Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 5180–5183, July 1986
Genetics

Evolution of the immunoglobulin κ light chain locus in the rabbit: Evidence for differential gene conversion events

(enhancer/junctional variation/joining segment of the κ light chain)

MARIE-ANDRÉE AKIMENKO, BERNARD MARIAMÉ, AND FRANÇOIS ROUGEON*

Institut Pasteur, Unité de Génétique et Biochimie du Développement, Département d'Immunologie, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France

Communicated by Jean Dausset, March 18, 1986

ABSTRACT **The rabbit κ light chain gene family is characterized by the presence of two constant region (C_κ) genes; the $C_{\kappa 1}$ gene encodes the constant region of the principal rabbit immunoglobulin light chain, the $C_{\kappa 2}$ gene being not or very poorly expressed in domestic rabbits. There exist four major *K1* alleles (*b4*, *b5*, *b6*, and *b9*), which are unequally expressed in heterozygous rabbits at the *K1* locus. Here, we compare the nucleotide sequences of the joining (*J*) clusters of the κ light chain gene (J_κ) linked to the *b4K2* locus and to the *b4* and *b9* alleles at the *K1* locus. As for C_κ genes, there is evidence for intergenic conversion between the $J_{\kappa 1}$ and $J_{\kappa 2}$ clusters as well as maximum divergence in the expressed *J* segments. The *b9* $J_{\kappa 1}$ cluster differs from its *b4* counterpart in that two out of the five J_κ segments (*J1* and *J2*) are expressed instead of only one. This implies that preferential expression of the *b4* allele as compared to the *b9* allele is not only correlated to the number of available J_κ pieces. The *b9 J2* segment is functional in spite of the presence of a termination codon immediately upstream of its coding region. Two major structural differences were observed between the *J-C* intron sequences of the *b9* and *b4* alleles; namely a 160-base-pair deletion of an A+T-rich sequence in *b9* (which also occurs in the *K2* locus) and a 10-base-pair deletion plus some substitutions in the region corresponding to the mouse κ intron activating element. These differences could underlie the lower transcriptional rate of the *b9* allele.**

The rabbit immunoglobulin κ light chain gene family constitutes an interesting model for the study of complex allele evolution and of the *cis* elements that control gene expression. In contrast to the situation in humans and mice, in domestic rabbits two κ chain loci, *K1* and *K2*, encode the κ chain bearing the nominal *b* allotype of the rabbit and the *bas* chain expressed in wild rabbits and in rabbits of the Basilea strain, respectively (1, 2).

In mature B cells of rabbits heterozygous for the *K1* locus, the two alleles are unequally expressed. The following "pecking order" is observed: *b4* > *b5* > *b6* > *b9* with a *b4*/*b9* allelic ratio of 80:20 (3–6). These alleles are characterized by a high level of divergence at the protein level (ranging from 22 to 33% of divergence). Structural analysis of the corresponding genes has suggested that most of the differences have been generated by gene conversion (7). Another remarkable feature of the *K1* locus is that, at least in the case of the *b4* allele, only one out of the five joining (*J*) segments of the κ light chain (J_κ) is functional (8, 9). The decrease in diversity that could result from this limited combinatorial potential is compensated for by an increased junctional variation during the *V-J* (*V*, variable) recombination because of nucleotide deletions in the J_κ segments and length heterogeneity of the V_κ germ-line segments (10). From nucleotide sequencing data of the human, mouse, and rabbit genes, a highly conserved region has been identified within the *J*-C_κ intron (*C*, constant) (11). This κ intron conserved region (KICR) correlates with a DNase I hypersensitive site (12–14). Furthermore, transient expression of mouse κ light chain genes in myeloma cells reveals that increasing deletions within a segment containing the KICR progressively abolish transcription from the mouse κ promoter (15).

In the present study we have extended the structural analysis of the *b9* allele to its J_κ cluster to compare the evolutionary rates of the J_κ and C_κ segments and to determine whether the difference in expression between the *b4* and *b9* alleles depends on differences in *cis* regulatory elements.

