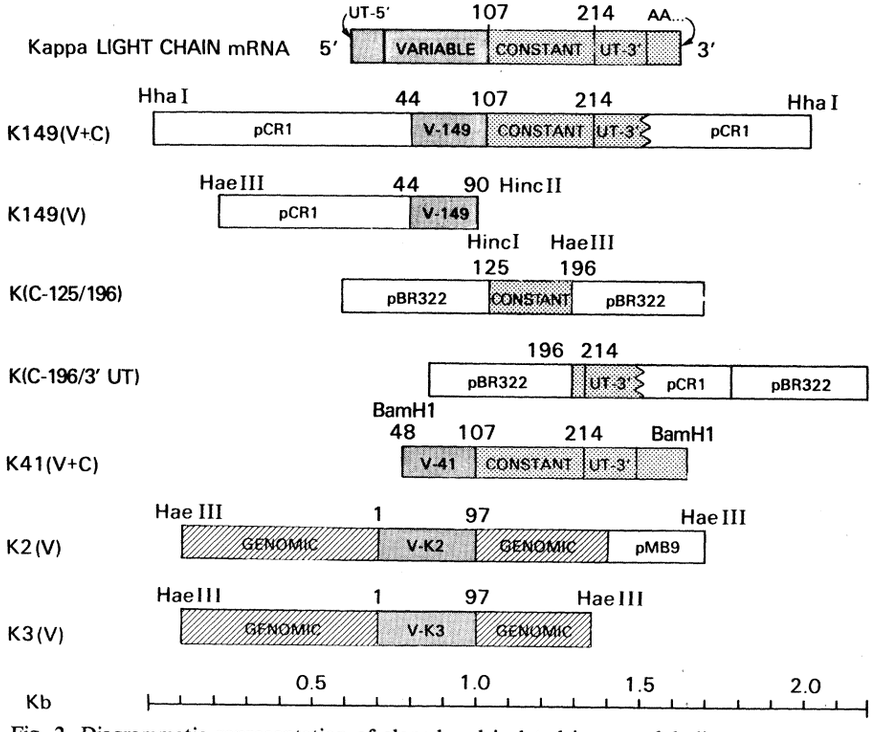


Fig. 2. Diagrammatic representation of cloned and isolated immunoglobulin gene sequences used as hybridization probes. The sequences are shown with reference to the kappa light chain mRNA with its 5′-untranslated (UT-5′) and poly(A) (AA . . .) ends (top line). Restriction enzyme cleavage sites and amino acid positions are indicated above each line. The scale at the bottom is in kilobases. K149 sequences are derived from a cloned cDNA copy of MOPC-149 mRNA (*13*). DNA fragments containing variable or constant region sequences obtained from this plasmid have been recloned in the plasmid pBR322 (*26*); K2 refers to cloned genomic variable region and portions of its flanking sequences (labeled genomic) (*10*). The letters pCR1, pMB9, and pBR322 refer to plasmid vector sequences; K41 represents a cloned cDNA sequence corresponding to the mRNA of another kappa light chain, MOPC-41. The MOPC-41 light chain sequences were cloned in plasmid pBR322 by the synthetic oligonucleotide linker method (*27*). Briefly, this method involves the synthesis of cDNA with total poly(A)-containing mRNA obtained from the MOPC-41 plasmacytoma as template, conversion of this cDNA into a double-stranded molecule by means of reverse transcriptase and subsequent S1 nuclease digestion, and finally ligation of the DNA fragment first to Bam H1 synthetic oligonucleotide linkers and then to pBR322. The transformants obtained in this fashion were screened by hybridization to the DNA fragment which would encode amino acids at residues 125 to 196 of the MOPC-149 constant region. One clone was found to have an 850-base pair insert. This clone, pBR322-K41, made a 618 ± 40 base pair R loop with MOPC-41 mRNA. The 850-base pair Bam H1 fragment from this clone was end-labeled with ^{32}P, and a portion of its sequence was determined by the method of Maxam and Gilbert (*28*). Nucleotide sequence analysis of this clone showed that one end of the cloned fragment encodes amino acids 48 to 68 of the MOPC-41 variable region. Thus pBR322-K41 contains both constant and variable region sequences of the MOPC-41 light chain.

