Reprinted from Nature, Vol. 302, No. 5909, pp. 575-581, 14 April 1983

Somatic generation of antibody diversity

Susumu Tonegawa

Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

In the genome of a germ-line cell, the genetic information for an immunoglobulin polypeptide chain is contained in multiple gene segments scattered along a chromosome. During the development of bone marrow-derived lymphocytes, these gene segments are assembled by recombination which leads to the formation of a complete gene. In addition, mutations are somatically introduced at a high rate into the amino-terminal region. Both somatic recombination and mutation contribute greatly to an increase in the diversity of antibody synthesized by a single organism.

A CENTRAL part in the immune system is played by antibodies which can recognize and distinguish specific molecular patterns of antigens. Because antigens are diverse in structure, the repertoire of antibodies must be large and is estimated to be at least one million and probably one to two orders of magnitude higher[1-4]. The way in which the genetic information required to code for this exceedingly large number of different but related proteins is stored in the genome and transmitted through generations has been the central issue of immunogenetics during the past few decades.

The basic unit of an antibody, or immunoglobulin (Ig), molecule is composed of two identical light (L) chains and two identical heavy (H) chains (Fig. 1). Extensive sequence analysis of many immunoglobulin chains revealed that the amino-terminal regions are highly variable while the carboxy-terminal regions are one of several types within a given species[5,6]. The variable (V) region is primarily responsible for antigen recognition and contains particularly variable subregions whose residues have been implicated in actual antigen contact (ref. 7; also see ref. 8 for a review). These subregions, which are referred to as complementarity-determining regions (CDR1, CDR2 and CDR3), are flanked by less variable subregions termed framework regions (FR1, FR2, FR3 and FR4). The constant (C) regions are responsible for a variety of effector functions such as complement fixation and Fc receptor binding and define the chain types or classes (Fig. 1).

The idea that the V and C regions of an immunoglobulin chain are encoded by two separate 'genes' in germ-line cells which undergo a rearrangement for joining in lymphocyte development was first suggested by Dreyer and Bennet[9]. This hypothesis was an attempt to accommodate two seemingly contradictory observations: (1) the extreme heterogeneity in the primary structure of the V regions within an individual animal and (2) the observation that an allotypic determinant (a serologically identifiable determinant commonly present on immunoglobulin chains of a particular type or class) segregates as a single mendelian trait[10]. The former result suggests that each individual carries a large number of structural genes for immunoglobulin chains while the latter result indicates that there may be only one gene for each of the several chain types or classes.

Thanks to restriction enzymes and recombinant DNA technology, the enigma of immunoglobulin genes has now been resolved at least in outline. It was found that an immunoglobulin polypeptide chain is encoded in multiple gene segments scattered along a chromosome of the germ-line genome[11-16]. These gene segments must be brought together to form a complete immunoglobulin gene active in B lymphocytes[11,16-24]. In addition, mutations are introduced somatically into an immunoglobulin gene at a high rate[17,19,25]. Both the recombination and mutation greatly diversify the genetic information carried in the germ-line genome. What follows is a brief review of these somatic events in mice. Recent studies on other mammalian systems have confirmed the unversality of the basic principles[26-29].

Organization of immunoglobulin gene sequences

Mouse immunoglobulin chains are encoded in three unlinked gene families: λ light-chain genes, κ light-chain genes and heavy-chain genes, residing on chromosomes 16 (ref. 30), 6 (refs 31, 32) and 12 (refs 31, 33, 34), respectively. In mice, λ chains form only about 5% of the total serum immunoglobulin light chains and are much less heterogeneous than κ chains or heavy chains. Thus, it is no coincidence that the first complete description of the sequence organization of an immunoglobin chain gene was made for the major λ chain subtype of mouse, namely λ1 chains (see Fig. 1 legend for the definition of λ

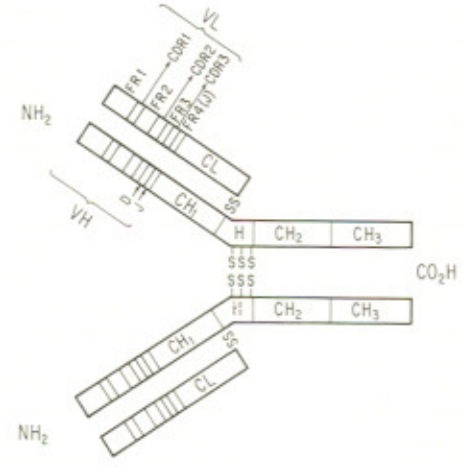

Fig. 1 Basic structure of a mouse IgG molecule. An immunoglobulin molecule is composed of two identical light chains and two identical heavy chains, each of which consists of the amino-terminal V region (V_L or V_H) and the carboxy-terminal C region (C_L or C_H). The V region contains amino acids primarily responsible for the antigen recognition which are enriched in the three complementarity-determining regions, CDR1, CDR2 and CDR3. The rest of the V regions are referred to as framework regions, FR1, FR2, FR3 and FR4. FR4 forms almost all of the J region, while the core portion of the heavy chain CDR3 is referred to as the D region. The C region executes a variety of effector functions. The light chains appear in two types, κ and λ. The mouse λ chains are further subdivided into three subtypes, λ1, λ2 and λ3. Each light-chain type or subtype is defined by a unique C region sequence. Mouse immunoglobulin molecules are conventionally grouped into several classes and subclasses defined by specific C_H region sequences. In the mouse, eight C_H regions, μ, δ, γ3, γ2a, γ1, γ2b, ϵ and α define as many classes or subclasses, IgM, IgD, etc. The C_H region is composed of three or four internal homology units C_{H1}, C_{H2}, etc. In addition, a cysteine-rich hinge (H) region resides between C_{H1} and C_{H2} of γ chains.

Fig. 2 Organization of various mouse immunoglobulin gene segments. The organization of various immunoglobulin gene segments before and after somatic rearrangement (to the left and right of the arrows, repectively) is shown. The rearranged state illustrated is only one of many possible cases in each gene family. The double slashes indicate that the distance between and the relative orientation of the two flanking sets of gene segment have not yet been determined. Within each set of gene segments connected by a solid line, the relative position of each gene segment has been determined and the orientation of all segments is identical reading 5′ to 3′ from left to right. The broken lines in the V_κ and V_H segment clusters indicate that the exact number and the relative orientation of the V-gene segments in each cluster are unknown. However, in a few cases studied, the relative positions of two or more adjacent V segments have been determined and their orientation has been shown to be identical (refs 55, 115 and G. Heinrich, unpublished). The broken line in the D-segment cluster and that between the C_μ- and C_δ-coding exons are only for illustrative purposes and the distance and the relative orientation of the involved gene segments have been determined. D_1 to D_{11} represent the D segments belonging to the SP2 and FL16 families (see ref. 53). The nine D segments of the SP2 family form a cluster spanning about 60 kb. The two D segments belonging to the FL16 family are located in the 5′ end region of the SP2-D cluster. The 12th D segment, D_{Q52}, has been mapped 696 bp 5′ to the J_{H1} segment[51]. The orientation of the SP2–FL16D cluster relative to and its distance from the V_H cluster or the D_{Q52}–J_H–C_H cluster are unknown. However, the fact that all of these D segments are deleted in many myelomas and Abelson virus-transformed B cells harbouring V–D–J rearrangements on both copies of chromosome 12 suggests that the D cluster is located between the V_H cluster and the D_{Q52}–J_H–C_H cluster. Likewise, it has been argued that the L_1V_1 segments are 5′ to the $J_3C_3J_1C_1$ unit and the L_2V_2 segments 5′ to the $J_2C_2J_4C_4$ unit[36,39]. The intron–exon structure of C_H segments other than the C_μ segments are abbreviated.

