Reprinted from
Cold Spring Harbor Symposia on Quantitative Biology, Volume XLI

Somatic Changes in the Content and Context of Immunoglobulin Genes

S. TONEGAWA, N. HOZUMI, G. MATTHYSSENS AND R. SCHULLER
Basel Institute for Immunology, Basel 5, Switzerland

In this paper, we will demonstrate that both the content and the context of immunoglobulin genes change during the differentiation of lymphocytes. By *content*, we mean the genetic information used for coding immunoglobulin chains. By *context*, we mean the relative location of this information in DNA. The possibility of such changes has been the subject of much speculation in immunology, but only recently has it become possible to attack the problem directly. The principal technological advance which makes this possible has been the development of procedures for the purification of immunoglobulin mRNA from murine plasmacytomas (Tonegawa and Baldi 1973; Mach et al. 1973; Brownlee et al. 1973; Schechter 1973; Honjo et al. 1974; Stavnezer et al. 1974).

Joining of Kappa-chain Genes during Ontogeny

Let us begin with changes in context. To study this, we need a way to determine the arrangement of immunoglobulin genes in DNA. Bacterial restriction endonucleases recognize and cleave specific sequences of base pairs within DNA duplexes. These enzymes have proven to be extremely useful in analyzing viral and bacterial genomes, but the application of these enzymes to eukaryotic genomes has been limited to cases where the arrangement of reiterated sequences is of interest. The sensitive assay of specific DNA sequences by hybridization with highly radiolabeled, purified mRNA has enabled us to extend the use of these enzymes to the study of the arrangement of unique genes such as those for immunoglobulins.

High-molecular-weight BALB/c mouse embryo DNA was digested to completion with restriction enzyme *Bam*H1. The resulting DNA fragments were fractionated according to size by electrophoresis in a 0.9% agarose gel. DNA eluted from gel slices was assayed for V- and C-gene sequences by hybridization with ^{125}I-labeled whole κ mRNA or half molecules containing the 3′ terminus. The assay is based on the fact that the sequences corresponding to V and C genes are on the 5′- and 3′-end halves of the mRNA, respectively (Fig. 1). Thus the 3′-end half is an RNA probe for C-gene sequences, and the V-gene sequences are determined indirectly from the difference in the two hybridization levels obtained with the whole RNA molecule and the 3′-end half. The details and basis of this assay procedure have been described elsewhere (Hozumi and Tonegawa 1976). In Figure 2, κ RNA probes prepared from MOPC-321 were used. Two DNA components of 6.0×10^6 and 3.9×10^6 m.w. hybridized with the whole RNA molecules, whereas only the 6.0×10^6 component hybridized with the 3′-end half. Hence, the C_κ gene is in the 6.0×10^6 component and the V_κ gene should be in the 3.9×10^6 component. The fact that the extent of hybridization in the larger component is nearly identical with the two RNA probes supports this interpretation.

Since V and C genes are in separate DNA fragments of sizes that are much larger than the size of either gene, they are probably some distance away from each other. However, the possibility

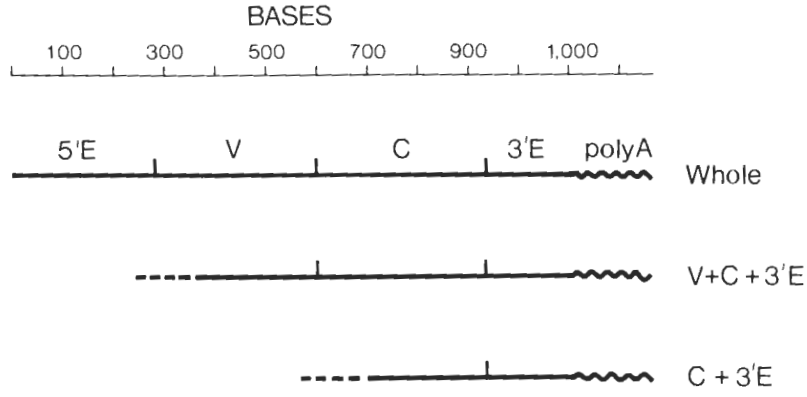

Figure 1. Schematic illustration of light-chain mRNA and its specific fragments. Light-chain mRNA is about 1250 nucleotides long and consists of four regions besides the poly(A) sequence at the 3′ end. These regions are designated by 5′E (external region at the 5′ end), V (variable region), C (constant region), and 3′E (external region towards the 3′ end). The relative lengths of the four regions are deduced from the size of V and C regions in light chains and from the data of Milstein et al. (1974). Purification, iodination, and isolation of 3′-end-containing fragments have been described previously (Tonegawa et al. 1974b; Tonegawa 1976). Size fractionation of the fragments was carried out by acrylamide gel electrophoresis in 99% formamide.

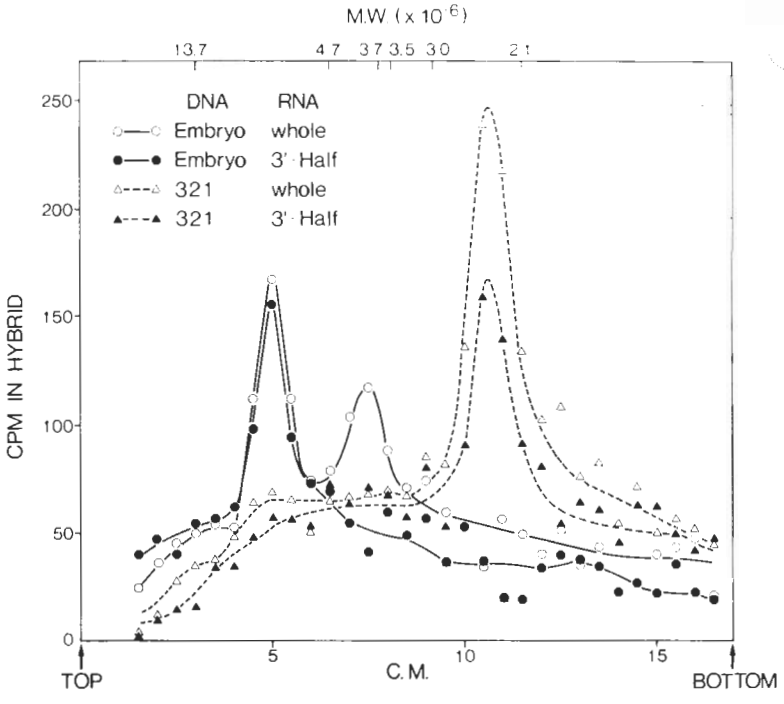

Figure 2. Gel electrophoresis pattern of *Bam*H1-generated mouse DNA fragments carrying V_κ- or C_κ-gene sequences. Results with mouse embryo and with MOPC-321 plasmacytoma DNAs are superimposed. Ten milligrams of high-molecular-weight embryo or MOPC-321 tumor DNA (prepared according to the method of Gross-Bellard et al. [1973]) in a buffer consisting of 0.006 M Tris-HCl pH 7.4, 0.006 M $MgCl_2$, and 0.006 M β-mercaptoethanol was incubated with 2×10^4 units (one unit is defined as the amount of the enzyme sufficient for digesting 1 μg phage λ DNA in 60 min at 37°C under the above conditions) of *Bam*H1 enzyme (purified according to the method of Wilson and Young [1975]) at 37°C for 4 hr. Completeness of digestion was monitored by digesting, in parallel, a mixture of mouse and phage λ DNAs (ratio 10:1) under similar conditions and analyzing the digestion product by analytical gel electrophoresis. The *Bam*H1-digested mouse DNA was concentrated with 2-butanol (Stafford and Bieber 1975), dialyzed against TA buffer (0.002 M Tris-CH_3CO_2 pH 8.0, 0.018 M NaCl, 0.002 M Na_2EDTA, 0.02 M $NaCH_3CO_2$) and electrophoresed in 0.9% agarose (Sigma, electrophoresis grade) cast in TA buffer. Electrophoresis was carried out in a Plexiglas tray (40 × 50 × 1 cm) at 4°C with 0.75 mA per cm² of gel for the first 20 hr and then 3 mA per cm² for 2 days. DNA was extracted from gel slices by melting agar in 2.5 M $NaClO_4$ (Fuke and Thomas 1970) at 60°C for 60 min. The melted agar was removed by passage through a small column of hydroxylapatite. Hybridization was carried out in 2 × SSC (0.3 M NaCl, 0.03 M Na_2 citrate) supplemented with 0.01 M Tris-HCl pH 7.5 and 0.001 M Na_2EDTA in a final volume of 25 μl. DNA corresponding to 1-mg input was annealed with ^{125}I-labeled whole MOPC-321 κ mRNA (1220 cpm) or its 3'-end half (600 cpm) at 70°C for 40 hr in a total volume of 25 μl. Specific activity of RNA was 5×10^7 cpm/μg. The processing of the hybridization mixture has been described previously (Tonegawa et al. 1974b).