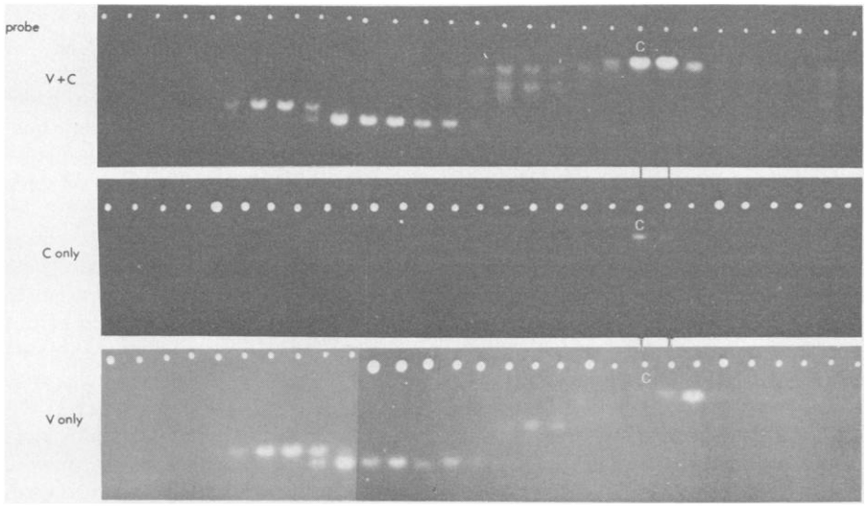


Fig. 3. Two-dimensional analysis of Eco R1 fragments of plasmacytoma MOPC-149 DNA annealed to MOPC-149 variable and constant region gene sequences. Each panel represents a set of Eco R1 DNA fragments separated (i) by RPC-5 chromatography, (ii) by agarose gel electrophoresis, and then (iii) blotted on nitrocellulose filters (*29*). Each filter was hybridized to ^{32}P-labeled variable plus constant (V + C), constant only (C), and variable only (V) gene segments as described (*11*). The probes are identified in Fig. 2. The C refers to the location of the fragment containing the constant region sequence and has been determined by the C-region probes referred to in Fig. 2 [K (C-125/196) and K (C-196/3′ UT)]. The significance of the band intensity is not yet clear, as it appears to be a complex function of many factors including homology, length, mismatch, and transfer efficiency of fragment to filter.

sponding panels in Fig. 4). While this technique does not detect small length variations (< 1 kilobase) in large fragments (> 15 kilobases), this result suggests that, when a rearrangement of constant and variable region genes occurs during differentiation of the lymphocyte as suggested by several studies (*6, 15*), this rearrangement—in these plasmacytomas—occurs outside, or very close to, the Eco R1 sites bordering the separate constant and variable region genes. Further cloning experiments with a variety of plasmacytomas should resolve this point.

Genomic DNA Is Apparently Rich in Variable Region Sequences

The two variable region probes we have used appear to have identified two nonoverlapping sets of variable region genes encoded on at least six to eight Eco R1 fragments of genomic DNA. It is likely, though not yet proved, that these fragments also encode extensive segments of the variable region gene. As MOPC-41 and MOPC-149 are rather typical examples of the mouse kappa light chain sequences, we may reasonably assume that probes corresponding to the 30 or so variable region subgroups already discovered will correspond to similar sets of variable region genes (*15*). This means that the germ line must encode, at a minimum, about 200 variable region genes. But analysis of kappa light chain sequences being characterized in newly generated BALB/c plasmacytomas indicates that the number of variable region isotypes (subgroups) has not been exhausted and could easily approach a hundred (*16*). Such a projection immediately carries the variable region gene numbers in the germ line close to the thousand or so necessary bases to encode the repertoire of the kappa variable region.

Molecular Basis for Germ Line and Somatic Instability of Antibody Genes

While genomic DNA may contain sufficient variable region genes so as to account for a major portion of the antibody repertoire, we must still explain the apparent flexibility of antibody genes in terms of their numbers and sequence diversity. Further, we must account for the apparent genetic lability of antibody genes as demonstrated by their instability in cultured cells (*17*).

An important clue comes from determining the relatedness of antibody genes and their surrounding sequences to one another. We have previously shown that the need for two stable β-globin genes in mammals fits well with evolutionary processes that have reduced to a minimum the homology of sequences that surround these globin genes (*12*). Both β-globin genes retain sequence homology over less than 1000 bases, a region smaller than their 15*S* transcripts (*18*) (even the large central segments of the intervening sequences that divide both β-globin genes differ between them). The effect of this limited homology would be to reduce the possibility of recombination between these segments, and, hence, reduce gene loss or expansion through unequal crossing over. This limited target size for recombination is illustrated by the electron micrograph of the two mouse β-globin genes annealed in the heteroduplex structure shown in Fig. 6A.