subtypes). In the germ line or most other cells which do not make a λ1 light chain, the coding potential of this polypeptide chain lies in three separate DNA segments referred to as V, J and C[17,19]. (I shall denote DNA segments in italics and the corresponding regions of the polypeptide chain in roman type.) The coding potential of the major portion of the signal peptide is in a fourth DNA segment, L. The V segment codes for the rest of the signal peptide and the ~95 amino-terminal residues (numbering of amino acids is according to Kabat *et al.*[35]) which compose the classical V region minus most of FR4, which is encoded by the J segment. The C segment codes for the entire C region. The L and V segments are separated by a 93-base pair long intron, I_1, and the J and C segments by a 1,150-bp long intron, I_2. The V and J segments are separated by a spacer of an unknown length. In the B lineage cell in which a λ1 light-chain gene is expressed, the spacer between the V and J segments has been removed so that V and J are fused tail to head[17,19]. Thus, a complete λ1 chain gene consists of three exons and two introns arranged in the following order: 5′L,I_1, VJ, I_2, C 3′.

Recently, the analysis has been extended to the genes coding for λ2 and λ3 subtype chains (Fig. 2). The λ gene family contains four pairs of J and C segments, J_1C_1, J_2C_2 and J_3C_3, corresponding to the three known λ subtypes, and J_4C_4 which is defective[36–40]. It also contains two pairs of L and V segments, L_1V_1 (ref. 19) and L_2V_2 (refs 14, 41). The V_1 segments are preferentially joined with either J_3 or J_1 and the V_2 segment with J_2 (refs 42–44), although the preference is not absolute (refs 39, 45, also B. Blomberg and H. Eisen, personal communication).

Studies on the κ and heavy-chain genes have confirmed that an immunoglobin gene sequence is split in the germ-line genome and that their gene segments are reorganized during B-cell development[11,15,16,18,20–24]. However, the details of the sequence organization are somewhat different in each gene family, reflecting the degree of heterogeneity of the three families of polypeptide chains. Thus, in contrast to the λ loci, the haploid genome contains a single copy of C_κ segment[18,46,47], a cluster of five homologous but different J_κ segments located from 4.0 to 2.5 kilobases (kb) 5′ to the C_κ segment[21,23] (the middle J_κ segment, $J_{\kappa 3}$, is defective), and a cluster of 90–300 different V_κ DNA segments[48], each of which carries its own L_κ segment (Fig. 2). The significance, if any, of the difference in the relative organization of J and C segments in the κ and λ loci is unclear. The V_κ–J_κ joining occurs in various different combinations among the κ chain-synthesizing clones of B cells to form complete κ-chain genes with a variety of V_κ regions[18,20–23,49].

In the heavy chains, the core portion of the CDR3 composed of 1–13 residues is encoded by a DNA segment referred to as D, which occupies an independent position in the germ-line genome[22,24,50–53]. The haploid germ-line genome (Fig. 2) contains a cluster of 100–200 different L_H–V_H segments[54,55], a cluster of about 12 D segments[50–53] (see Fig. 3), a cluster of 4 functional J_H segments[22,24,56] and a cluster of 8 copies of C_H segments, one each for the eight immunoglobulin classes or subclasses[57–60]. Each of the eight C_H-coding sequences is composed of multiple exons, each of which corresponds to an internal homology unit forming a structural domain[61,62]. In addition, the hinge regions of γ chains[61] and the intramembrane and cytoplasmic portions of all eight chains are encoded by independent exons[59,63–67].

In the cells in which a heavy-chain gene is expressed, one each of the V_H segments, D segments and J_H segments are joined tail to head to form a complete V_H region-coding DNA segment (Fig. 2)[22,24]. Because the V_H region is encoded in three gene segments and the joinings can occur in various combinations, the diversity that can be potentially generated by rearrangement is much greater for the heavy chains than for the light chains. The heavy-chain gene that is first formed is the one for the μ chain which defines IgM, the class of immunoglobulin that appears first in B-cell development among eight classes or subclasses. Although not directly related to the present subject, namely the diversity in the antibody recognition

Fig. 3 Nucleotide sequences of germ-line *D* segments and D regions of complete heavy-chain genes. On the basis of sequence homology, the 10 germ-line *D* segments sequenced to date can be classified into three families, FL16, SP2 and Q52. In addition, two more *D* segments thought to belong to the SP2 family have been identified but have not been sequenced[53]. Of the 16 somatic *D* sequences from the rearranged, complete heavy-chain genes expressed in myelomas or hybridomas that have been characterized by DNA sequencing, 13 can be assigned to one of the three germ-line *D* families based on the sequence homology. The other three somatic *D* sequences, those of myelomas H76, HOPC1 and M173, do not show any obvious homology to any of the identified germ-line *D* sequences and characteristically contain stretches of G-C pairs. In fact, even emong the homologous 13 somatic *D* sequences, none corresponds exactly to any of the germ-line *D* sequences. However, two lines of argument suggest that the repertoire of germ-line *D* segments is not much greater than the known 12. First, if additional germ-line *D* segments existed which corresponded exactly to the somatic *D* sequences, at least some of them should have been detected by the probes used because these probes cross-hybridized with less homologous *D* segments[52]. Second, in myelomas, hybridomas and other B and T lineage lymphocytes, joined *D–J* segments thought to be intermediates of *V–D–J* joining have been found on the allelic chromosome 12[51–53,78,79]. All eight randomly selected *D–J* joinings used 1 of the 12 known germ-line *D* segments and D_{Q52} and $D_{FL16.1}$ occur twice each. This suggests that the total repertoire of germ-line *D* segments should not be statistically much greater than the known 12. To determine the boundaries between V_H and *D* as well as between *D* and J_H in the assembled gene, the nucleotide sequences of three regions, namely the 3′ end of the germ-line V_H gene, the assembled V_H–*D*–J_H and the 5′ end of the germ-line J_H segment, were compared. For M141 (ref. 22), anti-PC myelomas (S107, M603, M167)[24] and anti-NP hybridomas (B1–8, S43)[105], these three sequences are available. The nucleotide sequences in the 3′ ends of these three different germ-line *V* segments are highly conserved. Assuming the other germ-line *V* segments have the same degree of sequence homology, the V_H–*D* boundaries of the complete genes were deduced for those *V* genes whose germ-line segments have not yet been sequenced, namely MPC11[111], H76[112], V_H2[113], V_{H49} (ref. 113), MC101[114] and 'M173'[52] and B38[115]. D sequences of H76, HOPC1 and M173 cannot be classified into any of these families. The sequence of $D_{FL16.1}$, or $D_{SP2.3}$, or D_{Q52} is used as the reference sequence of each family for comparison with the others. A dash indicates that the nucleotide in question is the same as that of the reference sequence.

sites, the expression of the other heavy chains (that is, δ, α, and so on) for the other class or subclass immunoglobulins requires additional sequence reorganization that takes place either at the level of DNA (switch recombination)[22,68–71] or during the processing of the primary RNA transcript[59,66,67,72]. For the reason given above, I shall not discuss them further.