that the enzyme cleaved V and C genes arranged contiguously near the boundary is not entirely eliminated. Cleavage sites are possible in the nucleotide sequences coding for the amino acids at positions 93–95 and 97–98 (McKean et al. 1973). However, the probability that either of these amino acid sequences provides the exact nucleotide sequences required is low.

The pattern of hybridization is completely different in the DNA from the homologous tumor (Fig. 2). Both RNA probes hybridized with a new DNA component of 2.4×10^6 m.w. There is no indication that either of these RNA probes hybridizes with other DNA components above the general background level. These results indicate that both V and C genes, or the entire sequences represented in the mRNA molecule (except for the unlabeled poly(A) sequence), are contained in the 2.4×10^6 m.w. component in the tumor genome. The whole RNA hybridizes with this component nearly twice as well as does the 3'-end half, thereby supporting this notion. The straightforward interpretation of these results is that V_κ and C_κ genes, which are some distance away from each other in the embryo genome, are brought together during differentiation to form a continuous nucleotide stretch in the plasma cells expressing this particular V_κ gene. The fact that neither the 6.0×10^6 nor the 3.9×10^6 component exists in the plasma-cell genome indicates that such rearrangement of immunoglobulin genes takes place in all of the homologous chromosomes in this plasma cell. Our interpretation of these results is schematically illustrated in Figure 3.

An alternative explanation of the results, namely, that accumulation of mutations or base modifications leading to either loss or gain of *Bam*H1 sites generated the observed pattern difference, is not impossible. In this view, there would have to be a *Bam*H1 site close to the V–C junction in embryo DNA. This *Bam*H1 site would have to be lost by mutation or base modification in the MOPC-321 tumor. By itself, such an alteration would cause the appearance of a single 9.9×10^6 m.w. component in the tumor. To achieve the molecular weight of the single component actually observed in the tumor (2.4×10^6), there would have to be new *Bam*H1 sites created by mutation or by base modification between the V_κ gene and the nearest site on either side. Since there is no reason why there should be any selective pressures involving *Bam*H1 sites, the occurrence of three mutations would seem to be quite unlikely.

There is also an alternative explanation for

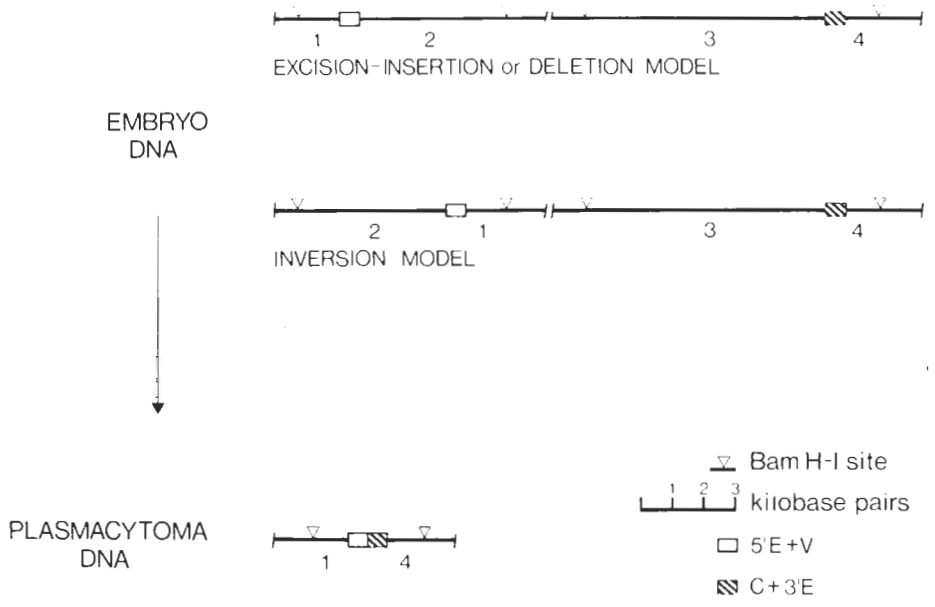

Figure 3. Schematic illustration of mouse *Bam*H1 DNA fragments carrying a MOPC-321 V_κ gene and a C_κ gene. Arrows indicate *Bam*H1 sites. 5'E and 3'E designate base sequences corresponding to untranslated regions of a κ-chain mRNA molecule at the 5' and 3' ends, respectively. V and C designate base sequences corresponding to variable and constant regions, respectively. The numbers 1 to 4 designate sequences flanked by the V or C gene and a nearest *Bam*H1 site. The relative positions of these sequences within each fragment are deduced from the present results within the framework of any one of the last three models depicted in Fig. 6.

the absence of the embryonic DNA components in the tumor, namely, that the V- to C-gene joining took place in only one of the homologous chromosomes, and that the other chromosome(s) was lost during propagation of the tumor. In view of the known chromosome abnormalities of murine plasmacytomas (Yoshida et al. 1968; Cohn 1968), we cannot eliminate this trivial possibility.

Arrangements of Lambda-chain Genes

We have also applied the kind of analysis described in the last section to mouse λ-chain genes. Figure 4 shows hybridization patterns of embryo DNA which had been digested with another restriction enzyme, *E. coli* R1 (*Eco*R1). Three DNA components of 4.7, 2.8, and 1.9 × 10^6 daltons hybridized with whole λ mRNA purified from the HOPC-2020 myeloma. Specificity and extent of hybridization obtained with whole and 3'-end-half λ mRNA indicate that only the 4.7 × 10^6 component carries a C gene (Fig. 4A). In Figure 4B, the hybridization pattern of the whole λ mRNA is compared with that of an RNA fragment which carries the 3' end and extends roughly three-quarters of the way toward the 5' end (Fig. 1). The latter RNA probe contains sequences in the V, C, and 3'-external regions but largely lacks those in the 5'-external region. Since the 2.8 × 10^6 component hybridizes with this RNA probe but not with the 3'-end half, it must carry only V-region sequences. Using analogous logic, the 1.9 × 10^6 component should carry sequences in the 5' external region. The relative extent of hybridization with the three RNA probes in each DNA component confirm these interpretations. Since amino acid sequence data indicate that there cannot be any *Eco*R1 site in the translated portions of the V_λ or C_λ genes (Weigert et al. 1970), these results prove unequivocally that, in the embryo genome, the two genes are some distance away from each other. Splitting of the sequences in the V and 5' external regions into two DNA fragments could be interpreted to mean that these two regions originate from two separate segments of DNA, but a more likely interpretation is that they originate from a continuous DNA segment, and that there is an *Eco*R1 site in the 5' external region near to the V region.