By contrast, recombination among closely related immunoglobulin variable region genes may be advantageous to the organism (*2a*). Specific variable region subgroups could be expanded or reduced in either the germ line or somatic genome (a simple diagrammatic representation of this process is shown in Fig. 7). Recombination would per se introduce new diversity, but—as pointed out by Gefter and Fox (*19*)—this diversity could be increased through a mismatch repair mechanism acting upon strand-displaced, recombinant heteroduplex structures (*20*). Mismatch repair, a mechanism by which mismatched base pairs in chromosomal DNA are enzymatically replaced by normal base pairs, would be especially relevant to the generation of diversity at either the germ line or somatic level. A series of closely homologous but nonidentical genes could recombine and presumably "correct" at random the differences between them, creating a new sequence as illustrated in Fig. 8. The differences in clustered complementarity determining regions would be reshuffled, creating novel combinations that, correspondingly, would differ most in complementarity determining regions.

That the structure of variable region genes differs from that of the fixed globin genes can be demonstrated by an analysis of heteroduplex structures formed between the two nonallelic kappa variable region sequences cloned from MOPC-149 genomic DNA. In contrast to the limited homology demonstrated by the globin genes (Fig. 6A), the two cloned kappa segments are homologous over virtually the entire 3000–base pairs sequence compared (Fig. 6B), a sequence that is approximately nine times longer than the variable region coding sequence. This relationship can be confirmed from what we already know of the sequences of these two variable region genes (*11*) that are obviously closely related to one another—although not identical—in the 500 base pairs of structural and flanking sequence that have been compared so far. The 3000-base-long

region of homology is a minimum value, limited in this analysis by the length of the smaller cloned kappa variable region fragment.

A high rate of recombination between homologous genes encoding small differences in base sequence might account for the extreme lability of the variable region phenotype recently identified in cultured cells by Cooke and Scharff (*21*). Obviously, this mechanism might operate in any population of rapidly dividing cells, either in the germ or somatic lines. A comparison of amino acid sequences of variable regions in polypeptides derived from the same patient with multiple myeloma, who was followed over a 10-year period, suggests that such alterations may have occurred in somatic cells in vivo (*22*). Further, the structural features of these genes would appear to facilitate recombination in both germ line and somatic cells, and in neither cell would special enzymes unavailable to other cells be required. Indeed, a high frequency of recombination and mismatch repair has been observed between homologous gene segments in SV40 and polyoma viruses, an indication that there is precedent for the enzymatic reactions required (*20*). The consequences of extending target size for recombination has been illustrated in prokaryotic organisms by expanding transfer RNA (tRNA) genes from two to three copies on a phage chromosome and, thereby, increasing the frequency of unequal crossing-over between these genes approximately 20- to 30-fold (*23*). Similar results have been obtained by increasing the genomic representation of integrated SV40 virus in mammalian cells (*24*).

Problem of Catastrophic Immunoglobin Gene Loss

If extensive homology exists among all kappa variable region genes, it is possible that an unequal crossover might result in a catastrophic loss of variable region genes. Indeed such an event may be responsible for the very limited lambda variable region repertoire in the mouse. One means of preventing extensive immunoglobulin gene loss would be to segregate variable region genes into small sequence-related subgroups. As both K2 and K3 are members of the same subgroup, the question of whether their extensive stretch of sequence homology extends beyond their subgroup can be answered by preparing hybridization probes consisting of both the cloned structural variable region genes, K2 and K3, and their flanking sequences (Fig. 2). If homology is extensively shared between subgroups, many fragments should cross-hybridize to this probe and be identifiable in a two-dimensional restriction fragment "fingerprint." In contrast, if the closely shared homology is restricted to the MOPC-149 subgroup, only those bands hybridizing with the MOPC-149 variable region probe would appear. The result (Fig. 5, bottom two panels) indicates that flanking sequence homology, like structural gene sequence homology, is largely restricted to the small sets of variable region genes that constitute the MOPC-149 subgroup. Extensive homology does not extend between gene sets, although more distantly related regions of reiterated ho-