Heptamer–nonamer and 12/23-base pair spacer signal

The mechanism of *V–J* or *V–D–J* joining is not known and the main possibilities will be discussed in more detail below. There are, however, specific sequences that seem to act as joining signals.

All germ-line *V* and *D* DNA segments are followed by either a heptamer, CACAGTG, or its analogue and a nonamer, ACA_5C_2, or its analogue separated by a short spacer of an apparently unconserved sequence (Fig. 4). Likewise, all germ-line *D* and *J* segments studied to date are immediately preceded by a consensus nonamer, G_2T_5GT, and a consensus heptamer, CACTGTG, separated by a spacer[21–24,51,52]. Note that the heptameric and nonameric sequences following a gene segment (for example, V_κ, V_H or *D*) are complementary to those preceding the gene segment (in this case J_κ, *D* and J_H, respectively) with which it recombines[21,23]. In addition, the heptamers are palindromic around the central A-T or T-A pair.

All V_κ, J_λ and *D* spacers are 12 bp long, while almost all J_κ, V_λ, V_H and J_H spacers are 23 ± 1 bp long (one exception is the apparently nonfunctional $J_{\kappa 3}$ whose spacer is 21 bp long) (see Fig. 4)[22,24,51,52]. Note that functionally meaningful recombinations occur only between two gene segments, one having a 12-bp spacer and the other a 23 ± 1-bp spacer.

It has been postulated that the recombinase contains two DNA-binding proteins, one recognizing the heptamer and nonamer with a 12-bp spacer and the other recognizing them with a 23 ± 1-bp spacer[22,24]. The complementarity of the two pairs of the signal sequence may merely reflect the subunit

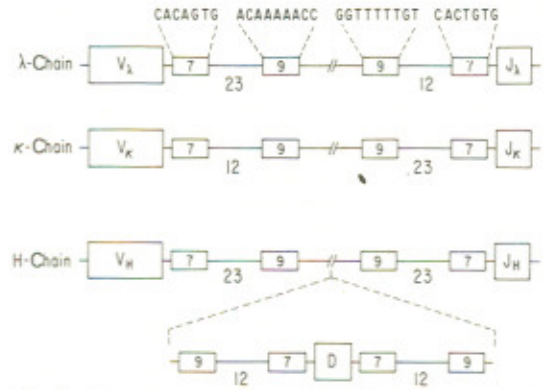

Fig. 4 Heptamer–nonamer and 12/23-bp spacer rule. See text for explanation.

structure of the recombinase. Alternatively, with the aid of the putative DNA-binding proteins, the two heptamers might form transient base pairs which will juxtapose the two recombining sequences[21,23].

Imprecise joining ends

The joining end(s) of an immunoglobulin gene segment is (are) imprecise and this contributes to antibody diversity. In κ chains, the amino acid position 96, which is usually encoded by the first J_κ triplet, is a hot spot[49,35]. At this position, most κ chains carry one of the four amino acid residues that can be fully encoded by one of the four germ-line J_κ segments[21,23,73]. However, the 96th codons of some of the rest of the κ chains could be generated if we assume that only the second and third bases or the third base alone are (is) supplied by the J_κ segment and the rest of the bases by the germ-line V_κ segments[21,23,73]. Furthermore, at least one κ chain case is known where the 96th residue is deleted entirely, and another case was found where an amino acid encoded by the 96th triplet of the V_κ

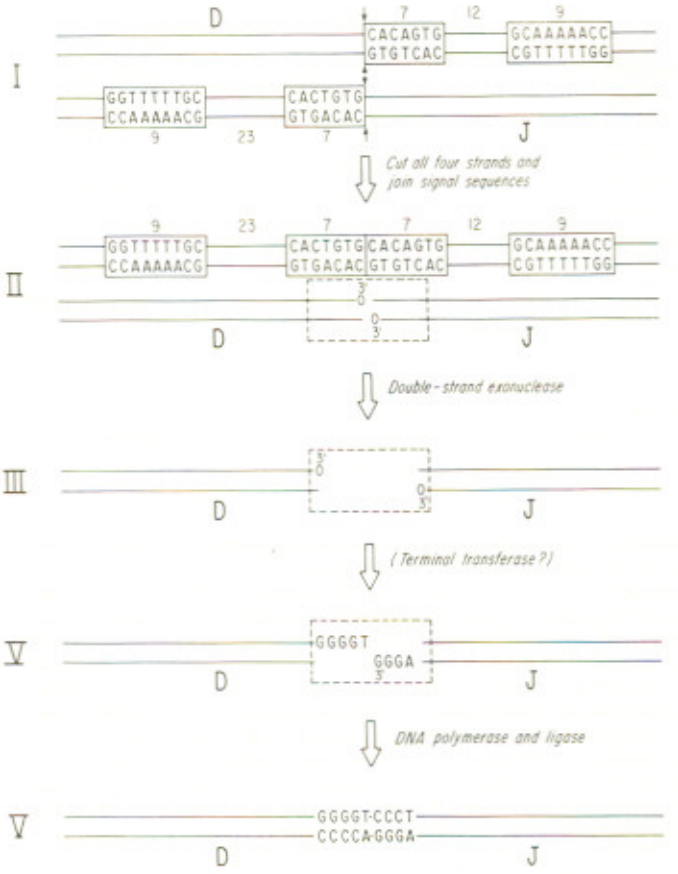

Fig. 5 Alt–Baltimore model for *D–J* joining[78]. At stage I, the two joining gene segments are lined up for joining with the help of the DNA-binding proteins. All four DNA strands are nicked exactly at those ends of the signal heptamers adjacent to the coding sequences (indicated by arrows). At stage II, the two heptamers are joined intact in a tail-to-tail fashion. The coding sequences do not join but are held by protein so that their proximity is retained. At stage III, some nucleotides may be removed from these ends by an exonuclease(s). At stage IV, one or more nucleotides may be added at these ends in a template-free fashion. This may be carried out by terminal deoxynucleotidyl transferase which is found in bone marrow as well as in thymus and polymerizes random deoxynucleotides at 3′ ends. To finish the joining process (stage V), a DNA polymerase replicates the added bases and a ligase seals the structure. It has been noticed that when no bases are added between *D* and J_H, there is an exact overlap of one or more bases from *D* and J_H at the joint[53]. This may be explained if we assume that replicative polymerization of nucleotides begins from the hybrid formed transiently by the overlapping sequences before terminal transferase comes into action. The apparent lack of insertion at the join of light-chain genes may be explained by the fact that V_L–J_L joining occurs later in B-cell development than V_H–*D*–J_H joining when terminal transferase is absent from the cells.

segment appears as an insertion[73]. These facts suggest strongly that the joining ends of the V_κ and J_κ segments can vary from one recombination event to another by several nucleotides[22]. In the case of J_H segments, this variability seems to extend over as many as 10 nucleotides[22].

