Some sequence similarities among cloned mouse DNA segments that code for λ and κ light chains of immunoglobulins

(minigenes/insertional mechanism/recognition signals)

TAI TE WU*, ELVIN A. KABAT†, AND HOWARD BILOFSKY‡

*Departments of Biochemistry and Molecular Biology, and Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois 60201;
†The National Cancer Institute, Bethesda, Maryland 20205, and Departments of Microbiology, Human Genetics and Development, and Neurology, and The
Cancer Center/Institute for Cancer Research, College of Physicians and Surgeons, Columbia University, New York, New York 10032; and ‡Bolt, Beranek, and
Newman, Inc., Cambridge, Massachusetts 02138

ABSTRACT A comparison between the cloned mouse DNA segments that were found to code for the λ and κ light chains of immunoglobulins established that there were seven short nucleotide sequences, two of which matched 6 out of 7, two 7 out of 8, and one 9 out of 10 bases; these sequences were located either at homologous amino acid positions or at positions displaced by four amino acids or less. They all occurred in the framework regions (FRs), five next to the complementarity-determining regions (CDRs). Three of these were unique and did not occur elsewhere in the immunoglobulin nucleotides sequenced thus far or in DNAs of phage φX174, phage G4, or simian virus 40. Five could serve as sites of joining by recombination or insertion of CDR to FR segments, and the invariant tryptophan that is the first residue of the second FR might serve as a sixth. These sites are consistent with the minigene or insertional hypotheses for the generation of antibody diversity but could also serve as points of recognition for a mutator enzyme or could serve to limit somatic mutation to the CDRs.

The vast amount of experimental data on amino acid sequences of immunoglobulin variable regions (V regions) (1) has provided a basis for trying to understand antibody diversity. From a statistical analysis of such sequences (2, 3) as they were being accumulated, the three complementarity-determining (hypervariable) regions or segments (CDRs) were recognized in the light (L) and in the heavy (H) chains and predicted to form the walls of the antibody combining site; this has been verified by x-ray crystallography in four different laboratories (refs. 4–7; see ref. 8). As a consequence of this analysis each immunoglobulin chain was divided into four framework regions (FRs) separated by the three CDRs as follows: FR1, CDR1, FR2, CDR2, FR3, CDR3, and FR4. It was recognized that a simple mechanism for assembling the chains could be the insertion of nucleotides coding for the CDRs into those coding for the FR segments to assemble a complete V-region gene (2, 3). When the various FR segments for each chain were arranged into sets (9), each set being made up of segments identical in sequence, it was noted that the members of a given FR1 set could be associated with different FR2 sets, FR3 sets, and FR4 sets. This independent assortment of FR sets led to the hypothesis (9) that the V-region gene is assembled from sets of germ line minigenes for the FR segments and inferentially from sets of minigenes for the CDRs. A minigene was defined as a DNA segment coding for a portion of the V region that shows some evidence of segregation as a functional unit independent of the rest of the DNA coding for the V region (10). It was further hypothesized that the minigenes were assembled somatically during differentiation. Because assortment was demonstrated only for FR segments, it would be independent of whether one or two

CDR residues assorted with a given FR. The CDRs could not be examined for independent assortment because there were too few that were identical.

Studies by Tonegawa and coworkers (11–13) and by Seidman *et al.* (14, 15) have led to the isolation from 12-day mouse embryo of cloned DNA gene segments coding for V_κ and V_λ chains from FR1 almost to the end of CDR3.

Bernard *et al.* (13) subsequently isolated from mouse embryo DNA a clone (Ig99λ) coding for the last two residues of CDR3 plus FR4, which they termed the J segment, followed by an intervening sequence of about 1.2 kilobase pair and then by the DNA coding for the constant (C) region of mouse V_λ; they also isolated another clone from adult myeloma DNA (Ig25λ) coding for the entire V region. Seidman *et al.* (16) have obtained similar results with clones coding for V_κ DNA; in embryo DNA, a nucleotide sequence for MOPC41 myeloma coded for amino acids 1–95 and another coded for amino acids 96 to the end of the V region (FR4 plus two residues of CDR3), whereas in adult myeloma DNA, the entire V region was assembled. Thus the J segments (FR4 plus two residues of CDR3) are minigenes whose DNA is joined to that coding for the rest of the V region between the twelfth day of embryonic life and the adult myeloma, confirming our postulated somatic assembly for this segment (9).