Are the V_λ and C_λ genes joined in the differentiated plasma cell as is the case with the κ-chain genes? In Figure 5, A and B compare hybridization patterns obtained with embryo and HOPC-2020 DNA, respectively. Although not as striking as the κ-chain case (see above), the two patterns do show some differences. In the plasmacytoma DNA, there seems to exist a new component of 4.3 × 10^6 m.w. which migrated in the gel slightly ahead of the 4.7 × 10^6 C-gene component. That these two components are distinct is suggested by two observations. First, the relative position of each of these two components to that of the internal marker, a DNA fragment carrying 18S ribosomal RNA, is different. Second, in the plasmacytoma DNA component, the extent of hybridization obtained with

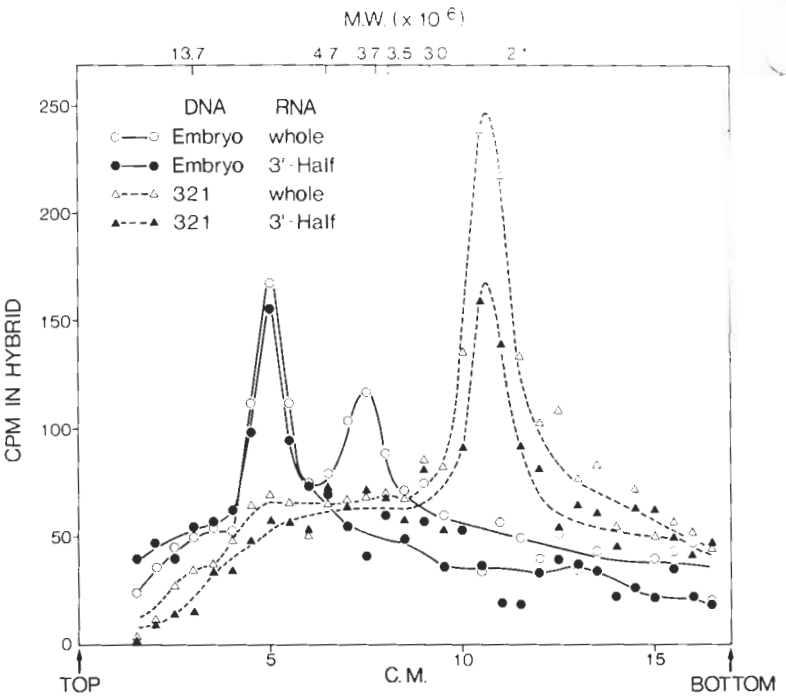

Figure 2. Gel electrophoresis pattern of *Bam*H1-generated mouse DNA fragments carrying V_κ- or C_κ-gene sequences. Results with mouse embryo and with MOPC-321 plasmacytoma DNAs are superimposed. Ten milligrams of high-molecular-weight embryo or MOPC-321 tumor DNA (prepared according to the method of Gross-Bellard et al. [1973]) in a buffer consisting of 0.006 M Tris-HCl pH 7.4, 0.006 M $MgCl_2$, and 0.006 M β-mercaptoethanol was incubated with 2×10^4 units (one unit is defined as the amount of the enzyme sufficient for digesting 1 μg phage λ DNA in 60 min at 37°C under the above conditions) of *Bam*H1 enzyme (purified according to the method of Wilson and Young [1975]) at 37°C for 4 hr. Completeness of digestion was monitored by digesting, in parallel, a mixture of mouse and phage λ DNAs (ratio 10:1) under similar conditions and analyzing the digestion product by analytical gel electrophoresis. The *Bam*H1-digested mouse DNA was concentrated with 2-butanol (Stafford and Bieber 1975), dialyzed against TA buffer (0.002 M Tris-CH_3CO_2 pH 8.0, 0.018 M NaCl, 0.002 M Na_2EDTA, 0.02 M $NaCH_3CO_2$) and electrophoresed in 0.9% agarose (Sigma, electrophoresis grade) cast in TA buffer. Electrophoresis was carried out in a Plexiglas tray (40 × 50 × 1 cm) at 4°C with 0.75 mA per cm^2 of gel for the first 20 hr and then 3 mA per cm^2 for 2 days. DNA was extracted from gel slices by melting agar in 2.5 M $NaClO_4$ (Fuke and Thomas 1970) at 60°C for 60 min. The melted agar was removed by passage through a small column of hydroxylapatite. Hybridization was carried out in 2 × SSC (0.3 M NaCl, 0.03 M Na_3 citrate) supplemented with 0.01 M Tris-HCl pH 7.5 and 0.001 M Na_2EDTA in a final volume of 25 μl. DNA corresponding to 1-mg input was annealed with ^{125}I-labeled whole MOPC-321 κ mRNA (1220 cpm) or its 3'-end half (600 cpm) at 70°C for 40 hr in a total volume of 25 μl. Specific activity of RNA was 5×10^7 cpm/μg. The processing of the hybridization mixture has been described previously (Tonegawa et al. 1974b).

that the enzyme cleaved V and C genes arranged contiguously near the boundary is not entirely eliminated. Cleavage sites are possible in the nucleotide sequences coding for the amino acids at positions 93–95 and 97–98 (McKean et al. 1973). However, the probability that either of these amino acid sequences provides the exact nucleotide sequences required is low.

The pattern of hybridization is completely different in the DNA from the homologous tumor (Fig. 2). Both RNA probes hybridized with a new DNA component of 2.4×10^6 m.w. There is no indication that either of these RNA probes hybridizes with other DNA components above the general background level. These results indicate that both V and C genes, or the entire sequences represented in the mRNA molecule (except for the unlabeled poly(A) sequence), are contained in the 2.4×10^6 m.w. component in the tumor genome. The whole RNA hybridizes with this component nearly twice as well as does the 3'-end half, thereby supporting this notion. The straightforward interpretation of these results is that V_κ and C_κ genes, which are some distance away from each other in the embryo genome, are brought together during differentiation to form a continuous nucleotide stretch in the plasma cells expressing this particular V_κ gene. The fact that neither the 6.0×10^6 nor the 3.9×10^6 component exists in the plasma-cell genome indicates that such rearrangement of immunoglobulin genes takes place in all of the homologous chromosomes in this plasma cell. Our interpretation of these results is schematically illustrated in Figure 3.

An alternative explanation of the results, namely, that accumulation of mutations or base modifications leading to either loss or gain of *Bam*H1 sites generated the observed pattern difference, is not impossible. In this view, there would have to be a *Bam*H1 site close to the V–C junction in embryo DNA. This *Bam*H1 site would have to be lost by mutation or base modification in the MOPC-321 tumor. By itself, such an alteration would cause the appearance of a single 9.9×10^6 m.w. component in the tumor. To achieve the molecular weight of the single component actually observed in the tumor (2.4×10^6), there would have to be new *Bam*H1 sites created by mutation or by base modification between the V_κ gene and the nearest site on either side. Since there is no reason why there should be any selective pressures involving *Bam*H1 sites, the occurrence of three mutations would seem to be quite unlikely.