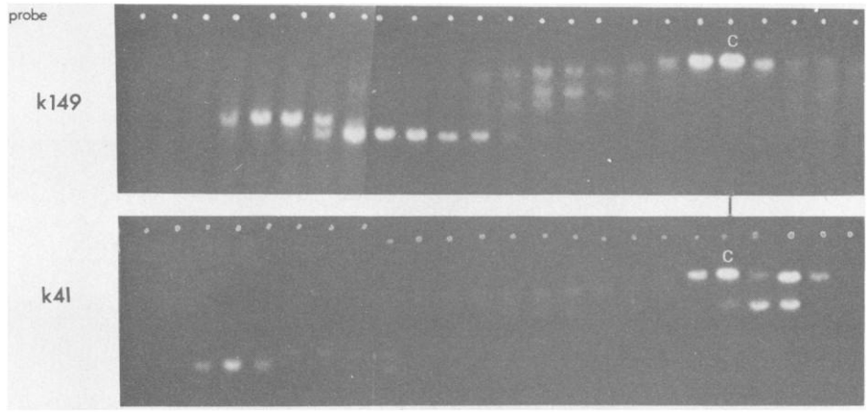


Fig. 4. Two-dimensional analysis of Eco R1 restriction fragments of MOPC-149 DNA hybridized to cloned MOPC-149 (K149) and MOPC-41 (K41) variable and constant region cDNA sequences. The two-dimensional separation and blot were carried out as described in the legend to Fig. 3. The probes were cloned MOPC-149 (K149) and MOPC-41 (K41) variable plus constant region cDNA sequences as indicated in Fig. 2.

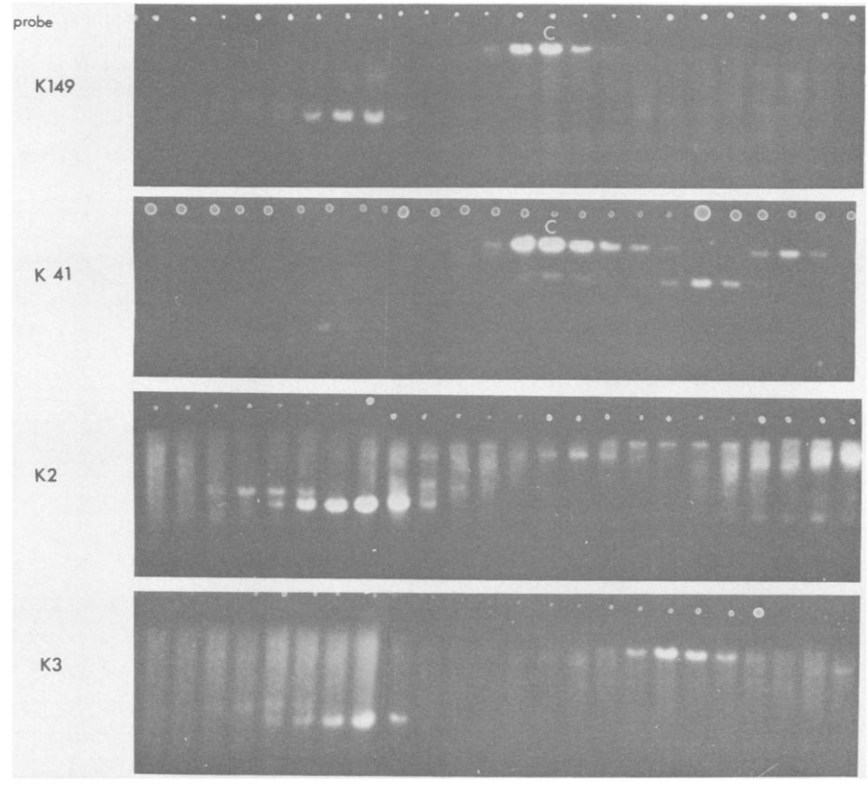


Fig. 5. Two-dimensional analysis of Eco R1 DNA fragments derived from embryonic DNA annealed to cloned MOPC-149 and MOPC-41 constant and variable region cDNA sequences. DNA was derived from 13- to 16-day-old BALB/c mouse embryos. The hybridization probes are as follows: K41, MOPC-41 variable plus constant region; K149, MOPC-149 variable plus constant region. These are identified diagrammatically in Fig. 2. The C identifies the fragment containing the constant region. The broad block of exposure of the last five lanes of the K2 hybridization represents an artifact of the annealing procedure.