MATERIALS AND METHODS

Restriction map analysis of the recombinant phage λ104 shows that the J_κ locus and part of the *J*-C_κ intron sequence associated with the *b9*C_κ region gene are contained in the 1.5-kilobase (kb), 1.7-kb, and 1.1-kb *Pst* I fragments (Fig. 1). These fragments were separated on a 3.5% acrylamide gel and purified by electroelution. To determine the nucleotide sequence of their extremities, the *Pst* I fragments were first subcloned into the *Pst* I restriction site of M13mp701 phage vector. They were also further digested with *Sau*3A, *Hae* III, and *Alu* I restriction enzymes; the resulting fragments were cloned into the M13mp8 phage vector and transfected into the JM101 *Escherichia coli* cells. Single-stranded DNA templates were prepared and sequence analysis was carried out by the Sanger dideoxynucleotide chain-termination technique (16, 17) using a complementary synthetic oligonucleotide primer and the M13 sequencing kit from Amersham.

RESULTS

The *b9* J_κ Cluster Encodes Two Functional *J* Segments. We have reported the isolation from a λ phage library of the recombinant λ104, containing the genomic *b9* C_κ gene (7). This recombinant has a cluster of five *J* gene segments separated from the *b9* C_κ gene by an intron sequence of 3.1 kb. The restriction maps of the *b4v* and *b9* alleles show a high degree of homology (Fig. 1) except for a deletion of 160 base pairs (bp) located in the intron sequence, which is also observed in the *K2* isotype (see below). Fig. 2 illustrates the strategy used for the sequence analysis of the J_κ locus and part of the intron containing the KICR determined by the M13-dideoxy method (16, 17). The nucleotide sequence of the *b9* J_κ gene segments reveals that only *J1* and *J2* possess both of the features characteristic of expressed J_κ segments (18) (Fig. 3). To the 5′ side of their coding sequences, they contain the correct signal sequences involved in the *V-J* recombination event (i.e., the nonamer GGTTTTTGT, a 23-bp spacer followed by the heptamer CACTGTG) and at their 3′ end the

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

Abbreviations: kb, kilobase(s); bp, base pair(s); C, constant; V, variable; J, joining; KICR, κ intron conserved region.

*To whom reprint requests should be addressed.

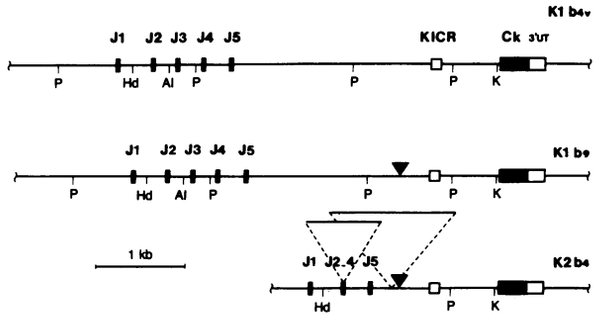

FIG. 1. Restriction maps of the *b9* and *b4v K1* alleles and of the *b4 K2* isotype. The J_κ and C_κ coding regions are represented by the solid rectangles and the 3'-untranslated region of the C_κ genes by the open rectangle. The open square located to the 5' side of the C_κ gene indicates the position of the KICR. The *Kpn* I (K), *Hin*dIII (Hd), and one of the *Pst* I (P) restriction sites are common to the three regions. Because of the two large deletions of the K2 region illustrated by the dotted triangles, the *Ava* I (AI) and the two other *Pst* I restriction sites shared by the *b4v* and *b9* alleles are absent in the *K2* map. The black triangle represents the deletion common to both *b9K1* and *K2* maps.