There is also an alternative explanation for

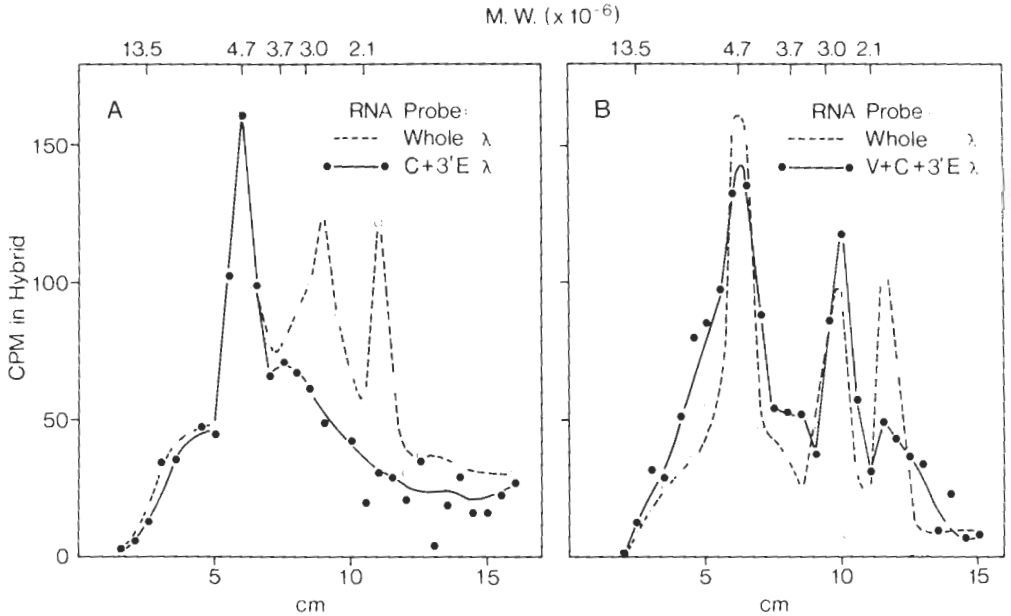

Figure 4. Gel electrophoresis patterns of embryo DNA digested with *Eco*R1 enzyme and assayed with HOPC-2020 λ-chain mRNA and its 3′-end-containing fragments. Embryo DNA was digested in R1 buffer (0.1 M Tris-HCl pH 7.5, 0.01 M $MgSO_4$, 0.05 M NaCl, 0.1 mM Na_2EDTA) with *Eco*R1 enzyme (purified according to Greene et al. [1974]) at 37°C for 4 hr. Electrophoresis and hybridization were carried out as described in the legend to Fig. 2. A single preparation of ^{125}I-labeled HOPC-2020 λ-chain mRNA (specific activity 8×10^7 cpm/μg) was used for hybridization and for preparation of 3′-end-containing fragments (see legend to Fig. 1). Input ^{125}I-counts per hybridization was 1300 cpm (whole mRNA), 1100 cpm (V + C + 3′E fragment), and 620 cpm (C + 3′E fragment).

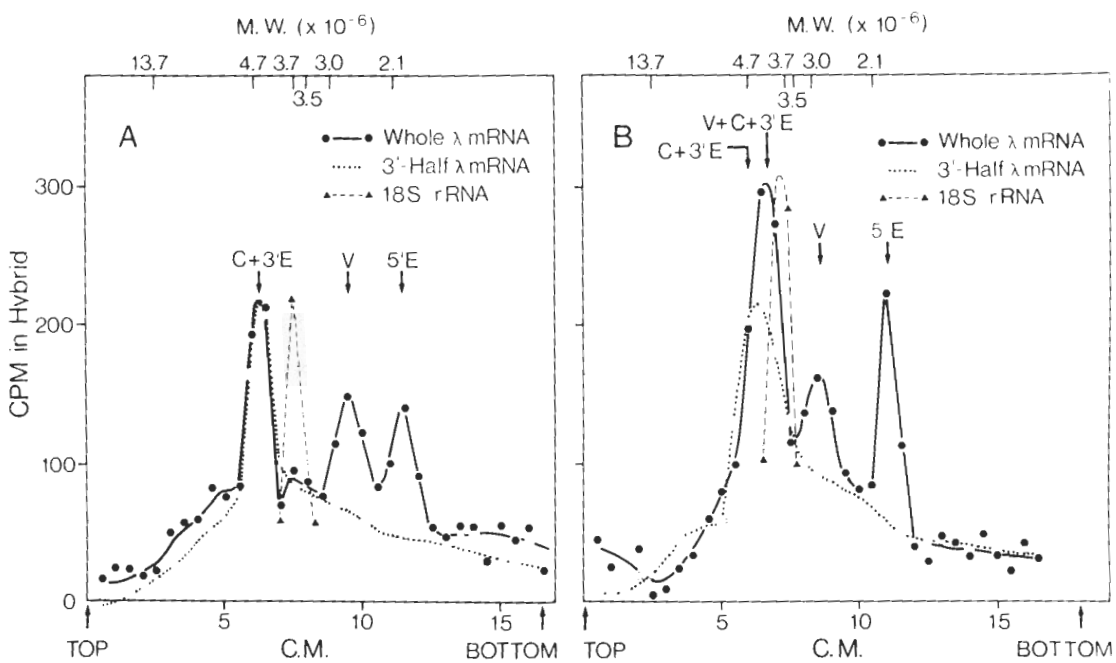

Figure 5. Gel electrophoresis patterns of embryo *(A)* and HOPC-2020 tumor *(B)* DNAs digested with *Eco*R1 enzyme and assayed with HOPC-2020 λ-chain mRNA. DNA was digested as described in the legend to Fig. 4. Electrophoresis and hybridization were carried out as described in the legend to Fig. 2. Input ^{125}I-counts were 1300 cpm (embryo DNA) and 1300 cpm (HOPC-2020 DNA). An aliquot (equivalent to 60 μg input DNA) of DNA from the indicated gel slices was fixed on a Millipore HAWP filter after denaturation and annealed to excess ^{125}I-labeled 18S ribosomal RNA.

whole RNA is nearly twice as high as that obtained with the 3′-end half, whereas both RNA probes hybridize with the embryo DNA component to nearly the same extent. Thus the results suggest that the V_λ and C_λ genes are some distance away in the embryo cells and that they seem to be joined in the plasmacytoma.

The striking difference in the two cases, κ and λ chains, is the fact that, unlike in MOPC-321, the embryonic components persist in the pattern of the HOPC-2020 DNA. Several explanations can be offered for this difference. In a BALB/c mouse, there seem to exist two λ-chain loci. These loci are represented by the λ chains produced by HOPC-2020 (λ_I) and MOPC-315 (λ_{II}). The sequence homology between the two groups of λ chains is extensive, albeit less than among themselves (Weigert et al. 1970; Dugan et al. 1973). If there are separate germ-line V_λ genes for the two groups of λ chains, mRNA coding for a λ_I chain could cross-hybridize with the $V_{\lambda II}$ gene. An alternative but trivial explanation is that the plasmacytomas originate by fusion of a committed B lymphocyte and a nonlymphatic cell, such as a macrophage (Warner and Krueger 1975). Most of the BALB/c plasmacytomas are subtetraploids (Cohn 1967; Yoshida et al. 1968). Whereas some tumors (HOPC-2020) retain at least one of the nonlymphatic chromosomes which carry the relevant immunoglobulin genes, others (MOPC-321) might throw out these chromosomes during cell propagation. Of course, it is also possible that the mechanism of V–C joining in λ chains is intrinsically different from that in κ chains.

Models for V–C Joining at DNA Level

What is the mechanism by which the integration of V- and C-gene sequences is brought about? Several models have been proposed, some of which are illustrated schematically in Figure 6. In the "copy-insertion" model, a specific V gene is duplicated and the copy is inserted at a site adjacent to a C gene (Dreyer et al. 1968). In this model, the embryonic sequence context of a V gene should be retained in the lymphocyte expressing that particular V gene. Since the embryonic DNA fragment carrying the MOPC-321 V gene does not seem to exist in