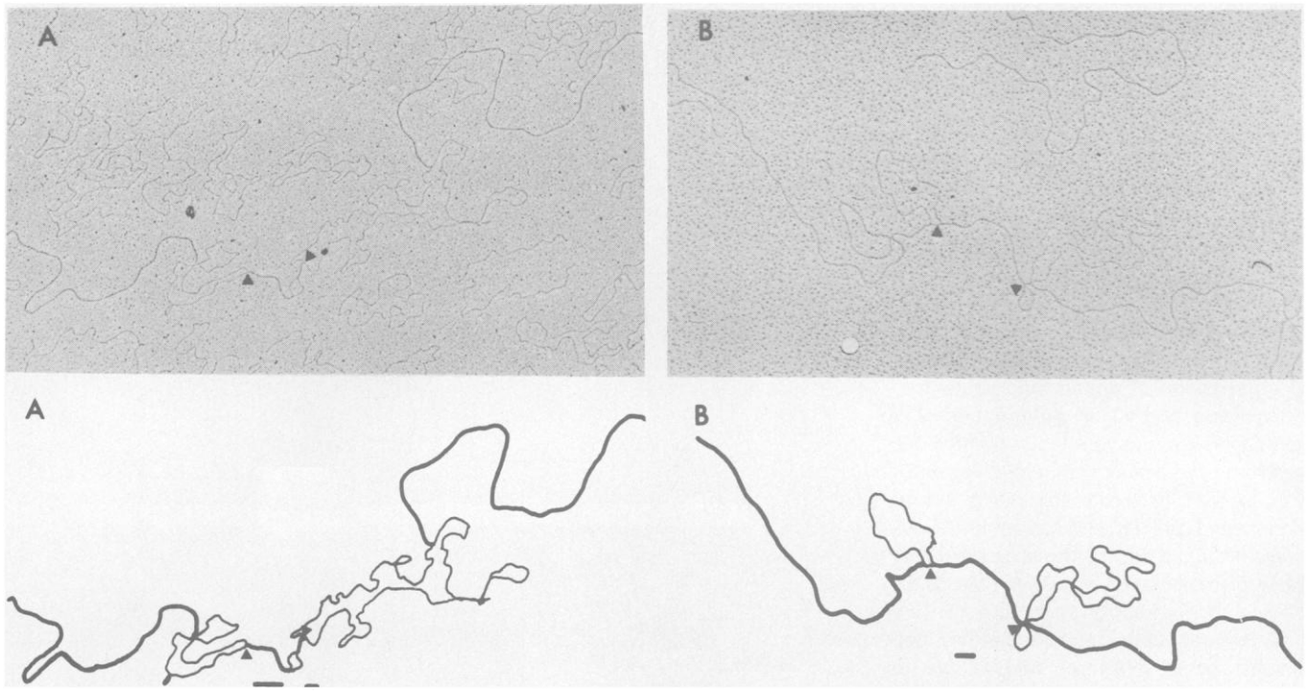
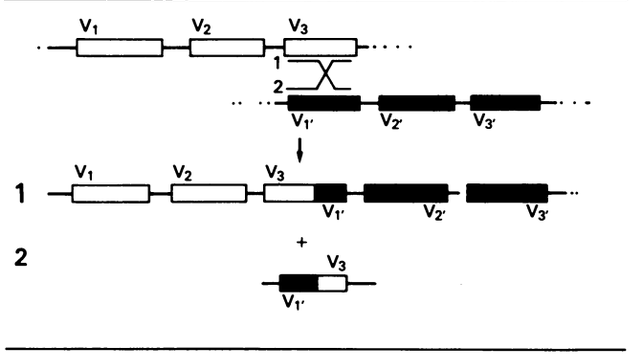


Fig. 6. A comparison of two β-globin and two MOPC-149–like variable region genes and their surrounding sequences by visualization of heteroduplex structures. In each case, heteroduplex structures were formed between two λgtWES·MβG2 and λgtWES·MβG3 and (B) λgtWES·K2 and λgtWES·K3, each carrying a different Eco R1 β-globin (MβG2-MβG3) or variable region gene-containing fragments (K2-K3). Filled triangles represent the limits of homologous regions; the straight lines beneath each line drawing represent, in approximate scale, the lengths of the β-globin and kappa variable region coding segments. (A) The comparison is between β-globin genes on a 14-kilobase (MβG3) and a 7-kilobase (MβG2) fragment of mouse genomic DNA. The annealed portion of the fragment (of the 7-kilobase compared) consists of about 1000 bases of homology (thick strands) interrupted by a single-stranded (thin strands) nonhomologous bubble corresponding to the central portion of large intervening sequences that interrupt both genes between codons 104 and 105. The details of this structure are described in and the electron micrograph is reproduced from Teimeier *et al.* (*12*). (B) One phage carries the 13-kilobase, K3 fragment; the other carries the 3-kilobase, K2 fragment. Each fragment contains a single, 291-base-pair-long variable region gene and flanking sequences (*11*). The sequences are compared over the 3-kilobase length of the smaller fragment. Twenty-six heteroduplex structures were identified and measured. The region of homology yielded a contour length of 2.47 ± 0.79 (± S.D.) kilobases and obviously represents a minimum value.