A striking case for the imprecise joining end was demonstrated by analyses of two joined *D–J* segments isolated from a myeloma QUPC52 and a cytolytic T-cell clone, B6. Both joinings occurred between D_{Q52} and J_{H2} and yet both the 3′ joining end of D_{Q52} and the 5′ joining end of J_{H2} are different in the two recombinations[52].

Reminiscent of the imprecise joining end of a *D* segment is the fact that the *D* segment of an assembled heavy-chain gene is frequently several nucleotides shorter than the apparently corresponding (except for one or a few base substitutions which are probably attributed to somatic mutations as discussed below) germ-line *D* segments at one or both ends (Fig. 3)[53]. It is almost certain that this type of sequence discrepancy between the germ-line and somatic *D* segments arises from joining imprecision.

Because all immunoglobulin gene segments seem to be translated in only one reading frame in order to give rise to a functional immunoglobulin chain, one implication of this imprecision is that these gene segments often join in an out-of-phase reading frame. This type of non-productive joining occurs frequently in lymphatic cells[53,73–77]. Thus, diversity is achieved at the expense of some waste.

Insertion at the joins

Two lines of observation suggest strongly that one or more nucleotides are inserted at the join of *D* and J_H segments and also probably at the join of V_H and *D* segments. In myelomas, hybridomas and other B and T lineage lymphocytes, joined *D–J* segments thought to be intermediates of *V–D–J* joining have been found on the allelic chromosome 12[51–53,78,79]. In five of the eight *D–J* segments studied by DNA sequencing, one to four base pairs of unknown origins appear at the junction of the *D* and J_H segment[53]. One possible origin of these extra base pairs is a second germ-line *D* segment[52,53]. If this interpretation is correct, we must assume that two *D* segments can join. However, an extensive search for a joined *D–D* segment has been unsuccessful (C. Wood and S.T., unpublished). Although these negative data do not eliminate the possibility of *D–D* joining, they do increase the attractiveness of an alternative interpretation, that nucleotides are inserted without a template[53,78].

This possibility has also been suggested by comparisons of the germ-line and somatic *D* sequences, which show that extra nucleotides are added at the joins of both V_H–*D* and *D*–J_H segments[53]. As illustrated in Fig. 3, one to several nucleotides at both ends of many somatic *D* sequences cannot be explained by the apparently corresponding germ-line *D* segments. Interestingly, these nucleotides are often a run of Gs or Cs[53]. A model of *V–J* and *V–D–J* joining which contains an explanation for the origin of these nucleotides has been proposed[78] and is described below.

Mechanism of *V–J* and *V–D–J* joining

Earlier studies using rabbit allotypic markers showed that the V and C regions of a single chain are almost always encoded in the corresponding DNA segments residing in *cis* configuration[80,116,117]. This suggested that the *V–J* and *V–D–J* joinings are intrachromosomal recombinations. *A priori*, an intrachromosomal recombination can occur according to any one of a combination of several different basic models. A specific gene segment may be duplicated and the copy inserted at a site adjacent to the second gene segment (copy–insertion)[81]. Alternatively, a specific gene segment may be excised into an episome-like structure and this in turn is integrated adjacent to the second gene segment (excision–insertion)[81]. The third possibility is that DNA in the interval between the two joining gene segments loops out, is excised, and is diluted out on subsequent cell multiplication (looping out–deletion)[81]. A fourth possibility is that the two gene segments to be joined are in the opposite orientation in the germ-line genome and the DNA between a particular gene segment (inclusive) and the second gene segment (exclusive) is inverted (inversion)[81]. Yet another possibility is that during the mitotic duplication of a chromosome an unequal crossing-over occurs between the 3′ end of a gene segment on one sister chromatid and the 5′ end of the second gene segment on the second sister chromatid (sister chromatid exchange)[82,83]. The first Southern experiment addressing this question, in which the fate of the sequence

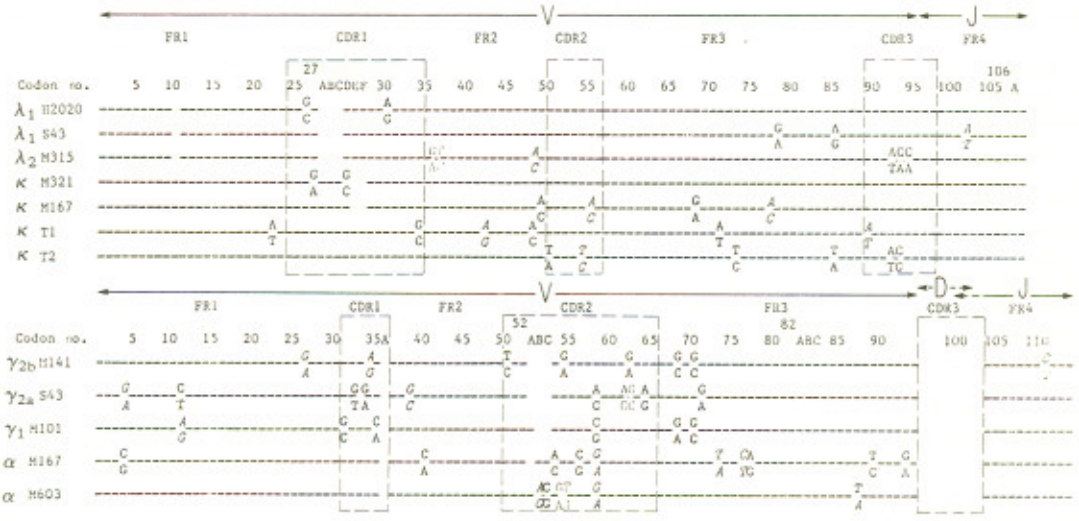

Fig. 6 Those mutations that have been documented by comparing the nucleotide sequences of a rearranged immunoglobulin gene and its corresponding germ-line gene segments are compiled. Where the corresponding nucleotides are identical, they are indicated by dashes; where different, the germ-line and somatic nucleotides on the coding strands are shown above and below, respectively. Neutral changes are indicated in italics. Data for the heavy-chain CDR3 are not given because of the uncertainty in the correspondence of a germ-line *D* segment to the somatic *D* region in question. References are: H2020λ1 (ref. 19), S43λ1 (ref. 109) and M315λ2 (ref. 109), M321 κ (O. Bernard, G. Heinrich and S.T., unpublished), M167 κ (refs 96, 97), T1 and T2 κ (ref. 110), M141γ2b (ref. 22), S43γ2a (ref. 105), M101γ1 (ref. 114), M167α and M603α (ref. 24). All rearrangements but T2 κ are normal. T2 κ is rearranged in a wrong reading frame. Numbering of codons is according to the system used by Kabat *et al.*[35] which includes alphabetical letters A, B, C, etc. to denote some codon positions in order to maximize the sequence homology among the *V* regions of various species.