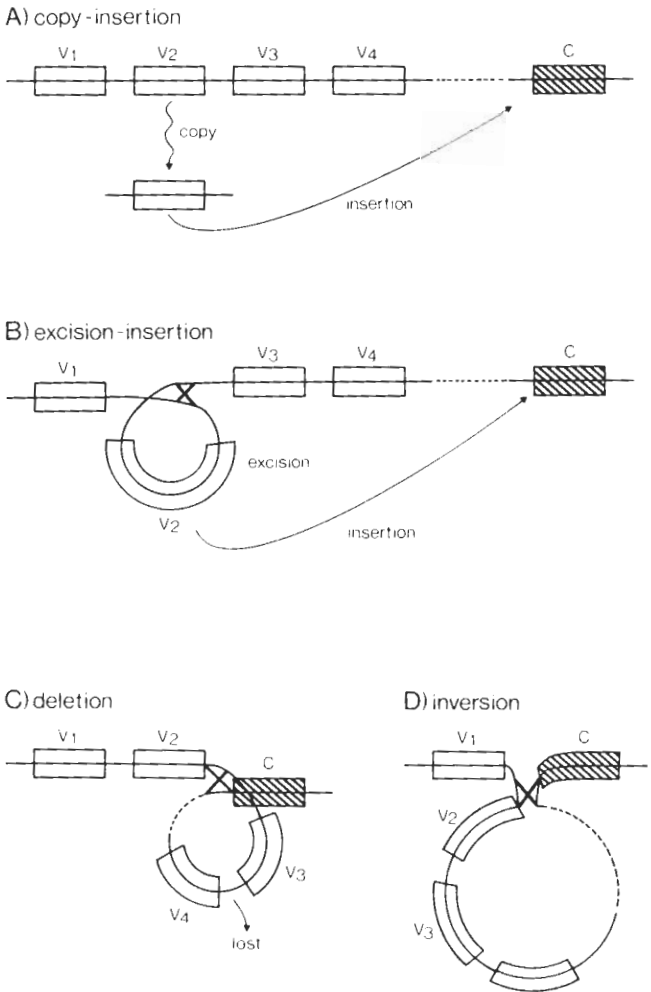

Figure 6. Models for V–C joining at DNA level. See

Capra and Kehoe 1974; Wigzell 1973). In the past several years, however, several groups of researchers have reported experimental results aimed at obtaining some insight into this formidable problem (Delovitch and Baglioni 1974; Storb 1974; Premkumar et al. 1974; Tonegawa et al. 1974a; Bernardini and Tonegawa 1974; Rabbitts et al. 1974; Leder et al. 1974).

The method commonly used has been to examine the kinetics of hybridization of a purified immunoglobulin mRNA with total cellular DNA, thereby counting the number of cross-hybridizing V genes. Earlier work with mRNA preparations of lower purity invariably generated hybridization kinetics with multiple components: a reiterated fraction and a unique or nearly unique fraction. Assuming that the reiterated fraction represents V-gene sequences, several authors interpreted their results as an indication of a germ-line basis for antibody diversity (Delovitch and Baglioni 1974; Storb 1974; Premkumar et al. 1974). However, using mRNAs coding for κ chains which code for V regions of various degrees of sequence homology, we demonstrated that the major proportion of V-region sequences is in the unique rather than the reiterated fraction (Tonegawa et al. 1974a,b). Based on competition hybridization, we could define a group of κ chains, comprising over 10% of all such chains for which sequence information was available, within which there was over 80% amino acid sequence homology. DNA sequences coding for the V regions of one κ chain from this group were reiterated only one- to twofold. This strongly suggested the existence of some somatic mechanism for antibody diversity. These findings were later confirmed by us as well as by others who used either mRNA of higher purity or DNA fractionated according to degree of reiteration (Leder et al. 1975; Rabbitts et al. 1975; Tonegawa 1976; Schuller 1976; Farace et al. 1976). These experiments showed that the *entire* sequences in both κ- and λ-chain mRNAs and in a heavy-chain mRNA are reiterated no more than a few times per haploid genome. However, as we have pointed out previously (Tonegawa et al. 1974b), the uniqueness of a V-gene sequence is, by itself, not unequivocal evidence for the somatic generation of antibody diversity. The extent of cross-hybridizability among V genes coding for different V regions must be experimentally determined. For instance, the extensive amino acid sequence homology among the V regions of a single "subgroup" does not necessarily mean that there is equally extensive nucleotide sequence homology among the corresponding V genes. Thus the kinetic data must be combined with a demonstration of nucleotide sequence homology. Using competition hybridization experiments with λ mRNAs, we demonstrated that extensive nucleotide sequence homology does in fact exist among the V regions of most BALB/c mouse myeloma λ chains. Since amino acid sequence studies strongly suggest that there is a minimum of 20 to 30 myeloma V_λ regions, we concluded that the number of germ-line V genes (a few per haploid genome) is too small to account for the diversity of these immunoglobulin chains (Tonegawa 1976).

New Evidence for the Somatic Generation of Antibody Diversity

In the past, all hybridization studies designed for counting the number of immunoglobulin genes were performed by hybridizing whole mRNA molecules or complementary DNA synthesized in vitro from this RNA with total cellular DNA. Since these probes contain not only V-gene but also C-gene sequences, it has not been possible to focus the experiments exclusively on V genes. As described above, treatment of embryo DNA with a restriction enzyme and subsequent fractionation of the DNA fragments in the preparative agarose gel permitted us to isolate and enrich the fraction containing V-gene DNA away from that containing C-gene DNA. We have therefore reexamined the validity of our previous conclusion using these DNA fractions, namely, that the number of V genes coding for BALB/c myeloma λ chains is no more than a few.

Cross-hybridizability of MOPC-104E and HOPC-2020 V_λ Genes

Let us assume that, as we concluded previously, the entire nucleotide sequences in the two λ mRNAs from MOPC-104E and HOPC-2020 plasmacytomas are indeed extensively homologous. Then the pattern of hybridizations with the two mRNA should be indistinguishable when *Eco*R1-digested embryo DNA is examined across the agarose gel. That this is the case is shown in Figure 7. As discussed earlier, the three DNA components (from large to small) carry C-gene, V-region, and 5′ external region sequences. The result is consistent with the assumption but does not prove it. It is possible that the enzyme cleaves regularly arranged multiple V_λ genes such that DNA fragments of identical size are generated, each of which contains a different V_λ gene. It is also possible, although less likely, that the 2.8×10^6 component consists of a single DNA fragment carrying sequences for multiple V regions.

We have isolated the 4.7×10^6 and 2.8×10^6 components. DNA in these components were denatured separately and hybridized to excess MOPC-104E or HOPC-2020 λ mRNAs under conditions that do not permit substantial renaturation of DNA. The mixture was then treated with S1 nuclease to digest all single-stranded nucleic acids. The remaining nucleic acid was treated with alkali to digest RNA in hybrid and then rehybridized with ^{125}I-labeled MOPC-104E or HOPC-

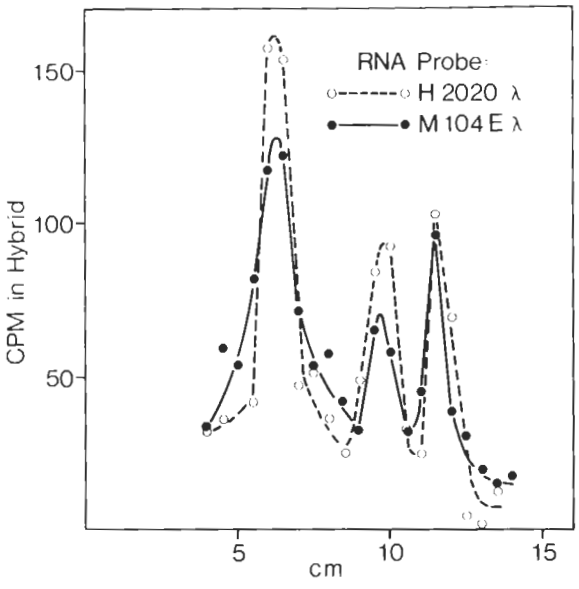

Figure 7. Gel electrophoresis patterns of embryo DNA digested with *Eco*R1 enzyme and assayed with MOPC-104E or HOPC-2020 λ-chain mRNA. Digestion of DNA is described in the legend to Fig. 4. Electrophoresis and hybridization are described in the legend to Fig. 2. Specific activity and input ^{125}I-counts were 8×10^7 cpm/μg and 1100 cpm, respectively, for MOPC-104E λ-chain mRNA, and 9×10^7 cpm/μg and 1300 cpm, respectively, for HOPC-2020 λ-chain mRNA.