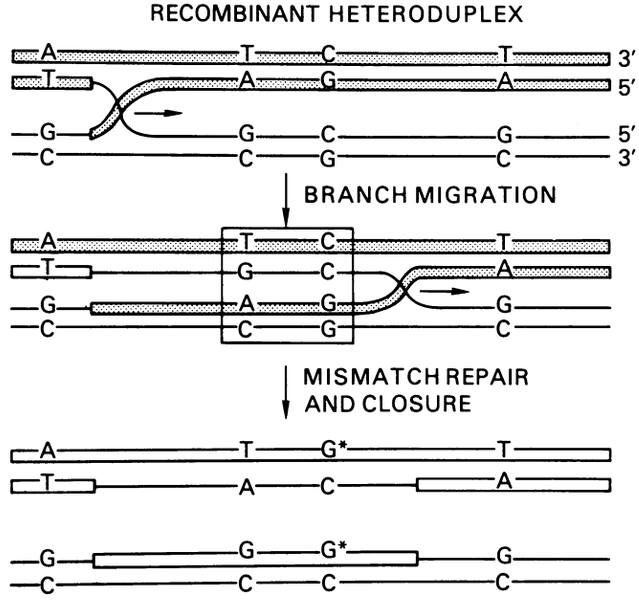


Fig. 7 (left). Gene expression and loss through unequal crossing-over. The diagram represents an unequal crossover event between two homologous, but nonallelic, hypothetical variable region sequences. The reciprocal products to be inherited by separate daughter cells represent an expanded set of genes (three genes to five) and a diminished set of genes (three genes to one). Such unequal crossovers between homologous, nonallelic genes were originally proposed by Edelman and Gally (*2a*) as a means of generating somatic diversity. Fig. 8 (right). Generation of diversity by recombination and mismatch repair. The top strands represent a recombinant heteroduplex structure in which single strands of DNA join two aligned DNA duplexes at a region of close, but not complete, homology. Points at which base sequences differ in the two strands are indicated. In the second set of strands, branch migration has occurred, matching extensive homologous regions from each duplex but also creating small regions of mismatched base pairs (that is, T·G, A·C, C·C, and G·G, as indicated). These mismatched base pairs are converted into authentic Watson-Crick base pairs by appropriate repair enzymes. There are two ways of correcting the mismatch. In one, the original sequence contained in that duplex will be restored. (The first set of mismatches is restored in the last set of strands.) In the other, a new sequence is created (the new sequence in the second set of mismatches is shown as G*). A role for mismatch repair in generating immunoglobulin gene diversity was originally suggested by Gefter and Fox (*19*). That such a mechanism actually can induce (or correct) mutations has been demonstrated with the use of viruses in both pro- and eukaryotic organisms (*20*).

mology are reflected by the higher hybridization background observed with gene and flanking region probes. These patterns do not differ in DNA prepared from embryo (Fig. 4) or from committed, MOPC-149 cells that produce light chains (not shown). Obviously such an arrangement would facilitate intrasubgroup recombination, thereby restricting gene loss (or gross amplification) to a single small set of variable region genes.

Conclusions

Immunoglobulin variable region genes encoding light chains of the kappa type are represented in multiple copies (probably > 200, but possibly close to 1000) in both embryonic cells and cells that produce the kappa light chains. These variable region genes appear segregated into small families of genes (roughly corresponding to subgroups) that are very closely related in sequence, but that differ largely—but not entirely—in their complementarity determining regions. In contrast to the fixed β-globin gene system, the homologies within these small variable region families extend thousands of bases into the sequences that surround the structural variable region genes. This extensive homology appears largely and most stringently restricted to the small set corresponding to a single family or subgroup and does not appear to be closely shared with other kappa subgroups.

The above conclusions are drawn from the results of initial cloning experiments involving the mouse kappa gene system. It is possible, even likely, that a continuing direct analysis of mouse immunoglobin genes will require the alteration of this picture. Nevertheless, from the foregoing we may begin to formulate a plausible and more detailed molecular mechanism for the generation of some—but not necessarily all—immunoglobin gene diversity. Evolution has generated a large number of variable region genes by duplication and segregation, creating many small families (subgroups) of closely and extensively homologous gene segments. These present large target sites for intragenic recombination within (but, not outside) each subgroup. This recombination can, in turn, generate diversity in both gene number and sequence by unequal crossing-over, direct recombination, and mismatch repair. Such a mechanism requires no ad hoc assumptions about special enzymes and can operate in germ line or somatic cells (or both). The immune repertoire is protected from catastrophic gene loss (such as may have occurred in the mouse lambda gene system) by segregation of homology within small sets of "inbreeding" genes.