immediately 5′ to the $J_{\lambda 1}$ segment and immediately 3′ to the $V_{\lambda 1}$ segment was examined, gave a result in favour of the looping out–deletion model[21]. A subsequent experiment demonstrating the absence of a 5′-flanking region of the J_κ cluster and a 3′-flanking region of a V_κ segment in some κ-producing myelomas confirmed the deletion model[84].

However, Zachau and his co-workers found that some myelomas retain a J_κ heptamer joined with a V_κ heptamer in tail-to-tail configuration[82,85]. In spite of the apparently reciprocal structural feature of this type of recombinant and the V_κ–J_κ recombinant, the joined signal sequences were not the reciprocal product of the recombination that brought together the V_κ and J_κ segments expressed in the myeloma[82,85]. These results suggested strongly that in the κ locus of some myelomas more than one recombination has occurred and that looping out–deletion cannot be the only mechanism. In agreement with these results, Van Ness *et al.* found that a 5′-flanking sequence of the J_κ cluster remains in a new sequence context in some myelomas and in some κ-bearing B cells[83]. On the basis of these results, Höchtl *et al.* and Van Ness *et al.* suggested the involvement of sister chromatid exchange in V_κ–J_κ joining[82,83].

However, the combination of inversion and deletion can also explain all existing findings[86] and a recent finding of a truly reciprocal product in κ myelomas[87] favours the involvement of inversion. Note that except for one *D* segment (D_{Q52}) and J_H segments, we remain ignorant about the relative orientation of various clusters of gene segments in the germ-line genome. A sister chromatid exchange between two joining gene segments of an opposite orientation will cause the loss of chromosomes on which these gene segments lie. Another testable criterion is that a segment of chromosome may be duplicated if a sister chromatid exchange occurs[86].

Several findings, all mentioned above, are relevant in considering the *V–J* and *V–D–J* joining mechanism at the fine structure level. First, the apparent by-product of V_κ–J_κ joining is composed of the heptamer 5′ to a J_κ joined directly with the heptamer 3′ to a V_κ in a tail-to-tail fashion. Second, the coding end of a given gene segment may differ by several nucleotides between two separate joining events. Third, in the case of heavy-chain genes, one to several nucleotides of unknown origin are apparently inserted at the join during the joining event. Based on these findings, Alt and Baltimore have proposed[78] the model outlined in Fig. 5. The details of the model are described in the legend.

Somatic mutation

The idea that mutation is a somatic diversifier of antibodies goes as far back as the middle of he 1950s[88,89]. During the subsequent two decades, there have been extensive debates about this hypothesis and the alternative germ-line hypothesis (that is, for each immunoglobulin chain sequence there is a germ-line gene encoding it)[3,4,46,90–94]. The first cogent evidence for the mutation theory was obtained by comparing the diversity of mouse λ1 chains and their genes. Cohn, Weigert and their co-workers determined the V_λ region sequences of 19 different BALB/c and NZB λ1 myelomas and found 12 of them to be identical (prototype sequences) and the other 7 different not only from each other but also from the prototype sequence by one to three residues (variant sequences)[25,42,95]. Brack *et al.*[17] and Bernard *et al.*[19] showed that the mouse has a single $V_{\lambda 1}$ segment per haploid genome and that its nucleotide sequence corresponds exactly to that of the prototype $V_{\lambda 1}$ region: the variant sequences must therefore have arisen by somatic mutations.

Several subsets of κ and heavy chains and their germ-line *V* segments have been analysed by cloning and sequencing. These results all confirmed that somatic mutations amplify the diversity encoded in the germ-line genome (Fig. 6). Some examples are briefly described below. The V_κ probe prepared from the κ gene expressed in MOPC167 atypically detects only one major band in the Southern gel blot analysis of total cellular DNA, allowing a relatively easy sequence comparison of the germ-line V_κ segment and its rearranged counterparts from two myelomas (MOPC167 and MOPC511)[96,97]. Four and five mutations were found in the MOPC167 V_κ and in the MOPC511 V_κ, respectively; none of the mutations in the two expressed genes were identical.

Recently, a set of V_κ regions referred to as the V_κ-21 group, which show a more typical κ diversity pattern, has been analysed. A large number (>30) of κ chains of this group were identified, sequenced and classified into seven different subgroups[98,99], where a subgroup is defined as a set of V regions sharing subgroup-associated residues[100]. Within a subgroup, the sequence variability of the V_κ regions is very similar to that of entire $V_{\lambda 1}$ regions: they consist of both prototype and variant sequences (see above). An earlier saturation hybridization experiment[101] and the recent sequencing analysis of the germ-line V_κ-21 DNA segments (G. Heinrich and S.T., unpublished) strongly suggested that each subgroup is defined by a single

germ-line V_κ segment coding for the prototype V region and that the variant V regions arose by somatic mutation.

Similar observations have been made in heavy chains and, in particular, those V_H regions associated with the binding of phosphorylcholine (PC). Of the 19 V_H regions of anti-PC antibodies for which complete or nearly complete amino acid sequences are available, 10 (five μ chains and four α chains) use a single sequence, denoted T15 (prototype sequence). The remaining nine (five γ chains and four α chains) of these V_H regions differ from each other and from the prototype sequence by one to eight residues (variant sequences)[102]. A cDNA probe containing the prototype V_H sequence detects four different germ-line V_H segments, one of which corresponds exactly to the prototype amino acid sequence. An analysis of sequence homology[103] and, in some cases, of the restriction maps in the 5′ upstream regions[104], allowed the conclusion that the same germ-line gene also gave rise to all nine variant sequences.

In the case of anti-NP^b [(4-hydroxy-3 nitrophenyl) acetyl] antibodies[105], it was shown by comparison of the 5′ upstream regions that a germ-line V_H segment should have generated a somatic V_H segment γ chain that differed by 10 bases.