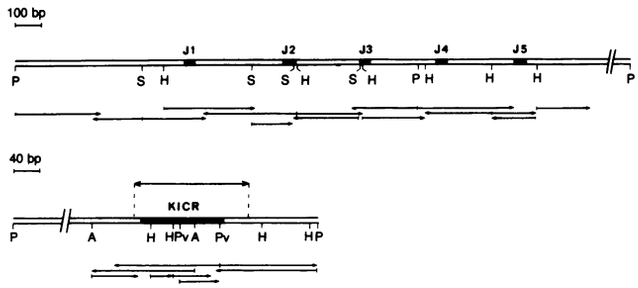

FIG. 2. Sequencing strategies for the *b9* J_κ locus and part of the *J-C* intron. Solid area: *J* coding segments (*Upper*) and KICR (*Lower*). The restriction fragments were subcloned in M13mp701 vectors and subjected to the sequencing procedure of Sanger. Direction and extent of sequencing are indicated by horizontal arrows. P, *Pst* I; S, *Sau*3A; H, *Hae* III; A, *Alu* I; Pv, *Pvu* II.

splicing site of the *J-C* intron (i.e., the dinucleotide GT). In addition their coding sequences are homologous to those of the J_κ human and mouse sequences (19, 20). These results allow us to conclude that only the *J1* and *J2* segments are functional.

The *J3* segment lacks the correct splice site and the spacer of the *J4* segment is composed of only 14 bp. The nonamer and the heptamer of the *J5* segment differ by one nucleotide deletion and two substitutions from the consensus sequences, respectively. Furthermore, at position 6 of the J5 amino acid sequence, the glycine codon, shared by all the expressed J_κ segments, is replaced by a glutamic acid codon. It is noteworthy that, in contrast to the *b9* J_κ locus, of the five b4 J_κ segments only the J2 segment is functional; the spacer between the nonamer and the heptamer preceding the J1 segment has a deletion of 8 bp that renders this segment defective.

J_κ coding sequences are frequently separated from the heptamer by one or two nucleotides. Since the location of *V-J* recombination cut is variable, it can occur upstream or downstream of the extra nucleotide. The rabbit *J2* segments of *K1* and *K2* are characterized by the presence of three extra nucleotides, constituting a complete codon. The three extra nucleotides of the *b9 J2* segment correspond to the termination codon TGA. This means that to utilize the *J2* segment in the *b9* κ chain, the *V-J* recombination can only take place from the second base of the triplet. In *b4* where the TGA codon is replaced by the tyrosine codon, the protein sequence analysis of 27 *b4* κ chains has revealed that the tyrosine extra codon has never been found, indicating that the *V-J* junction event had not occurred upstream from the nucleotide triplet (8). Two cDNA clones of *b9* allotype have been isolated (21). The J_κ segments associated with the *b9* constant region correspond to the *J1* and *J2* genomic segments. In the cDNA containing the *J2* segment, the entire TGA triplet of the genomic sequence has been removed. Therefore, the presence of the termination codon in the reading frame of the *J2* segment does not prevent the use of this segment.

The *b4* and *b9* Allelic Genes Present the Highest Divergence in the Protein Coding Segments. We have reported (7) that highly divergent *b4* and *b9* C_κ gene sequences are embedded in regions of high homology (the allelic C_κ coding regions are 83.4% homologous, while the 5' and 3' surrounding regions are 93.4 and 95.7% homologous, respectively). Furthermore, the nonallelic $C_{\kappa 1}$ and $C_{\kappa 2}$ genes show several segmental homologies. From these two observations, we have concluded that nonreciprocal intergenic conversion occurred in the evolution of the rabbit C_κ genes. We were interested to know if, 3.1 kb upstream from the C_κ genes, the J_κ loci have evolved