References and Notes

1. D. W. Talmage, *Annu. Rev. Med.* **8**, 239 (1957); F. M. Burnet, *Aust. J. Sci.* **20**, 67 (1957).
1a. L. E. Hood, M. Potter, D. J. McKean, *Science* **170**, 1207 (1970); L. E. Hood, E. Loh, J. Hubert, P. Barstad, B. Eaton, B. Early, J. Fuhrman, W. Johnson, M. Kronenberg, J. Schilling, *Cold Spring Harbor Symp. Quant. Biol.* **41**, 817 (1976); D. J. McKean, M. Bell, M. Potter, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
2. J. Lederberg, *Science* **129**, 1649 (1959); L. Szilard, *Proc. Natl. Acad. Sci. U.S.A.* **46**, 293 (1960); O. Smithies, *Nature (London)* **199**, 1231 (1963); S. Brenner and C. Milstein, *ibid.* **211**, 242 (1966); E. S. Lennox and M. Cohn, *Annu. Rev. Biochem.* **36**, 365 (1967); O. Smithies, *Science* **157**, 267 (1967); J. A. Gally and G. M. Edelman, *Nature (London)* **227**, 341 (1970); N. K. Jerne, in *Immune Surveillance*, R. T. Smith and M. Landy, Eds. (Academic Press, New York, 1970), p. 345; M. Cohen, *Ann. N.Y. Acad. Sci.* **190**, 529 (1971); N. K. Jerne, *Eur. J. Immunol.* **1**, 1 (1971).
2a. G. M. Edelman and J. A. Gally, *Proc. Natl. Acad. Sci. U.S.A.* **57**, 353 (1967).
→ W. J. Dreyer, and J. C. Bennett, *ibid.* **54**, 864 (1965).
→ C. H. Faust, H. Diggelman, B. Mach, *ibid.* **71**, 2491 (19 → T. Honjo, S. Packman, D. Swan, M. Nau, B. Norman, *ibid.*, p. 3 → P. Leder, T. Honjo, S. Packman, D. Swan, M. Nau, B. Norman, *ibid.*, p. 5109; J. Stavnezer, R. C. C. Huang, E. Stavnezer, J. M. Bishop, *J. Mol. Biol.* **88**, 43 (1974); U. Storb, *Biochem. Biophys. Res. Commun.* **57**, 31 (1974); S. Tonegawa, C. Steinberg, S. Dube, A. Bernadini, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4027 (1974); P. Leder, T. Honjo, D. Swan, S. Packman, M. Nau, B. Norman, in *Molecular Approaches to Immunology*, E. E. Smith and D. W. Ribbons, Eds. (Academic Press, New York, 1976), p. 173; T. H. Rabbitts, J. M. Jarvis, C. Milstein, *Cell* **6**, 5 (1975); T. H. Rabbitts and C. Milstein, *Eur. J. Biochem.* **52**, 125 (1975); M.-G. Farace, M.-F. Aellen, P.-A. Briand, C. H. Faust, P. Vassalli, B. Mach, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 727 (1976); T. Honjo, S. Packman, D. Swan, P. Leder, *Biochemistry* **15**, 2780 (1976); T. Honjo, D. Swan, M. Nau, B. Norman, S. Packman, F. Polsky, P. Leder, *ibid.*, p. 2775; S. Tonegawa, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 203 (1976); _ → _, C. Brack, N. Hozumi, R. Schuller, *ibid.* **74**, 3518 (1977).
5. C. Brack and S. Tonegawa, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5652 (1977).
→ N. Hozumi and S. Tonegawa, *ibid.* **73**, 3628 (1976); S. Tonegawa, N. Hozumi, G. Matthyssens, R. Schuller, *Cold Spring Harbor Symp. Quant. Biol.* **41**, 877 (1976).
7. T. H. Rabbitts and A. Forster, *Cell* **13**, 319 (1978).
8. L. E. Hood, W. R. Gray, W. J. Dreyer, *Proc. Natl. Acad. Sci. U.S.A.* **55**, 826 (1966); D. D. Brown, in *Molecular Genetics and Developmental Biology*, M. Sussman, Ed. (Prentice-Hall, Englewood Cliffs, N.J., 1972), p. 101; H. Wigzell, *Scand. J. Immunol.* **2**, 199 (1973); L. Hood, E. Loh, J. Hubert, P. Barstad, B. Eaton, P. Early, J. Fuhrman, N. Johnson, M. Kronenberg, J. Schilling, *Cold Spring Harbor Symp. Quant. Biol.* **51**, 817 (1976).