Table 1. Cross-hybridization of Two Lambda mRNAs

DNA fragment	RNA: first hybridization	RNA: second hybridization	^{125}I-cpm in second hybridization
C	MOPC-104E	MOPC-104E	299
		HOPC-2020	255
C	HOPC-2020	MOPC-104E	285
		HOPC-2020	286
C	none	MOPC-104E	81
		HOPC-2020	62
V	MOPC-104E	MOPC-104E	216
		HOPC-2020	192
V	HOPC-2020	MOPC-104E	228
		HOPC-2020	210
V	none	MOPC-104E	62
		HOPC-2020	61

Mouse embryo DNA (7 mg) was digested with *Eco*R1 and fractionated on a 0.9% agarose gel. DNA fragments containing C_λ and V_λ genes were first hybridized to an excess of λ mRNA in 6 × SSC/50% formamide at 45°C $C_0t = 4.2 \times 10^{-2}$; $C_0te = 10^2$. The hybrid was treated with 0.6 kU/ml of S1 nuclease in 0.3 M NaCl, 0.05 M NaAc, 1 mM $ZnSO_4$ pH 4.6 for 30 min at 35°C. RNA was digested with alkali. The second hybridization was to ^{125}I-labeled λ mRNA (≈ 5×10^7 cpm/μg RNA) and was carried out at 70°C in 2 × SSC under DNA excess conditions. Input ^{125}I-labeled RNA was 1500 cpm. The hybrid was then treated with RNase-mix (RNase A 20 μg/ml + T_1 RNase 2 units/ml) at 37°C for 20 min and trichloroacetic acid (TCA)-precipitated. Data represent the average of two measurements.

2020 λ mRNA, under conditions of DNA excess. The relative extent of hybridization in the second step was compared in all four possible combinations both for the C-gene and V-region DNA fragments (Table 1). The fact that in each of the two heterologous combinations the extent of hybridization is indistinguishable from that of the corresponding homologous combination confirms our previous contention that sequences in the two λ mRNAs are extensively homologous in the V regions. Analogous experiments with the C-gene fragments serve as a positive control.

Hybridization Kinetics of Individual DNA Components

Two points were experimentally tested with the hope of strengthening our previous conclusion that the myeloma V_λ gene is unique or nearly so. First, are the V_λ gene sequences any more reiterated than the C_λ gene sequences? Second, what is the resolution of the kinetics of hybridization? Can we, for instance, really distinguish 10 to 20 copies from a few copies?

To these ends, a large amount (30 mg) of embryo DNA was digested by *Eco*R1, and DNA fragments were fractionated in preparative agarose gel electrophoresis followed by sonication. Fractions containing the C segment, V segment, and 5′E segment were pooled from several gels, each of which displayed a hybridization pattern very similar to that in Figure 4. To make the results obtained with these DNA fractions and those obtained with unfractionated DNA comparable, a parameter of enrichment of a specific sequence, e, was defined as follows: $e = (a \cdot b \cdot I)/S$, where I is the amount of input DNA subjected to electrophoresis, S is the amount of DNA in the pooled fractions containing specific sequences, a is the overall recovery of DNA during electrophoresis and subsequent extraction from the gel, and b is the recovery of the specific sequences in the pool. In practice, a is obtained by dividing the sum of recorded O.D. from all gel fractions by the input O.D.; b is obtained by dividing the area under the hybridization peak corresponding to pooled fractions by total peak area. Since, as described elsewhere (Hozumi and Tonegawa 1976), the hybridization was carried out under conditions in which the final level increases sublinearly as the amount of the sequences increase, the value of b obtained by the above procedures is a lower limit.

A series of hybridization curves obtained with various fractionated DNAs as well as with unfractionated DNA are illustrated in Figure 8. The source of λ mRNA was MOPC-104E. Values of e were calculated for each of the fractionated DNA, and the hybridization level is plotted against "C_0te," which is C_0t corrected for sequence enrichment. The final levels of hybridization attained are different for each DNA component and are in good agreement with the relative lengths of the regions depicted in Figure 1. For easier comparison, the data in Figure 8 are replotted in Figure 9 with

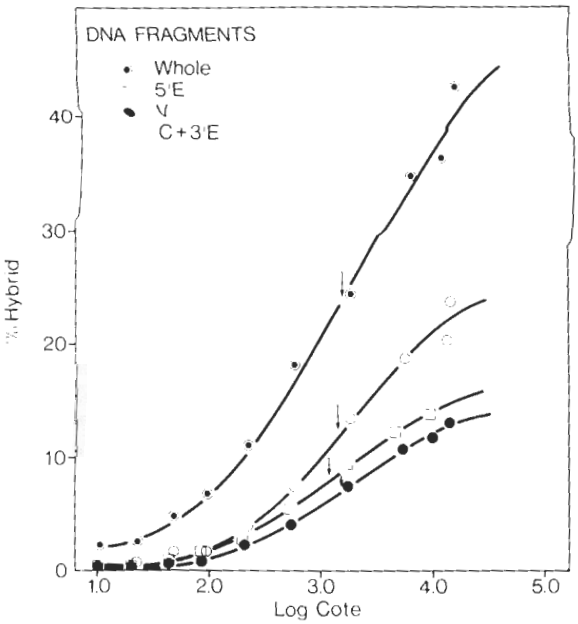

Figure 8. Kinetics of hybridization of ^{125}I-labeled MOPC-104E λ-chain mRNA with whole or fractionated embryo DNA. Hybridization was carried out in 2 × SSC supplemented with Tris-HCl pH 7.0, 0.01 M at 70°C. Other procedures for hybridization and processing of hybridization mixture were as described previously (Tonegawa et al. 1974b). A single preparation of ^{125}I-labeled MOPC-104E mRNA (specific activity 4×10^7 cpm/μg) was used in all experiments. Intrinsic RNase-resistant counts (2.0% of input) are subtracted. The following numbers are listed for whole DNA, 5'E-containing fragments, V-containing fragments, and C + 3'E-containing fragments in this order. DNA concentration (C_0) was 22.0, 2.2, 2.8, and 3.1 mg/ml. Enrichment factor (*e*) was 1, 9.4, 7.4, and 6.5. Ratio of DNA to RNA was 2.2×10^7, 2.2×10^6, 2.3×10^6, and 2.7×10^6. Arrows indicate the positions of $C_0te_{1/2}$ which were determined taking the maximum hybridization levels as 48%, 16%, 14%, and 24%.

multiple ordinate scales. It is clear from Figure 9 that the points obtained with the three fractionated DNAs fall virtually on a single sigmoid curve with $C_0te_{1/2} = 1600$. This corresponds to a nominal reiteration frequency of 3 (Melli et al. 1971). The points obtained with unfractionated DNA also fall on this curve, except for a possible slight deviation of those at lower C_0te values. The MOPC-104E RNA used here was not subjected to the acrylamide electrophoresis step. The slight deviation is therefore most likely attributable to impurities whose genes are reiterated (Rabbitts et al. 1975; Tonegawa 1976; Farace et al. 1976). These results demonstrate unequivocally that V_λ genes are no more reiterated than C_λ genes, a conclusion which is in agreement with our previous results using highly purified whole λ mRNAs. The results, also in conformity with our previous conclusions, indicate that both genes are reiterated no more than a few times per haploid genome.

In Figure 10, two C_0t curves obtained with two independent preparations of the C-gene fraction are compared with a C_0t curve obtained with unfractionated DNA. Note that, unlike in Figures 8 and 9, C_0t values are not corrected for the *e* factor, which was 5.8 and 6.5 for the two C-gene fractions. Whereas the points corresponding to the two C-gene fractions fall virtually on a single curve, those corresponding to unfractionated DNA are clearly *shifted to higher* C_0t values. These results demonstrate that, with the method employed, we can distinguish two genes whose reiteration frequencies differ by at least sixfold. In fact, a threefold difference in gene reiteration frequency, that is one-half of the shift observed in Figure 10, would seem to be distinguishable.