Several groups (17, 18) have used amino acid sequences of mouse V_κ and V_H chains to show assortment of a J segment (two or three residues of CDR3 plus all of FR4), and Max *et al.* (19) have shown by determining the nucleotide sequence of the 3.7-kilobase pair intervening sequence of MOPC41 the presence of five J segments each encoding a different FR4 plus two residues of CDR3.

A major question remains as to whether the minigene hypothesis (9) holds for the rest of the V region. A detailed analysis of sequence data and the recognition of an FR2 that was identical in 1 human, 20 mice, and 13 rabbits (10) was interpreted as supporting this hypothesis.

We have aligned the nucleotide sequences for the V_λ (11–13) and V_κ (14–16) clones lacking the J segment to correspond to the amino acid sequences and have looked for identical sequences of six or more nucleotides with possibly one mismatch; seven such sequences were found. Three of these nucleotide segments were unique to V regions and were not found in the coding or noncoding strands of phage φX174 (20), simian virus 40 (SV40) (21, 22), phage G4 (23), or elsewhere in the sequences of immunoglobulin clones (11–16, 24). Five occurred in positions such that they could serve as points of joining by recombination or insertion of the appropriate CDR and the FR; the

Abbreviations: CDR, complementarity-determining region; FR, framework region; V, variable; C, constant; H, heavy; L, light; SV40, simian virus 40; HH, hyperhomologous; MOPC, plasmacytomas induced in BALB/c mice with paraffin oil (43).

invariant Trp, for which only one codon exists and which is the first residue of FR2, could serve as a sixth recognition signal. Thus V_κ and V_λ frameworks contain the same nucleotide sequences at the proper locations for joining all of the FR and CDR segments up to and including the J minigene.

Tonegawa *et al.* (11) tabulated homologies of portions of the nucleotide sequences of the bottom strand of CDR1 and the top strands of CDR2 and CDR3 in their V_λII gene, and Ben-Sasson (25) has extended these similarities in nucleotide sequences to two other regions in V_λII that he terms hyperhomologous (HH). The fourth region related to V–C joining and the fifth was in the intervening sequence 32 nucleotides away from it. He also lists a different set of five HH sequences from the data on V_κ (14, 15). Ben-Sasson points out the similarity of the V_λ HH nucleotide segments to prokaryote recombination site sequences IS1 and IS2 (26, 27) and those of V_κ HH to *att* (28) and hypothesizes that the HH sequences are sites of extensive recombination that, together with an error-prone repair system (refs. 29, 30; compare refs. 31–34), would generate diversity in the CDR somatically after an initial V–C joining step. Our findings, if the constancy of these nucleotide segments is maintained as

further clones are sequenced, could also provide recognition sites for mutator enzyme attachment.

MATERIALS AND METHODS

We have aligned three mouse λ L chain DNA sequences—Ig 13λ (11), Ig 99λ, and Ig 303λ (13)—and five mouse κ sequences—K2, K3, the one deduced from the partial mRNA sequence of MOPC149 (14, 15), and the cloned embryo and adult DNA sequences of MOPC41 (16)—together with amino acid positions (Fig. 1). The nucleotide sequences for Ig 13λ and Ig 99λ correspond to the L chain V region sequence up to amino acid residue 96, (amino acid numbering as in ref. 1) together with 24 bases beyond position 96. That for Ig 303λ corresponds to amino acid positions 1 through 95. Those for K2 and K3 correspond to amino acid positions 1 through 97, together with 21 bases beyond position 97. The DNA sequence deduced from the partial mRNA sequence of MOPC149 corresponds to amino acid positions 46 to 89. The DNA sequence for cloned embryo MOPC41 corresponds to amino acid positions 1 to 95 followed by an intervening sequence and that for the cloned adult MOPC41 corresponds to amino acid positions 1 to 95. The