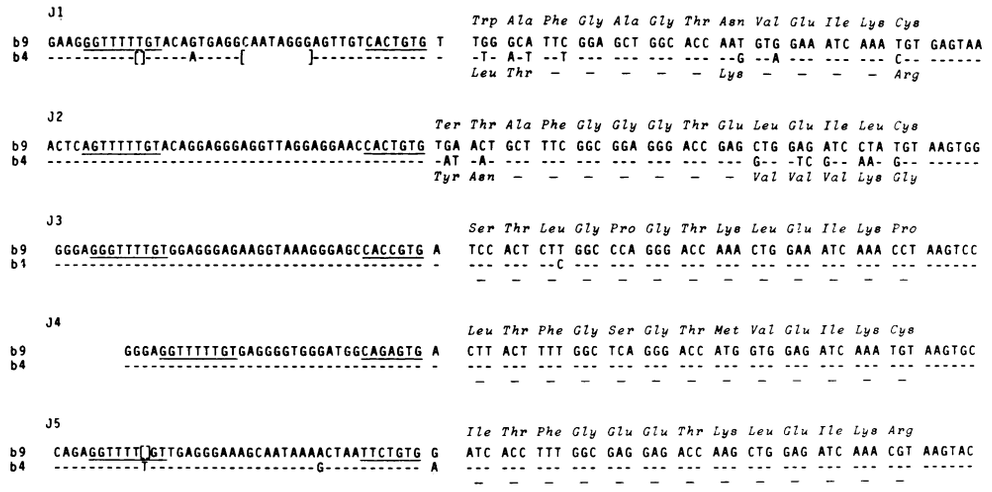

FIG. 3. Comparison of the germ-line *b4* and *b9* J_κ coding sequence and their respective signal sequence involved in the *V-J* recombination. The nonamer and the heptamer elements are underlined. The amino acid sequences of the *J* segments are deduced from their respective nucleotide sequence. In the *b4* amino acid and nucleotide sequences, the dashes represent homologous positions with the *b9* allele, and the brackets indicate gaps introduced to maximize the homology.