Weigert and colleagues have determined the amino acid sequences of a number of BALB/c and NZB mouse myeloma λ chains. They have already found nine different V_λ-region sequences (Weigert and Riblet, this volume; M. Weigert, pers. comm.). Most of these V_λ-region sequences are as homologous to that of MOPC-104E as is HOPC-2020, suggesting that the corresponding V genes cross-hybridize with MOPC-104E λ mRNA. We have shown above that the number of germ-line V_λ genes is no more than a few, using a method which can distinguish a threefold difference in gene reiteration frequency. Thus the experimentally determined gene number seems to be smaller than the number of different V_λ regions we already know. Furthermore, a simple statistical calculation which can be applied to the sequence data and is described elsewhere (Tonegawa 1976) indicates that the repertoire of mouse V_λ regions must be much larger than the nine known sequences. The results thus strongly suggest the existence of somatic processes by which V_λ-region sequences are diversified from a few germ-line genes.

The Existence of Multiple V Genes

The hybridization studies described in the last few sections demonstrated that groups of closely related V regions are somatically generated from a few germ-line genes. They do not, however, give us any reliable estimate of the *total* number of germ-line V genes. Given the enormous diversity of V regions, however, the existence of multiple germ-line genes is a highly likely possibility. This notion is also suggested by the genetic analysis of idiotypic markers on mouse heavy chains (Eichmann 1973).

A direct demonstration of multiple V genes for mouse κ chains is illustrated in Figure 11, which shows hybridization patterns obtained with mouse embryo DNA and whole κ RNAs from MOPC-21 and MOPC-321. These two κ chains show little homology in their V regions (Milstein and Svasti 1971; McKean et al. 1973)—they belong to two

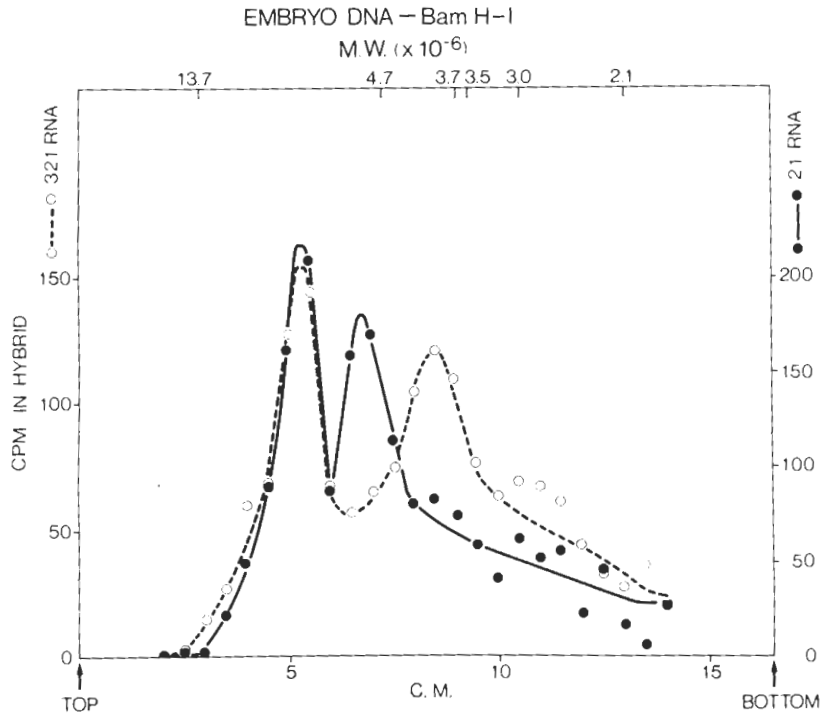

Figure 11. Gel electrophoresis patterns of embryo DNA. DNA was digested with *Bam*H1 and electrophoresed as described in the legend to Fig. 2. Hybridization was carried out with two κ-chain mRNAs. Specific activities were 5×10^7 cpm/μg for MOPC-321 κ-chain mRNA and 8×10^7 cpm/μg for MOPC-21 κ-chain mRNA.

in the germ line. On the other hand, there exist multiple germ-line V genes. We would like to know, of course, the total number of germ-line V genes.

Amino acid sequences of immunoglobulin-chain V regions can be classified into subgroups such that the variability within a subgroup is less than in the group as a whole. Obviously, the greater the number of subgroups, the less variability within the subgroups. The precise degree of variability tolerated within a subgroup is usually decided rather arbitrarily. Cohn et al. (1974), however, have defined subgroups in a less arbitrary fashion: If two V regions differ by even one "frame" residue (i.e., a residue outside the hypervariable regions), they belong to different subgroups. According to Cohn, all known V regions associated with the λ_I chain belong to a single subgroup. We have shown that there is probably only one germ-line V gene for λ_I. MOPC-321 and MOPC-21 κ chains belong to different subgroups, and we have shown that their V_κ genes are derived from different germ-line genes. It will be interesting to see how far this correlation between genes and subgroups is maintained. We think that the number of subgroups a la Cohn is an upper limit for the number of germ-line V genes. If this is so, the mouse genome contains from 2 to 100 V_κ genes.

CONCLUDING REMARKS

The dynamic nature of the genetic material is an essential element of evolution. Genes can mutate and recombine. If they could not, there would be no higher forms of life. Both of these processes are under genetic control and are themselves subject to natural selection, i.e., populations will acquire optimal rates of mutation and recombination. But does the collection of cell populations which we call a multicellular organism utilize these processes in any meaningful way? In other words, do they play an important role during the development of a single individual? We have shown that this is the case for the immune system. This is not, however, a unique example. Some vertebrate tumor viruses appear to behave like temperate bacteriophages and are perhaps the most primitive examples. The behavior of genes for ribosomal RNA is the hitherto most sophisticated example. Note that these are all examples of *contextual* changes. We do not know of any cases involving changes of genetic *content*. But neither do we know of any system which is as complex as the immune system except one—the nervous system. Jerne (1973) has already pointed out the striking analogy between these two systems.

Acknowledgments

We acknowledge the expert technical assistance offered by Mr. G. R. Dastoornikoo, Ms. S. Stutz and Ms. B. Durrer.

REFERENCES

BERNARDINI, A. and S. TONEGAWA. 1974. Hybridization studies with an antibody heavy chain mRNA. *FEBS Letters* **41:** 73.

BROWNLEE, G. G., E. M. CARTWRIGHT, N. J. COWAN, J. M. JARVIS and C. MILSTEIN. 1973. Purification and se-

quence of messenger RNA for immunoglobulin light chains. *Nature New Biol.* **244:** 236.

CAPRA, J. D., and J. M. KEHOE. 1974. Antibody diversity: Is it all coded for by the germ line genes? *Scand. J. Immunol.* **3:** 1.

COHN, M. 1968. Natural history of the myeloma. *Cold Spring Harbor Symp. Quant. Biol.* **32:** 211.

COHN, M., B. BLOMBERG, W. GECKELER, W. RASCHKE, R. RIBLET and M. WEIGERT. 1974. First order considerations in analyzing the generation of diversity. In *The immune system: Genes, receptors, signals* (ed. E. E. Sercarz et al.), p. 89. Academic Press, New York.

DELOVITCH, T. and C. BAGLIONI. 1974. Immunoglobulin genes: A test of a somatic vs. germ line hypothesis by RNA/DNA hybridization. *Cold Spring Harbor Symp. Quant. Biol.* **38:** 739.

DREYER, W. J., W. R. GRAY and L. HOOD. 1968. The genetic molecular and cellular basis of antibody formation: Some facts and a unifying hypothesis. *Cold Spring Harbor Symp. Quant. Biol.* **32:** 353.

DUGAN, E. S., R. A. BRADSHAW, E. S. SIMMS and H. N. EISEN. 1973. Amino acid sequence of the light chain of a mouse myeloma protein (MOPC-315). *Biochemistry* **12:** 5400.