As shown in Fig. 6, somatic mutations are not restricted to the *V* segments but have been shown to occur in *J* segments and *D* segments. I have tacitly assumed that the base substitutions observed between the germ-line and rearranged gene segments arise by mutations. The fact that these somatic changes occur by single base substitutions in the overwhelming majority of cases (63 out of a total of 67 changes) supports this assumption rather than either of the two alternative mechanisms previously proposed: reciprocal recombinations between two homologous *V* 'genes'[92,93] and gene conversion (refs 106–108 and M. Gefter and M. Fox, personal communication). Both of these mechanisms require two or more copies of homologous *V*-gene segments. Because the evidence for one germ-line $V_{\lambda 1}$ segment is good, neither of these mechanisms can apply to this gene system. (Note that the germ-line $V_{\lambda 2}$ segment[14] cannot supply any of the bases required for the substitutions.) In the other gene systems hitherto studied, it is more difficult to eliminate these mechanisms because not all cross-hybridizing germ-line *V*-gene segments have been characterized by sequencing. Nevertheless, the studies on the more closely related set of germ-line *V* segments hitherto carried out have provided no supporting evidence[103,109].

Do lymphocytes have a hypermutation system which will alter bases specifically in the *V* genes[89,90], or are these base changes merely a consequence of an ordinary level of mutation (spontaneous mutation) followed by selection acting on the protein[3,4]. The multiplication of B cells is stimulated by antigen, and those B cells bearing some mutant antibodies that fit better with an antigen may preferentially multiply. An alternative means of preferential B-cell stimulation is the idiotype–antiidiotype interaction[118]. Two observations, however, strongly suggest that the former hypothesis is correct. First, the frequency of mutation seems to be very high (Fig. 6). Second, the regions in which mutations are introduced are highly restricted to the *V–J* or *V–D–J* segments and, more significantly, to a few hundred base-pair untranslated regions surrounding the *V–J* or *V–D–J* segment. Because mutations in a particular untranslated region cannot be selected by a mechanism acting on the protein product, the above findings argue strongly against the hypothesis that these mutations arose by random spontaneous changes followed by selection operating on the gene products.

The existence of a hypermutation mechanism is not, of course, incompatible with the selection of the mutants by a mechanism acting on the protein. In fact, the density of mutations is about three times greater in CDRs than in FRs (Fig. 6). This could be explained without assuming selection if the putative hypermutation apparatus has a built-in mechanism to generate more mutations in CDRs. However, an alternative and more likely explanation is that those mutants with changes in CDRs are selected at the protein level because of the primary role of CDRs in the interaction with antigens.

In some immunoglobulin genes, there exists a curiously high degree of preference for a purine on the coding strand to be mutated (ref. 119 and M. Weigert, personal communication). Thus, in the translated regions of two $\lambda 1$ chain (H2020 and S43), three κ chain (MOPC167, MOPC321 and T1) and three heavy-chain genes (M141, S43 and M101), 36 out of a total of 41 mutations occurred at purine residues in the coding strand. This may be considered as additional evidence for the existence of a hypermutational mechanism. However, the same tendency does not apply to the other four genes studied to date (M315λ2, T2κ, M167α and M603α), where only 13 out of 30 mutations are purines.

As immunoglobulin gene sequences undergo two major somatic changes during B-cell differentiation—*V–J* or *V–D–J* joining and mutation—a question arises as to the temporal order of these events. Because there are far fewer mutations in the V_H regions of μ chains which are expressed first than in those of γ chains which are expressed later, Gearhart *et al.*[102] and Bothwell *et al.*[105] concluded that *V–D–J* joining precedes mutation. They further suggested that somatic mutations may be linked in time and even in mechanism to switch recombination. If this hypothesis is correct, it may mean that all μ chains expressed on the surface of virgin B cells are products of rearranged germ-line V_H gene segments and that the mutation mechanism is called into action only after these B cells meet antigens.

Three considerations call this hypothesis into question, however. First, μ chains with mutated V_H regions have been found (C. Berek, personal communication). Second, the idea that somatic mutation is linked in mechanism to switch recombination is refuted by the finding of α chains with V_H regions encoded by germ-line sequences[102]. Third, and perhaps most important, we know very little about the extent of selection by antigen or other means (that is, anti-idiotype) and its effect on the relative occurrence of germ-line and somatically mutated V_H gene segments in the populations of various isotypes. It is possible, for instance, that mutation takes place during *V–D–J* joining. If the recombination accompanies mutations relatively infrequently and there is not much selection for mutants among the population of IgM-bearing cells, then most of the μ chains would carry germ-line-encoded V_H regions. Subsequent selection may drastically increase the fraction of mutants among IgG or IgA populations.

In three cases studied to date, unrearranged V_κ segments isolated from three myelomas have been shown not to carry mutations (ref. 96 and G. Heinrich and S.T., unpublished). In contrast, mutations are common among rearranged V_κ segments. At first glance, these results may suggest that mutations occur during or after *V–J* joining, but this interpretation is subject to the same criticism raised above concerning selection. Interestingly, in one of two cases where a V_κ segment has rearranged to a J_κ segment in an out-of-phase reading frame, six mutations were found in the rearranged V_κ segment[110]. This V_κ segment does not produce a complete κ chain and is therefore presumably unselected. Analysis of more V_κ segments of this sort in comparison with unrearranged V_κ segments in myelomas may be very useful to our understanding of the possible relationship between *V–J* joining and mutation.

Virtually nothing is known about the molecular mechanism of the putative hypermutation. *A priori* mutations can be introduced either during semiconservative replication or repair synthesis. Because of the relative ease in accounting for the occurrence of these mutations in restricted areas of DNA, models presented to date generally favour the latter possibility. Brenner and Milstein suggested that a specific sequence near the *V–C* junction and in the *C* segment may act as a substrate for a DNA-nicking enzyme (note that it was assumed that *V* and *C* sequences are contiguous)[90]. The nick is converted to a gap by a 3′ exonuclease which is subsequently repaired by an error-prone polymerase and sealed by a ligase. In the light of new

information, this model obviously must be modified. One specific modification suggested by Selsing and Storb is that mutations are tied in mechanism to *V–J* (and perhaps to *V–D–J*) joining[96].

Conclusion: four somatic diversifiers

It is clear from the above that the principal mechanisms that diversify the germ-line-encoded genetic information for immunoglobulin chains have now been identified. The somatically generated diversity derives from four sources. The first type of diversity may be referred to as combinatorial diversity. This diversity arises from the fact that V_L and V_H regions are encoded by two and three different gene segments, respectively, each of which exists in multiple copies of different sequences, and that the joining can occur in various combinations in the population of lymphocytes. Thus, if the genome carries 2 V_λ, 3 J_λ, 300 V_κ and 4 J_κ segments, the maximum number of different V_L regions that could be somatically generated by combination is 1,206 (2×3 plus 300×4). Likewise, if there exist 200 V_H, 12 D and 4 J_H segments, the maximum number of different V_H regions would be 9.6×10^3 ($200\times12\times4$). In reality, some of these combinations may not occur or occur very rarely (for example, $V_{\lambda1}$–$J_{\lambda2}$).