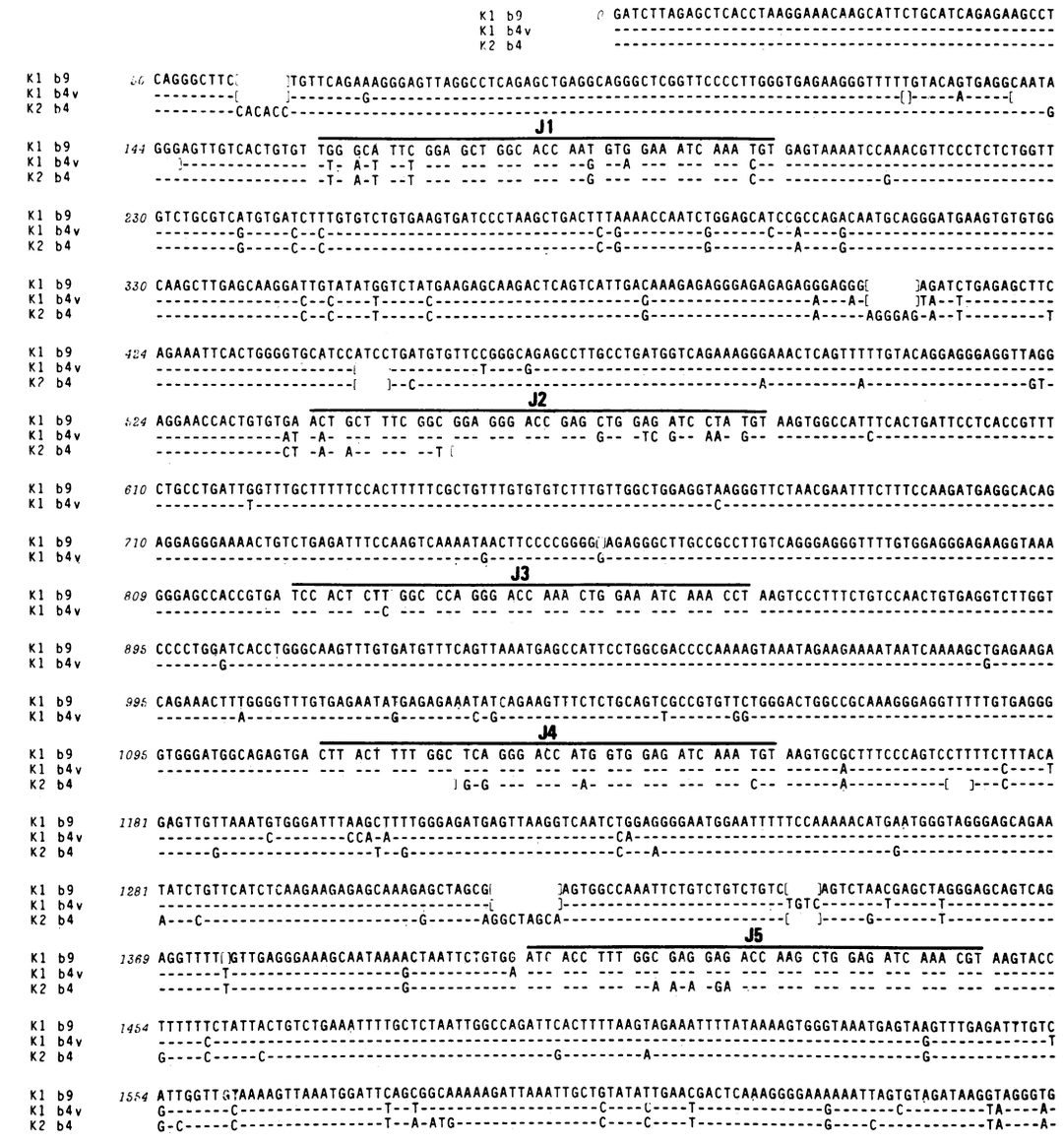

FIG. 4. Nucleotide sequence of the *K1* and *K2* J_κ loci. The *K1 b4v* and *K2 b4* sequences are aligned to maximize the homology with the *K1 b9* sequence; Dashes indicate identity with the *b9* sequence; gaps and deletions are delimitated by brackets. The *b9* nucleotide sequence is numbered from the first base presented. Sequence data for the *K1 b4v* and *K2 b4* regions are taken from Heidmann and Rougeon (8) and Emorine and Max (22), respectively.

in the same way. In Fig. 4, the nucleotide sequence of the *b9 K1* gene has been aligned with that of the *b4v K1* allele (8) and of the *b4 K2* nonallelic form (22). The *K2* locus characterized in a *b4b4v* rabbit (2) has an identical nucleotide sequence to the published *b4 K2* sequence (unpublished data). Gaps have been introduced in either sequences wherever required to maximize the sequence homology. Fig. 5 shows the percent homology calculated for the coding and noncoding sequences between the *b4v* and *b9* J_κ loci. As for the $C_{\kappa 1}$ allelic genes, maximum divergence is observed in the *J1* and *J2* protein coding sequences while the nonexpressed *J3*, *J4*, and *J5* segments are highly conserved.

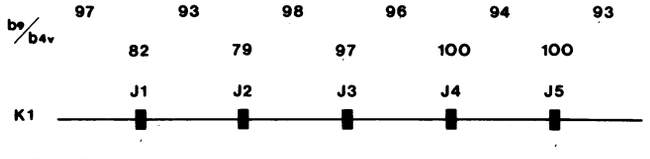

FIG. 5. Percent homologies were calculated separately for each J_κ coding segment (*Lower*) and intersegment (*Upper*) between the *b4* and *b9* alleles by 100 × (number of homologous bases/number of bases compared). Each gap is scored as a single difference.