FIG. 1. Alignment of cloned mouse λ and κ L chain DNA sequences. The sequence designated by λ represents the sequences of Ig13λ (11) and Ig99λ (12), which code for the V region minus the J segment (or FR4 with one or two amino acid residues of CDR3), together with 24 nucleotides beyond amino acid position 96. The base differences among them are indicated below the line. In addition, the sequence of Ig303λ (13) is also included up to amino acid position 95. The sequence designated by κ represents the sequences of K2 and K3 (14, 15), which code for the V region minus FR4 together with 21 nucleotides beyond amino acid position 97, the corresponding DNA sequence for the partial mRNA sequence of MOPC149 from amino acid position 46 to 89 (14, 15), the sequence of the embryonic MOPC41 (16), which codes for the V region minus the J segment together with 27 nucleotides beyond amino acid position 95, and the sequence of adult MOPC41 (16) up to the first and second nucleotides coding for Pro at position 95. The corresponding amino acid positions are aligned. Gaps used for alignment are given by dashes. Short nucleotide sequences that are identical and similar in the three λ and three κ sequences are underlined. FR1 consists of amino acid residues 1–23; FR2, 35–49 (ref. 1, but see text); and FR3, 57–88.

maximum length of these sequences was 318 bases. They were aligned according to amino acid positions (1), (Fig. 1), and a search for identical nucleotides was conducted among all eight sequences. These sequences were also misaligned by five amino acid positions or less to see whether additional identical nucleotides could be found. If such nucleotides occurred in clusters, they might be of genetic importance. Because in φX174 each frame may call for a different function (20), it is conceivable that frames other than the coding frame might serve other functions such as recognition signals, etc.

We further used known sequences of φX174 (20), SV40 (21, 22), and G4 (23) for comparison. Because these DNAs would presumably be totally unrelated to immunoglobulin genes, the presence of similar or identical stretches of nucleotide segments in these viral DNAs and in cloned immunoglobulin genes might suggest lack of specificity or, conceivably, depending on their locations, some related function.

RESULTS

As shown in Fig. 1, there are seven short nucleotide sequences that match at least six bases with possibly one mismatch among the three mouse λ and five κ light chain DNA sequences and are located either at homologous amino acid positions or off by four amino acid positions or less. Their corresponding amino acid sequences are also listed except for the last two pairs, which are located either at the junction of CDR3 and the intervening DNA sequences or beyond the coding DNA segments (Fig. 1). These segments are listed in Table 1.

Sequence 1 is identical for the three mouse λ chains and three of the five mouse κ chains, with the fifth nucleotide being T, while the adult and embryonic DNA sequences of MOPC41 have a C instead. They are all in register with the amino acid positions, and begin at the second nucleotide for Val of λ chains and Met of κ chains at position 4 (Fig. 1). Sequence 2 is also in register between the λ and κ chains; it begins at the second nucleotide for Leu of λ and Ile of κ at position 21. The fifth nucleotides are T for the three mouse λ chains and the adult and embryonic sequences of MOPC41, whereas the remaining three κ chains have an A instead. Sequence 3 is out of register by four nucleotides. It is identical for the three mouse λ chains and for κ chains K2, K3, and MOPC149, with the third nucleotide being G, whereas the adult and embryonic sequences of MOPC41 have an A instead. In λ it begins at the third nucleotide for Thr at position 45, and in κ at the second nucleotide for Leu at po-

sition 47. Sequences 4 and 5 are again in register. Sequence 4 begins at the first nucleotide for Gly at position 57, and the sixth nucleotides are different. Sequence 5 begins at the first nucleotide for Tyr at position 86, and the fifth nucleotides are different. Sequences 6 and 7 are out of register, the former by 1 and the latter by 11 nucleotides. Sequence 6 is located at the junction of CDR3 and the intervening noncoding region, and the sixth nucleotides are different. Sequence 7 is identical for the three mouse λ chains and for κ chains K2, K3, and MOPC149, with the third nucleotide being G, whereas the adult and embryonic sequences of MOPC41 have an A instead; it is in the intervening noncoding region beyond CDR3.

Three of the seven sequences—2, 5, and 6 (Table 1)—were not found anywhere else in the cloned mouse DNA sequences, including intervening noncoding sequences and those coding for the J segment and C regions of L (12) and H (24) chains of immunoglobulins or in φX174 (20), SV40 (21, 22), and G4 (23). They are also the longer sequences. Sequences 1 and 4 are shorter, and similar sequences were found elsewhere. A sequence similar to no. 1, TGACGCAG, with the fifth nucleotide being G, was found twice in φX174 (20) at positions 18 and 797. At both places, the sequence was used to code for two different proteins: proteins A and B at position 18, and proteins D and E at position 797. In the A protein, it consisted of the second and third nucleotides coding for Met and the codons for Thr and Gln; in the B protein, the third nucleotide coding for His, the codons for Asp and Ala, and the first nucleotide coding for Glu; in the D protein, the third nucleotide coding for Thr, the codons for Asp and Ala, and the first nucleotide coding for Glu; and in the E protein, the second and third nucleotides coding for Leu and the codons for Thr and Gln. A sequence identical to no. 4 of K2, K3, and MOPC149, GGTGTGCC, was found in SV40 (21, 22) on the complementary strand at position 1214, which is noncoding.