Both the second and the third types of diversity occur in the joint regions and are referred to as junctional site diversity and junctional insertion diversity. The junctional site diversity arises at the V_L–J_L, V_H–D and D–J_H junctions because the joining ends are imprecise. The junctional insertion diversity seems to occur only in the V_H–D and D–J_H junctions where one to several nucleotides are inserted apparently in a template-independent fashion. Superimposed on these three somatic diversifiers based on recombination is somatic mutation, which alters bases throughout the sequences coding for the variable region.

The existence of the somatic diversifiers has now been established. While it is certain that these mechanisms are important for the function of the immune system, we remain ignorant about the details of the molecular mechanisms and their control in B-cell development. We are now in a position to resolve these issues.

The work was supported in part by grants from the National Institute of Allergy and Infectious Diseases (AI-17879-02), the American Cancer Society (NP-372), the Cancer Research Institute, Inc. (New York), and Abbott Laboratory.

1. Jerne, N. K. *Proc. natn. Acad. Sci. U.S.A.* **41,** 849 (1955).
2. Jerne, N. K. *A. Rev. Microbiol.* **14,** 341 (1960).
3. Jerne, N. K. *Eur. J. Immun.* **1,** 1 (1971).
4. Cohn, M. *Cell. Immun.* **1,** 461 (1970).
5. Hilschmann, N. & Craig, L. C. *Proc. natn. Acad. Sci. U.S.A.* **53,** 1403 (1965).
6. Putnam, F. W. & Easley, C. W. *J. biol. Chem.* **240,** 1626 (1965).
7. Wu, T. T. & Kabat, E. A. *J. exp. Med.* **132,** 211 (1970).
8. Davies, D. R., Padlan, E. A. & Segal, D. *A. Rev. Biochem.* **44,** 639 (1975).
9. Dreyer, W. J. & Bennett, J. C. *Proc. natn. Acad. Sci.U.S.A.* **54,** 864 (1965).
10. Mage, R. *et al.* in *The Antigens* Vol. 1 (ed. Sela, M.) 299 (Academic, New York, 1973).
11. Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73,** 3628 (1976).
12. Tonegawa, S., Brack, C., Hozumi, N. & Schuller, R. *Proc. natn. Acad. Sci. U.S.A.* **74,** 3518 (1977).
13. Brack, C. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **74,** 5652 (1977).
14. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **75,** 1485 (1978).
15. Seidman, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **75,** 3881 (1978).
16. Rabbitts, T. H. & Foster, A. *Cell* **13,** 319 (1978).
17. Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. *Cell* **15,** 1 (1978).
18. Lenhard-Schuller, R., Hohn, B., Brack, C., Hirama, M. & Tonegawa,S. *Proc. natn. Acad. Sci. U.S.A.* **75,** 4709 (1978).
19. Bernard, O., Hozumi, N. & Tonegawa, S. *Cell* **15,** 1133 (1978).
20. Seidman, J. G. & Leder, P. *Nature* **276,** 790 (1978).
21. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. *Nature* **280,** 288 (1979).
22. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. *Nature* **286,** 676 (1980).
23. Max, E. E., Seidman, J. G. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **76,** 3450 (1979).
24. Early, P., Huang, H., Davs, M., Calame, K. & Hood, L. *Cell* **19,** 981 (1980).
25. Cohn, M. *et al.* in *The Immune System, Genes, Receptors, Signals* (eds Sercarz, E. E., Williamson, A. R. & Fox, C. F.) 89 (Academic, New York, 1974).
26. Rabbitts, T. H., Matthyssens, G. & Hamlyn, P. H. *Nature* **284,** 238 (1980).
27. Hieter, P. A., Korsmeyer, S. J., Waldmann, T. A. & Leder, P. *Nature* **290,** 369 (1981).
28. Sheppard, H. W. & Gutman, G. A. *Cell* **29,** 121 (1982).
29. Matthyssens, G. & Rabbitts, T. H. *Proc. natn. Acad. Sci. U.S.A.* **77,** 6561 (1980).
30. D'Eustachio, P., Bothwell, A. L. M., Takaro, T. K., Baltimore, D. & Ruddle, F. H. *J. exp. Med.* **153,** 793 (1981).
31. Hengartner, H., Meo, T. & Muller, E. *Proc. natn. Acad. Sci. U.S.A.* **75,** 4494 (1978).
32. Swan, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **76,** 2735 (1979).
33. Meo, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77,** 550 (1980).
34. D'Eustachio, P., Pravcheva, D., Marcu, K. & Ruddle, F. H. *J. exp. Med.* **151,** 1545 (1980).
35. Kabat, E. A., Wu, T. T. & Bilofsky, H. in *Sequence of Immunological Chains* NIH Publ. No. 80-20008 (1979).
36. Blomberg, B., Traunecker, A., Eisen, H. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **78,** 3765 (1981).
36. Miller, J. & Storb, U. *Proc. natn. Acad. Sci. U.S.A.* **79,** 4681 (1981).
38. Blomberg, B. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **79,** 530 (1982).
39. Blomberg, B. & Tonegawa, S. *UCLA Symp. molec. cell. Biol.* **26** (in the press).
40. Miller, J., Selsing, E. & Storb, U. *Nature* **295,** 428 (1982).
41. Hozumi, N., Brack, C., Pirrotta, V., Lenhard-Schuller, R. & Tonegawa, S. *Nucleic Acids Res.* **5,** 1779 (1978).
42. Weigert, M., Cesari, I. M., Yondovich, S. J. & Cohn, M. *Nature* **228,** 1045 (1970).
43. Dugan, E. S., Bradshaw, R. A., Simms, E. S. & Eisen, H. N. *Biochemistry* **12,** 5400 (1973).
44. Azuma, T., Steiner, L. & Eisen, H. *Proc. natn. Acad. Sci. U.S.A.* **78,** 569 (1981).
45. Elliot, B. W. Jr, Eisen, H. N. & Steiner, L. *Nature* **299,** 559 (1982).
46. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. *Science* **202,** 11 (1978).
47. Gough, N. M., Cory, S. & Adams, J. M. *Nature* **281,** 394 (1979).
48. Cory, S., Tyler, B. M. & Adams, J. M. *J. molec. appl. Genet.* **1,** 103 (1981).