DISCUSSION

New Evidence for Nonreciprocal Intergenic Conversion of the *K1* by the *K2* Gene. The *b4v J1* segment and a major part of the noncoding *J1-J2* intersegment (up to the nucleotide position 406) present a high degree of homology with the corresponding *K2* sequence. In particular, the *K2* and *b4v J1* segments only differ by one silent substitution (see Fig. 4). These findings constitute strong evidence for nonreciprocal conversion of the *b4 K1* by the *K2* gene. Regions of extensive homology between the two nonallelic sequences are also found in the *J-*C_κ intron. Compared to the *b4* intron, the *b9* and *K2* introns have exactly the same 160-bp deletion of a segment rich in adenosine and thymidine nucleotides located 1.1 kb upstream of the C_κ region (see Fig. 1). The identity of the two nucleotide sequences extends 30 nucleotides to the 5′ and 3′ side of the deleted region. This finding suggests that sequences can be deleted through conversion events. In

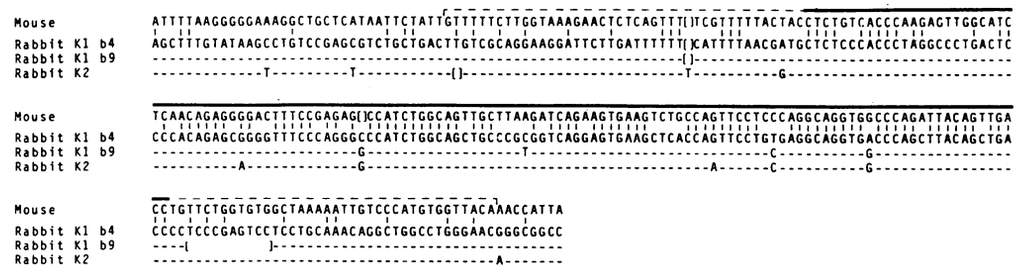

FIG. 6. Nucleotide sequence of the enhancer region of the *J-C*$_\kappa$ intron. The mouse and *b4* rabbit nucleotide sequence comparison is taken from Emorine *et al.* (11); vertical bars represent nucleotide identities between these two sequences. The *b9* and *K2* sequences have been aligned with these sequences. The sequence below the thick line corresponds to the KICR as defined by Emorine *et al.* (11). The interrupted line limits the region that contains the mouse κ enhancer described by Queen and Stafford (15).

addition, it can be noted that the most conserved region of the *b* alleles (from the nucleotide position 578 to 1156) correlates with the position of the segment deleted in the $J_{\kappa 2}$ locus. These last data are further evidence for a conversion of the *K1* by the *K2* locus.

The present data extend our previous studies on complex alleles of the rabbit immunoglobulin gene family. Based on structural analysis, the C_κ and J_κ germ-line genes present the following comparable characteristics of evolution: (*i*) homogeneity of the nonfunctional regions and (*ii*) sequence homology in limited regions of the *K1* locus with the *K2* locus, presumably introduced by intergenic conversion, increasing the diversity in the *K* alleles. The outcome is that heterozygous rabbits for the *K1* locus have a larger combinatorial potential than rabbits homozygous for this locus.

What Regulates the Preferential Expression of the *b* Alleles? Structural analysis of *cis* elements involved in κ chain expression have shown that there are two functional J_κ segments in the *b9 K1* locus that theoretically give a recombination potential twice as great as for the *b4* allele (one functional J_κ gene). In *b* heterozygous rabbits, one could, therefore, expect preferential expression of the *b9* allele. Nevertheless, heterozygous rabbits produce four times as many *b4* as *b9 K* chains (6). Wood and Coleclough (23) have shown that the frequency with which a *J* segment is used correlates with the proximity of the first dinucleotide TG encountered in the mouse and human J_κ and mouse J_λ segments to the consensus heptamer. According to this model, the *b9 J1* and *J2* segments would be used at the same frequency since the first TG appears within the three bases following the heptamer. They are even closer to the heptamer than the TG of the *b4 J2* segment. Therefore, and at least in this case, the difference in expression observed between the *b* alleles cannot be explained by the proximity of the first TG dinucleotide. Nevertheless, in *b9* and *b4*, it is interesting to note that in the defective *J3*, *J4*, and *J5* segments, the first TG dinucleotide is located the furthest away from the heptamer.