The two remaining sequences, 3 and 7, are the shortest (Table 1), and they were found elsewhere 8 and 10 times, respectively. However, if the third nucleotide is required to be a purine—i.e., G or A—they were found elsewhere only 4 and 3 times respectively. A sequence identical to no. 3 of λ, K2, K3, and MOPC149, TGGTCTA, was found in φX174 at position 2833 (20) and G4 at position 4461 (23). It consisted of the codons for Trp and Ser and the first nucleotide coding for Asn located near the COOH terminus of the G protein. A sequence identical to no. 3 of the adult and embryonic MOPC41, TGATCTA, was

---

Table 1. Similar short nucleotide segment pairs among the eight mouse λ and κ L chain DNA sequences

Pair no.	λ DNA sequences	Amino acid positions	κ DNA sequences	Amino acid positions	Matches between λ and κ sequences	Number of similar sequences found elsewhere	Other nucleotide arrangements that could give same amino acid sequences	
							λ	κ
1	TGACTCAG	4–6	TGACTCAG C	4–6	7/8 or 8/8	2	8	2
2	TCACTTGTCG	21–24	TCACATGTCG T	21–24	9/10 or 10/10	0	8	6
3	TGGTCTA	45–47	TGGTCTA A	47–49	6/7 or 7/7	4 or 8	96	16
4	GGTGTTCC	57–59	GGTGTGCC C	57–59	7/8	1	4	4
5	TATTTCTGT	86–88	TATTACTGT	86–88	8/9	0	8	8
6*	CCACAATGA	96–99	CCACAGTGA†	96–99	8/9	0	?	?
7*	ATGACAT	98–100	ATGACAT A	102–104	6/7 or 7/7	3 or 10	?	?

Identical bases are underlined.

* These sequences were found either at the junction of CDR3 and the intervening DNA sequences or in the intervening DNA sequences beyond CDR3.

† Seidman *et al.* (16) have noted that a portion of this sequence CACAGTG can form an inverted repeat.

found on the complementary strand of Igl3λ (11) and Ig99λ (13), located 33 nucleotides beyond amino acid position 96, in the intervening noncoding region. A sequence similar to no. 3, TGCTCTA, with the third nucleotide being C, was found in SV40 (21, 22) at position 1403, which probably codes for both the VP2 and VP3 proteins; and in the coding regions of Igl3λ (11), Ig99λ, and Ig303λ (13), consisting of the third nucleotide coding for Cys 88 and codons for Ala 89 and Leu 90.

A sequence identical to no. 7 of λ, K2, K3, and MOPC149—ATGACAT—was found in SV40 (21, 22) at position 988, which probably codes for both the VP2 and VP3 proteins, and in the intervening noncoding segment 58 nucleotides in front of the sequence coding for the variable region of K3 or 12 nucleotides in front of the sequence coding for the precursor segment of K3 (14). A sequence identical to no. 7 of the adult and embryonic MOPC41—ATAACAT—was found in SV40 (21, 22) on the complementary strand at position 4161 in the main body coding for the T antigen. A sequence similar to no. 7, ATTACAT, with the third nucleotide being T, was found in G4 (23) at position 3205, consisting of the third nucleotide coding for Lys and codons for Leu and His in the F protein; and at 5092, consisting of codons for Ile and Thr and the first nucleotide for Ser in the H protein. Another sequence similar to no. 7, ATCACAT, with the third nucleotide being C, was found in SV40 (21, 22) at position 3889, which is noncoding, in the coding regions of K2 and K3 (14) consisting of codons for Ile-21 and Thr-22 and the first nucleotide for Cys-23, at the beginning of the J3 segment (19) consisting of codons for Ile and Thr and the first nucleotide for Phe, and in the intervening noncoding segment 22 nucleotides in front of the sequence coding for mouse CH1 (24).