9. M. Cohn, B. Blomberg, W. Geckler, W. Raschke, R. Riblet, M. Weigert, in *The Immune System: Genes, Receptors, Signals*, E. E. Sercăz, A. R. Williamson, C. F. Fox, Eds. (Academic Press, New York, 1974), p. 89; M. Weigert and R. Riblet, *Cold Spring Harbor Symp. Quant. Biol.* **51**, 837 (1976).
10. The restriction endonucleases referred to in this article, Eco R1 and Bam H1, are representatives of a class of enzymes that cleave double-stranded DNA at specific base sequences. In this case, each enzyme would be expected to cleave total mouse genomic DNA into about 10^6 different fragments, and these are commonly referred to as restriction fragments.
11. J. G. Seidman, A. Leder, M. H. Edgell, F. I. Polsky, S. Tilghman, D. C. Tiemeier, P. Leder, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3881 (1978).
12. D. C. Tiemeier, S. M. Tilghman, F. I. Polsky, J. G. Seidman, A. Leder, M. H. Edgell, P. Leder, *Cell* **14**, 237 (1978).
13. Immunoglobulin subgroups refer to sets of light or heavy chain polypeptides having very similar amino acid sequences in the variable region. The precise definition of a subgroup has not been agreed upon generally, but involves an approximately 90 percent homology with differences largely restricted to the region determining complementarity.
14. J. G. Seidman, M. H. Edgell, P. Leder, *Nature (London)* **271**, 582 (1978).
15. E. A. Kabat, T. T. Wu, H. Bilofsky, *Variable Regions of Immunoglobulin Genes* (Medical Computer Systems, Bolt Beranek & Newman, Inc., Cambridge, Mass., 1976); M. Potter, *Adv. Immunol.* **25**, 141 (1977).
16. M. Potter, personal communication.
17. K. Adetugbo, C. Milstein, D. S. Secher, *Nature (London)* **265**, 299 (1977); M. D. Scharff, B. Birshtein, B. Dharmgrongartama, L. Frank, T. Kelly, W. M. Kuehl, D. Marguilies, S. L. Morrison, J.-L. Preudhomme, S. Weitzman, in *Molecular Approaches to Immunology*, E. E. Smith and D. H. Ribbons, Eds. (Academic Press, New York, 1975), p. 13.
18. S. Tilghman, P. Curtis, D. Tiemeier, P. Leder, C. Weissmann, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1309 (1978).
19. M. Gefter and M. Fox, personal communication.
20. J. Wildenberg and M. Meselson, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2202 (1975); C.-J. Lai and D. Nathans, *Virology* **66**, 70 (1975); L. K. Miller, B. E. Cooke, M. Fried, *Biochemistry* **73**, 3073 (1976).
21. W. D. Cooke and M. Scharff, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5687 (1977).
22. D. Capra, personal communication.
23. R. L. Russell, J. N. Abelson, A. Landy, M. L. Gefter, S. Brenner, J. D. Smith, *J. Mol. Biol.* **47**, 1 (1970).
24. M. Botcham, personal communication.
25. L. Hood, D. McKean, V. Farnsworth, M. Potter, *Biochemistry* **12**, 741 (1973).
26. E. Bolivar, R. L. Rodriguez, P. J. Green, M. C. Behach, H. L. Heyneker, H. W. Boyer, J. H. Crosa, S. Falkow, *Gene* **2**, 95 (1977).
27. P. H. Seeburg, J. Shine, J. A. Martial, J. D. Baxter, H. M. Goodman, *Nature (London)* **270**, 486 (1977).
28. A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
29. E. M. Southern, *J. Mol. Biol.* **98**, 58 and 503 (1975); D. C. Tiemeier, S. M. Tilghman, P. Leder, *Gene* **2**, 173 (1977).

30. We thank Drs. David McKean, Michael Potter, and Stuart Rudikoff for making the results of their work available to us prior to publication; Dr. Potter for providing the original myeloma tumors used in these studies; Marjorie Sullivan for assistance and advice regarding the electron microscopic analyses; and Terri Broderick for technical assistance.