EICHMANN, K. 1973. Idiotype expression and the inheritance of mouse antibody clones. *J. Exp. Med.* **137:** 603.

FARACE, M. G., M. F. AELLEN, P. A. BRIAND, C. H. FAUST, P. VASSALLI and B. MACH. 1976. No detectable reiteration of genes coding for mouse MOPC41 immunoglobulin light-chain mRNA. *Proc. Nat. Acad. Sci.* **73:** 727.

FUKE, M. and C. A. THOMAS, JR. 1970. Isolation of open-circular DNA molecules by retention in agar gels. *J. Mol. Biol.* **52:** 395.

GALLY, J. A. and G. M. EDELMAN. 1970. Somatic translocation of antibody genes. *Nature* **227:** 341.

GREENE, P. J., M. C. BETLACH and H. W. BOYER. 1974. The *E. coli* RI restriction endonuclease. In *Methods in molecular biology* (ed. R. B. Wickner), vol. 7, p. 88. Marcel Dekker, New York.

GROSS-BELLARD, M., P. OUDET and P. CHAMBON. 1973. Isolation of high-molecular-weight DNA from mammalian cells. *Eur. J. Biochem.* **36:** 32.

HONJO, T., S. PACKMAN, D. SWAN, M. NAU and P. LEDER. 1974. Organization of immunoglobulin genes: Reiteration frequency of the mouse chain constant region gene. *Proc. Nat. Acad. Sci.* **71:** 3659.

HOZUMI, N. and S. TONEGAWA. 1976. Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Nat. Acad. Sci.* **73:** 3632.

JERNE, N. K. 1973. Structural analogies between the immune system and the nervous system. In *1st Aharon Kazir-Katchalsky Conference on the Stability and Origin of Biological Information* (ed. I. R. Miller), p. 201. John Wiley and Sons, New York.

KABAT, D. 1972. Gene selection in hemoglobin and in antibody-synthesizing cells. *Science* **175:** 134.

LEDER, P., T. HONJO, D. SWAN, S. PACKMAN and B. NORMAN. 1975. The organization and diversity of κ and λ genes. In *Miami Winter Symposia: Molecular approaches to immunology* (ed. E. E. Smith and D. W. Ribbons), vol. 9, p. 173. Academic Press, New York.

LEDER, P., T. HONJO, S. PACKMAN, D. SWAN, M. NAU and B. NORMAN. 1974. The organization and diversity of immunoglobulin genes. *Proc. Nat. Acad. Sci.* **71:** 5109.

MACH, B., C. FAUST and P. VASSALLI. 1973. Purification of 14S messenger RNA of immunoglobulin light chain that codes for a possible light-chain precursor. *Proc. Nat. Acad. Sci.* **70:** 451.

MÄKELÄ, O. 1968. The specificity of antibodies produced by single cells. *Cold Spring Harbor Symp. Quant. Biol.* **32:** 423.

MCKEAN, D., M. POTTER and L. HOOD. 1973. Mous immunoglobulin chains. Pattern of sequences variation among kappa chains with limited sequence differences. *Biochemistry* **12:** 760.

MELLI, M., C. WHITFIELD, K. V. RAO, M. RICHARDSON and J. O. BISHOP. 1971. DNA-RNA hybridization in vast DNA excess. *Nature New Biol.* **231:** 8.

MILSTEIN, C. and J. SVASTI. 1971. Expansion and contraction in the evolution of immunoglobulin gene pools. In *Progress in immunology* (ed. E. Amos), vol. I, p. 33. Academic Press, New York.

MILSTEIN, C., G. G. BROWNLEE, E. M. CARTWRIGHT, J. M. JARVIS and N. J. PROUDFOOT. 1974. Sequence analysis of immunoglobulin light chain messenger RNA. *Nature* **252:** 354.

OHNO, S. 1974. Aneuploidy as a possible means employed by malignant cells to express recessive phenotypes. In *Chromosomes and cancer* (ed. J. German), p. 77. Wiley, New York.

PERNIS, B., G. CHIAPPINO, A. S. KELUS and P. G. H. GELL. 1965. Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *J. Exp. Med.* **122:** 853.

PREMKUMAR, E., M. SHOYAB and A. R. WILLIAMSON. 1974. Germ line basis for antibody diversity: Immunoglobulin V_H- and C_H-gene frequencies measured by DNA-RNA hybridization. *Proc. Nat. Acad. Sci.* **71:** 99.

RABBITTS, T. H., J. M. JARVIS and C. MILSTEIN. 1975. Demonstration that a mouse immunoglobulin light chain messenger RNA hybridizes exclusively with unique DNA. *Cell* **6:** 5.

RABBITTS, T. H., J. O. BISHOP, C. MILSTEIN and G. G. BROWNLEE. 1974. Comparative hybridization studies with an immunoglobulin light chain mRNA fraction and non-immunoglobulin mRNA of mice. *FEBS Letters* **40:** 157.

RAFF, M. C., M. FELDMANN and S. DE PETRIS. 1973. Monospecificity of bone marrow derived lymphocytes. *J. Exp. Med.* **137:** 1024.

SCHECHTER, I. 1973. Biologically and chemically pure mRNA coding for a mouse immunoglobulin. L-chain prepared with the aid of antibodies and immobilized oligothymidine. *Proc. Nat. Acad. Sci.* **70:** 2256.

SCHULLER, R. 1976. Purification of immunoglobulin heavy chain mRNA's and their use in hybridization studies. M. S. thesis, University of Alberta, Edmonton, Canada.

STAFFORD, D. W. and D. BIEBER. 1975. Concentration of DNA solutions by extraction with 2-butanol. *Biochim. Biophys. Acta.* **378:** 18.

STAVNEZER, J., R. C. C. HUANG, E. STAVNEZER and J. M. BISHOP. 1974. Isolation of messenger RNA for an immunoglobulin kappa chain and enumeration of the genes for the constant region of kappa chain in the mouse. *J. Mol. Biol.* **88:** 43.

STORB, U. 1974. Evidence for multiple immunoglobulin genes. *Biochem. Biophys. Res. Comm.* **57:** 31.

TONEGAWA, S. 1976. Reiteration frequency of immunoglobulin light chain genes. Further evidence for somatic generation of antibody diversity. *Proc. Nat. Acad. Sci.* **73:** 203.

TONEGAWA, S. and I. BALDI. 1973. Electrophoretically homogeneous myeloma light chain mRNA and its translation *in vitro*. *Biochem. Biophys. Res. Comm.* **51:** 81.

TONEGAWA, S., A. BERNARDINI, B. J. WEIMANN and C. STEINBERG. 1974a. Reiteration frequency of antibody genes. Studies with κ chain mRNA. *FEBS Letters* **40:** 92.

TONEGAWA, S., C. STEINBERG, S. DUBE and A. BERNARDINI. 1974b. Evidence for somatic generation of antibody diversity. *Proc. Nat. Acad. Sci.* **71:** 4027.

WARNER, T. F. C. S. and R. G. KRUEGER. 1975. The multi-

cellular (as opposed to the monoclonal) origin of the murine myeloma cell. *J. Theoret. Biol.* **54:** 175.

WEIGERT, M. G., I. M. CESARI, S. J. YONKOVICH and M. COHN. 1970. Variability in the lambda light chain sequences of mouse antibody. *Nature* **228:** 1045.

WIGZELL, H. 1973. Antibody diversity: Is it all coded for by the germ line genes? *Scand. J. Immunol.* **2:** 199.

WILSON, G. A. and F. E. YOUNG. 1975. Isolation of a sequence-specific endonuclease (Bam I) from *Bacillus amyloloquefaciens* H. *J. Mol. Biol.* **97:** 123.

YOSHIDA, T. H., H. T. IMAI and M. POTTER. 1968. Chromosomal alteration and development of tumors. XIX. Chromosome constitution of tumor cells in 16 plasma cell neoplasms of BALB/c mice. *J. Nat. Cancer Inst.* **41:** 1083.