Fig. 1 also shows that sequence 1 is located in FR1, and 2 at the end of FR1 with two nucleotides in CDR1. Sequence 3 is in FR2 near the junction of FR2 and CDR2. Sequence 4 is at the junction of CDR2 and FR3, being in FR3. Sequence 5 is in FR3 at the junction of FR3 and CDR3. Sequence 6 is at the junction of CDR3 and the intervening noncoding region, and two nucleotides beyond the point of joining of the J segment in both the mouse λ (13) and κ (19) chains. Sequence 7 is in the intervening sequence beyond CDR3.

In an attempt to evaluate further the significance of these nucleotide sequences, we have examined all of the alternative amino acid sequences of human, mouse, rabbit, and other species of κ and λ light chains available (1, 35) corresponding to nucleotide segments 1 through 5; these were converted into nucleotide sequences. In most instances, especially for those amino acid sequences occurring in many chains, these translated sequences gave choices that would maintain the integrity of nucleotide segments 1 through 5. In instances in which a given amino acid sequence was reported only once, the matching of nucleotide segment 1 was reduced from 7/8 to 6/8. Nucleotide segment 2, which coded through the first two nucleotides of Arg-24, was shortened in some instances so that the matching sequence terminated with Cys-23 and matches of nucleotide sequences from positions 21–23 were 8/8, 7/8, and 6/8; one rare instance consisting of a single sequence matched at only 5/8 nucleotides, but if the two nucleotides of Arg-22 were included it matched 7/10. The alternative amino acid sequences did not reduce the matching for nucleotide segment 3, which remained at 6/7 or 7/7; those for segments 4 and 5 were maintained at 7/8 and 8/9, respectively, except for rare sequences for which only one occurrence had been reported for which some 6/8 and 7/9 nucleotides were identical. This comparison could not be made for nucleotide segments 6 and 7, which were in the intervening sequences beyond the coding segment for CDR3 in clones from 12-day-old mouse embryo DNA.

DISCUSSION

The finding of seven nucleotide sequences that are identical in the Vλ and Vκ clones is of considerable interest. The λ and κ chains are unlinked and may even occur on different chromosomes (36). Thus the identity of these nucleotide sequences and their locations, if not a result of chance, could be of substantial significance. Five of these sequences—nos. 2, 3, 4, 5, and 6—were positioned so that they could serve as recognition sequences for the joining by recombination or insertion of CDR and FR nucleotide segments required by the minigene hypothesis (9, 10). Sequences 5 and 6 were unique in that they did not occur in other immunoglobulin V- and C-coding regions or in intervening sequences or in φX174, G4, and SV40. Sequences 1, 2, 4, 5, and 6 correspond to the amino acid sequence but may be out of register with the codon, whereas sequence 3 was associated with amino acid residues 45–47 in Vλ and 47–49 in Vκ. This displacement may also prove very significant in terms of antibody structure. Although CDR2 is considered to be residues 50–56 from the aligned sequences, this assignment is based on variability plots (1–3) on all L chain sequences. Mouse Vλ chains are unusually restricted; indeed 12 were shown to have the identical sequence throughout the V region. The remaining chains differed only at residues in the three CDRs except for position 48 (refs. 37–39; compare ref. 8). They all have an insertion of three residues in CDR1 (1). Position 48 has been found not to be invariant in mouse Vλ (refs. 37–39; compare ref. 1) and it has been noted (8) that position 48 might, because of the insertion, be part of CDR2 in mouse Vλ. Thus the finding that the identical nucleotide sequence in mouse Vκ and Vλ corresponds exactly to where it should be if CDR2 of Vλ consists of residues 48–56 and that of the other V regions consists of 50–56. Sequence 4 is again back in register and is suitably positioned for joining by recombination or insertion of CDR2 to FR3 in both Vκ and Vλ. If this is a coincidence it is an extraordinary one. It would be extremely important to have a high-resolution x-ray crystallographic study of a mouse Vλ, preferably a VλI, to ascertain whether CDR2 consists of residues 48–56.