Elements involved in the κ chain transcription activation have been identified in the *J-C* intron of immunoglobulin genes (24–27). In Fig. 6, we have aligned the rabbit *b4*, *b9*, and *K2* KICR and their surrounding regions together with the 200-bp sequence containing the mouse activating element defined by Queen and Stafford (15). The *b4* and *b9* sequences differ by four substitutions within the KICR and by a deletion of 10 bp in *b9* located at the end of the KICR but still within the region corresponding to the mouse activating element. However, two of the *b9* substitutions are identical with the *K2* sequence and increase the homology with the mouse sequence. Whether these differences between the two allelic nucleotide sequences result in modification of the transcription level cannot be resolved by the structural analysis. Studies on the cellular expression of plasmid constructions with the different *b* enhancers should provide information concerning their relative efficiency for transcriptional activation.

We would like to thank M. Goodhardt for critical reading of the manuscript and discerning advice and to C. Fabre for typing the manuscript. We are also grateful to J. Igolen (Unité de Chimie Organique, Institut Pasteur, Paris) for providing the synthetic primer used in DNA sequencing. B.M. was the recipient of a long-term European Molecular Biology Organization Fellowship. This work was supported by grants from the Centre National de la Recherche Scientifique (ATP 955199), the Fondation pour la Recherche Médicale Française.

1. Benammar, A. & Cazenave, P. A. (1982) *J. Exp. Med.* **156,** 585–595.
2. Heidmann, O. & Rougeon, F. (1983) *EMBO J.* **2,** 437–441.
3. Dubiski, S. (1967) *Nature (London)* **214,** 1365.
4. Chou, C. T., Cinader, B. & Dubiski, S. (1974) *Cell. Immunol.* **11,** 304–313.
5. Kindt, T. J. (1975) *Adv. Immunol.* **21,** 35–86.
6. Mage, R. G. (1975) *Transplant. Rev.* **27,** 84–99.
7. Akimenko, M. A., Heidmann, O. & Rougeon, F. (1984) *Nucleic Acids Res.* **12,** 4691–4701.
8. Heidmann, O. & Rougeon, F. (1983) *Cell* **34,** 767–777.
9. Emorine, L., Dreher, K., Kindt, T. J. & Max, E. E. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5709–5713.
10. Heidmann, O. & Rougeon, F. (1984) *Nature (London)* **311,** 74–76.
11. Emorine, L., Kuehl, M., Weir, L., Leder, P. & Max, E. E. (1983) *Nature (London)* **304,** 447–449.
12. Parslow, T. G. & Granner, D. K. (1982) *Nature (London)* **299,** 449–451.
13. Chung, S.-Y., Folsom, V. & Wooley, J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 2427–2431.
14. Parslow, T. G. & Granner, D. K. (1983) *Nucleic Acids Res.* **11,** 4775–4792.
15. Queen, C. & Stafford, J. (1984) *Mol. Cell. Biol.* **4,** 1042–1049.
16. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
17. Messing, J. & Vieira, J. (1982) *Gene* **19,** 269–276.
18. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. (1979) *Nature (London)* **280,** 288–294.
19. Hieter, P. A., Maizel, J. V. & Leder, P. (1982) *J. Biol. Chem.* **257,** 1516–1522.
20. Max, E. E., Seidman, J. G. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3450–3454.
21. McCartney-Francis, N., Skurla, R. M., Mage, R. G. & Bernstein, K. E. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1794–1798.
22. Emorine, L. & Max, E. E. (1983) *Nucleic Acids Res.* **11,** 8877–8890.
23. Wood, D. L. & Coleclough, C. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 4756–4760.
24. Banerji, J., Olson, L. & Schaffner, W. (1983) *Cell* **33,** 729–740.
25. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. (1983) *Cell* **33,** 717–728.
26. Queen, C. & Baltimore, D. (1983) *Cell* **33,** 741–748.
27. Picard, D. & Schaffner, W. (1984) *Nature (London)* **307,** 80–82.