49. Weigert, M., Gatmaitan, L., Loh, E., Schilling, J. & Hood, L. *Nature* **276,** 785 (1978).
50. Schilling, J., Clevinger, B., Davie, J. M. & Hood, L. *Nature* **283,** 35 (1980).
51. Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa, S. *Nature* **290,** 562 (1981).
52. Kurosawa, Y. *et al. Nature* **290,** 565 (1981).
53. Kurosawa, Y. & Tonegawa, S. *J. exp. Med.* **155,** 201 (1982).
54. Givol, D. *et al. Nature* **292,** 426 (1981).
55. Kemp, D. *et al. J. molec. appl. Genet.* **1,** 245 (1981).
56. Gough, N. M. & Bernard, O. *Proc. natn. Acad. Sci. U.S.A.* **78,** 509 (1981).
57. Tucker, P. W., Liu, C.-P., Mushinski, J. F. & Blattner, F. R. *Science* **209,** 1353 (1980).
58. Roeder, W., Maki, R., Traunecker, A. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **78,** 474 (1981).
59. Maki, R. *et al. Cell* **24,** 353 (1981).
60. Shimizu, A., Takahashi, N., Yaoita, Y. & Honjo, T.*Cell* **28,** 499 (1982).
61. Sakano, H. *et al. Nature* **277,** 627 (1979).
62. Early, P. W., Davis, M. M., Kaback, D. B., Davidson, N. & Hood, L. *Proc. natn. Acad. Sci. U.S.A.* **76,** 857 (1979).
63. Rogers, J. *et al. Cell* **20,** 303 (1980).
64. Early, P. *et al. Cell* **20,** 313 (1980).
65. Alt, F. *et al. Cell* **20,** 293 (1980).
66. Moore, K. *et al. Proc. natn. Acad.Sci. U.S.A.* **78,** 1800 (1981).
67. Knapp, M. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **79,** 2996 (1982).
68. Davis, M. M. *et al. Nature* **283,** 733 (1980).
69. Maki, R., Traunecker, A., Sakano, H., Roeder, W. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **77,** 2138 (1980).
70. Obata, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **78,** 2437 (1981).
71. Kataoka, T., Miyata, T. & Honjo, T. *Cell* **23,** 357 (1981).
72. Marcu, K. B. *Cell* **29,** 719 (1982).
73. Weigert, M. *et al. Nature* **283,** 497 (1980).
74. Altenburger, W., Steinmetz, M. & Zachau, H. G. *Nature* **287,** 603 (1980).
75. Max, E. E., Seigman, J. G., Miller, H. & Leder, P. *Cell* **21,** 793 (1980).
76. Walfield, A., Selsing, W., Arp, B. & Storb, U. *Nucleic Acids Res.* **9,** 1101 (1981).
77. Hozumi, N. *et al. Proc. natn. Acad. Sci. U.S.A.* **78,** 7019 (1981).
78. Alt, F. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **79,** 4118 (1982).
79. Early, P. & Hood, L. *Cell* **24,** 1 (1981).
80. Dubinski, S. *J. Immun.* **103,** 120 (1969).
81. Tonegawa, S., Hozumi, N., Matthyssens, G. & Schuller, R. *Cold Spring Harb. Symp. quant. Biol.* **41,** 877 (1977).
82. Höchtl, J., Müller, C. R. & Zachau, H. *Proc. natn. Acad. Sci. U.S.A.* **79,** 1383 (1982).
83. Van Ness, B. G., Coleclough, C., Perry, R. P. & Weigert, M. *Proc. natn. Acad. Sci. U.S.A.* **79,** 262 (1982).
84. Seidman, J. G. *et al. Proc. natn. Acad. Sci. U.S.A.* **77,** 6022 (1980).
85. Steinmetz, M., Altenburger, W. & Zachau, H. *Nucleic Acids Res.* **8,** 1709 (1980).
86. Lewis, S., Rosenberg, N., Alt, F. & Baltimore, D. *Cell* **30,** 807 (1982).
87. Höchtl, J. & Zachau, H. G. *Nature* **302,** 260–263 (1983).
88. Burnet, M. *Nature* **210,** 1308 (1966).
89. Lederberg, J. *Science* **129,** 1649 (1959).
90. Brenner, S. & Milstein, C. *Nature* **211,** 242 (1966).
91. Hood, L. & Talmage, D. *Science* **18,** 325 (1970).
92. Gally, J. A. & Edelman, G. M. *Nature* **227,** 341 (1970).
93. Whitehouse, H. L. K. *Nature* **215,** 371 (1967).
94. Smithies, O. *Science* **157,** 267 (1967).
95. Weigert, M. & Riblet, R. *Cold Spring Harb. Symp. quant. Biol.* **41,** 837 (1976).
96. Selsing, E. & Storb, U. *Cell* **25,** 47 (1981).
97. Gershenfeld, H. K., Tsukamoto, A., Weissman, I. L. & Joho, R. *Proc. natn. Acad. Sci. U.S.A.* **78,** 7674 (1982).
98. Weigert, M., Gatamaitan, L., Loh, E., Schilling, J. & Hood, L. *Nature* **276,** 785 (1978).
99. McKean, D. J., Bell, M. & Potter, M. *Proc. natn. Acad. Sci. U.SA.* **75,** 3913 (1978).
100. Potter, M.*Adv. Immun.* **25,** 141 (1977).
101. Valbuena, O., Marcu, K., Weigert, M. & Perry, R. *Nature* **276,** 780 (1978).
102. Gearhart, P. J., Johnson, N. D., Douglas, R. & Hood, L. *Nature* **291,** 29 (1981).
103. Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. *Cell* **25,** 59 (1981).
104. Kim, S., Davis, M., Sinn, E., Patten, P. & Hood, L. *Cell* **27,** 573 (1981).
105. Bothwell, A. L. M. *et al. Cell* **24,** 625 (1981).
106. Baltimore, D. *Cell* **24,** 592 (1981).
107. Dildrop, R., Brüggemann, M., Radbruch, A., Rajewsky, K. & Beyreuther, K. *EMBO J.* **5,** 635 (1982).
108. Schreier, P. H., Bothwell, A. L. M., Müller-Hill, B. & Baltimore, D. *Proc. natn. Acad. Sci. U.S.A.* **78,** 4495 (1981).
109. Bothwell, A. L. M. *et al. Nature* **290,** 65–67 (1981).
110. Pech, M., Höchtl, J., Schnell, H. & Zachau, H. G. *Nature* **291,** 668 (1981).
111. Zakut, R., Cohen, J. & Givol, D. *Nucleic Acids Res.* **8,** 3591 (1980).
112. Bernard, O. & Gough, N. M. *Proc. natn. Acad. Sci. U.S.A.* **54,** 864 (1980).
113. Hood, L. *et al. Cold Spring Harb. Symp. quant. Biol.* **45,** 887 (1980).
114. Honjo, T. *et al. Cold Spring Harb. Symp. quant. Biol.* **45,** 913 (1980).
115. Early, P., Nottenburg, C., Weissman, I. & Hood, L. *Molec. cell. Biol.* **2,** 829 (1982).
116. Kindt, T. J., Mandy, W. J. & Todd, C. W. *Biochemistry* **9,** 2028 (1970).
117. Landucci-Tosi, S., Mage, R. G. & Dubiski, S. *J. Immun.* **104,** 641 (1970).
118. Jerne, N. K. *Ann. Immun.* **125C,** 373 (1974).
119. Baltimore, D. *Cell* **26,** 295 (1981).