It might be argued that the amino acid sequence corresponding to these nucleotides in Vλ and Vκ was such as to dictate the identity of the sequences. This is not so, however, because the degeneracy of the code is such that other choices of codons could provide the same sequence. Indeed, when the identical sequence begins at a second or third codon additional choices may become possible. The numbers of possibilities for preserving the amino acid sequence without maintaining the nucleotide sequence in Vκ and Vλ chains are given in the last two columns of Table 1. It is clear that finding the identical nucleotide sequences in these segments of Vκ and Vλ chains is not determined by the amino acid sequence. Most strikingly, the chance of sequence 3 being completely identical in three of the five Vκ and all three Vλ is extremely low, especially in view of their different locations and amino acid sequences. This makes even more significant the possibility that each could serve to join FR2 and CDR2.

Sequence 5 is again back in register. Sequences 6 and 7 are in the intervening sequences beyond residue 95 of CDR3, and 6 could serve as a site for the joining of CDR3 to J.

The significance of sequence 1 cannot be evaluated on the basis of the data available.

Even the limited nucleotide sequences available have thus far provided recognition signals that could serve exactly for the joining by recombination or insertion of the 5′ end of CDR1 to FR1, of the 5′ and 3′ ends of CDR2 to FR2 and FR3, respectively, and of the 5′ end of CDR3 to FR3—all exactly as required by the minigene hypothesis. Sequence 6 in the intervening sequence also occurs in proximity to the site of joining

of the 5' end of the J minigene to the 3' end of the CDR3 sequence (16–18). It will be important to see if additional DNA clones will show whether or not the same nucleotides coding in V_κ and V_λ for the FR segments that adjoin the CDRs will be preserved. It is of interest that these nucleotide sequences differ from one another and from those at the junctions of the C_H and hinge domains (24) and the L chain precursor (11, 13–16).

The site of joining of CDR1 to FR2 has not emerged from the existing limited data unless it were to be just the invariant Trp. In the original formulation (2) of the insertion hypothesis it was pointed out that Trp might be a recognition site for joining of CDR1 to FR2. Because there is but one codon for Trp and because splicing has been shown to occur within a single codon (40), it could serve as this recognition sequence. Invariant Trp is the first residue of FR2 in all V_L and V_H chains (35).

The nucleotide segments described in the present study, while consistent with the insertional (2, 3) or minigene (9, 10) hypotheses, are also compatible with somatic mutation hypotheses (25, 29, 30, 37) in that they could provide recognition sites for mutator enzyme attachment or boundaries limiting somatic mutation to the CDRs. Analysis (41) of the assortment data (9, 10) suggests somatic mutation is not important in the framework of immunoglobulin chains (cf. ref. 42).

If additional κ and λ clones and further studies establish that these nucleotide sequences are recognition sequences, the degeneracy of the code as viewed from the standpoint of the translated amino acid sequences will be but an apparent degeneracy. Indeed, the choices of amino acids permitted by the code will essentially be dictated by the necessity to incorporate other genetic information or instructions within a given coding sequence while permitting the amino acid chain to assume the proper three-dimensional structure.

1. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1976) *Variable Regions of Immunoglobulin Chains* (Bolt, Beranek & Newman, Cambridge, MA).
2. Wu, T. T. & Kabat, E. A. (1970) *J. Exp. Med.* 132, 211–250.
3. Kabat, E. A. & Wu, T. T. (1971) *Ann. N.Y. Acad. Sci.* 190, 382–393.
4. Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P. & Saul, F. (1973) *Proc. Natl. Acad. Sci. USA* 70, 3305–3310.
5. Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974) *Proc. Natl. Acad. Sci. USA* 71, 4298–4302.
6. Epp, O., Colman, P., Fehlhammer, H., Bode, W., Schiffer, M., Huber, R. & Palm, W. (1974) *Eur. J. Biochem.* 45, 513–524.
7. Edmundson, A. B., Ely, K. R., Girling, R. L., Abola, E. E., Schiffer, M., Westholm, M., Fausch, M. D. & Deutsch, H. F. (1974) *Biochemistry* 13, 3816–3827.
8. Kabat, E. A. (1978) *Adv. Protein Chem.* 32, 1–75.
9. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1978) *Proc. Natl. Acad. Sci. USA* 75, 2429–2433.
10. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1979) *J. Exp. Med.* 149, 1299–1313.
11. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, A. (1978) *Proc. Natl. Acad. Sci. USA* 75, 1485–1489.
12. Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. (1978) *Cell* 15, 1–14.
13. Bernard, O., Hozumi, N. & Tonegawa, S. (1978) *Cell* 15, 1133–1144.
14. Seidman, J. G., Leder, A., Edgell, M. H., Polsky, F., Tilghman, S. M., Tiemeier, D. C. & Leder, P. (1978) *Proc. Natl. Acad. Sci. USA* 75, 3881–3885.
15. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. (1978) *Science* 202, 11–17.
16. Seidman, J. G., Max, E. E. & Leder, P. (1979) *Nature (London)* 280, 370–375.
17. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. & Hood, L. (1978) *Nature (London)* 276, 785–790.
18. Rao, D. N., Rudikoff, S., Krutzsch, H. & Potter, M. (1979) *Proc. Natl. Acad. Sci. USA* 76, 2890–2894.
19. Max, E. E., Seidman, J. G. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3450–3454.
20. Sanger, F., Coulson, A. R., Friedmann, T., Air, G. M., Barrell, B. G., Brown, N. L., Fiddes, J. C., Hutchinson, C. A., III, Shocombe, P. M. & Smith, M. (1978) *J. Mol. Biol.* 125, 225–246.
21. Reddy, V. B., Thimmappay, B., Dhar, R., Subramanian, K. N., Zain, B. S., Pan, J., Ghosh, P. K., Celma, M. L. & Weissman, S. M. (1978) *Science* 200, 494–502.
22. Fiers, W., Contreras, R., Haegeman, G., Rogiers, R., Van de Voorde, A., Van Heuverswyn, H., Van Herreweghe, J., Volckaert, G. & Ysebaert, M. (1978) *Nature (London)* 273, 113–120.
23. Godson, G. N., Barrell, B. G., Staden, R. & Fiddes, J. C. (1978) *Nature (London)* 276, 236–247.
24. Sakano, H., Rogers, J. H., Hüppi, K., Brock, C., Traunecker, A., Maki, R., Wall, R. & Tonegawa, S. (1979) *Nature (London)* 277, 627–633.
25. Ben-Sasson, S. A. (1979) *Proc. Natl. Acad. Sci. USA* 76, 4598–4602.
26. Ohtsubo, H. & Ohtsubo, E. (1978) *Proc. Natl. Acad. Sci. USA* 75, 615–619.
27. Musso, R. E. & Rosenberg, M. (1977) in *DNA Insertion Elements, Plasmids and Episomes*, eds. Bukhari, A. I., Shapiro, J. & Adhya, A. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 597–600.
28. Landy, A. & Ross, W. (1977) *Science* 197, 1147–1160.
29. Brenner, S. & Milstein, C. (1966) *Nature (London)* 211, 242–243.
30. Baltimore, D. (1974) *Nature (London)* 248, 409–411.
31. Witkin, E. M. (1976) *Bacteriol. Rev.* 40, 869–907.
32. Gillin, F. D. & Nossal, N. G. (1976) *J. Biol. Chem.* 251, 5219–5224; 5225–5232.
33. Laval, J. (1978) *Biochimie* 60, 1123–1134.
34. Devoret, R. (1978) *Biochimie* 60, 1135–1140.
35. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1979) *Sequences of Immunoglobulin Chains, Medical Computer Systems* (Bolt, Beranek & Newman, Cambridge, MA), in press.
36. Valbuena, O., Marcu, K. B., Croce, C. M., Huebner, K., Weigert, M. & Perry, R. P. (1978) *Proc. Natl. Acad. Sci. USA* 75, 2883–2887.
37. Cohn, M., Blomberg, B., Geckeler, W., Raschke, W., Riblet, R. & Weigert, M. (1974) in *The Immune System*, eds. Sercarz, E. E., Williamson, A. R. & Fox, C. F. (Academic, New York), pp. 89–117.
38. Appella, E. (1971) *Proc. Natl. Acad. Sci. USA* 68, 590–594.
39. Cesari, J. M. & Weigert, M. (1973) *Proc. Natl. Acad. Sci. USA* 70, 2112–2116.
40. Crick, F. (1979) *Science* 124, 264–271.
41. Kabat, E. A. (1979) in *Cells of Immunoglobulin Synthesis*, eds. Pernis, B. & Vogel, H. J. (Academic, New York), pp. 33–44.
42. Ohta, T. (1978) *Proc. Natl. Acad. Sci. USA* 75, 5108–5112.
43. Potter, M. (1972) *Physiol. Rev.* 52, 631–719.