Benvenuto Pernis and Henry J. Vogel (eds.) (27)

CELLS OF IMMUNOGLOBULIN SYNTHESIS

New York, Academic Press, 1979, pp. 15-32.

Somatic Recombination and Structure of an Immunoglobulin Gene

SUSUMU TONEGAWA, CHRISTINE BRACK,
MINORU HIRAMA, AND
RITA LENHARD-SCHULLER
Basel Institute for Immunology
Basel, Switzerland

INTRODUCTION

Is the organization of DNA sequences altered in cell differentiation? If so, an obvious implication is that such a somatic change might play a key role in cell determination. One of the best systems to which this question can be addressed is that of immunoglobulin genes. Twelve years ago, Dryer and Bennett proposed that a single immunoglobulin chain is encoded in two separate DNA segments, one for the amino terminal half (V region) and the other for the carboxyl terminal half (C region), and that synthesis of a complete chain is preceded by rearrangement of the DNA segments (1). The hypothesis was extended by Gally and Edelman to incorporate a mechanism for somatic generation of the V region diversity (2). Direct investigation into the issue became possible upon discovery of bacterial restriction enzymes. Using a restriction enzyme to digest total cellular DNA of embryo and myeloma cells, Hozumi and Tonegawa produced experimental results that were compatible with the Dryer and Bennett hypothesis (3). Development of the *in vitro* recombinant technique opened a way for further studies on this problem. This paper deals with isolation and characterization of DNA clones containing mouse λ type light chain genes. The study

ISBN 0-12-551850-1

not only provided a direct evidence for somatic recombination but also revealed a surprising feature of the structure of an immunoglobulin gene.

SCHEME FOR ISOLATION OF Ig GNES FROM CELLULAR DNA

Various steps involved in the gene cloning experiment are schematically illustrated in Fig. 1. We first digested highly polymerized total cellular DNA with restriction endonuclease *Eco*RI and fractionated the resulting DNA fragments by electrophoresis in a slab agarose gel. We then identified the DNA fragments carrying specific Ig gene sequences by hybridization. This step was carried out in two different

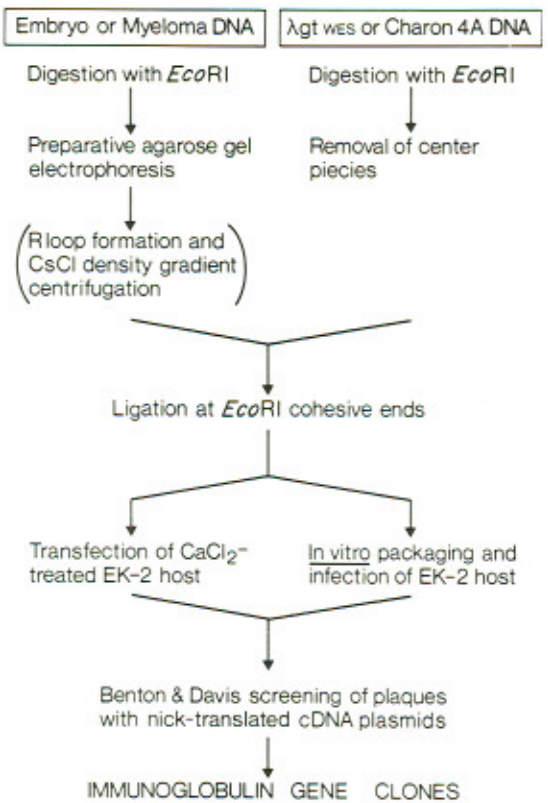

Fig. 1. Scheme for isolation of immunoglobulin genes. See text for explanation.

ways. Earlier, we had extracted DNA fragments from the gel slices and carried out hybridization in liquid using ^{125}I-labeled, purified Ig mRNA (3). More recently, we synthesized complementary DNA on the purified mRNA template by reverse transcriptase and constructed a plasmid composed of pCRI and Ig DNA sequences. Using chimeric plasmids as the hybridization probes, we applied the blotting technique developed by Southern (4) to a strip of slab gel (5). This procedure obviated the time-consuming extraction step necessary in the liquid hybridization method. Results of the Southern gel blotting of embryo and myeloma DNA's are illustrated in Fig. 2.

The gel electrophoresis step gave 10- to 100-fold enrichment for various Ig gene-positive fragments. In some cases, we further enriched the relevant DNA fragments by incubating the duplex DNA with excess Ig mRNA in 70% formamide and subjected the mixture to equilibrium centrifugation in a CsCl gradient. Since RNA–DNA hybrids are more stable than DNA duplexes in aqueous formamide under certain conditions, one can construct a structure in which duplex DNA is partially denatured and one of the two strands is replaced by complementary RNA (R-loop stucture) (6). Such hybrid molecules are denser than either double- or single-stranded DNA, and therefore can be separated by equilibrium density gradient centrifugation (7). Efficiency of enrichment depends largely on the relative size of the DNA fragments and mRNA molecules. For 4 kb DNA fragments and ~1 kb λ chain mRNA, we obtained enrichment of about fortyfold.

As the cloning vector we used either λgt_{WES} (8) or Charon 4A (9). These phages were attenuated by several genetic tricks to be used specifically in cloning of eukaryotic DNA fragments (EK-2 vectors). When digested with *Eco*RI, phage DNA's generate left and right arms as well as one or two DNA fragment(s) that originate from the center section of the genome and are dispensible for phage growth. The center pieces were removed by agarose gel electrophoresis. Vector and mouse DNA were mixed and ligated with T4-ligase at the *Eco*RI cohesive ends.

We used two different methods in generating phage plaques from recombinant DNA molecules. The first method is to transfect a $CaCl_2$-treated *Escherichia coli* with the naked recombinant DNA molecules (10). The second method is to package the DNA molecules *in vitro* into phage λ coats and to plate viable phage particles on an ordinary *E. coli* K12 (11).

Identification of the Ig gene sequence-positive plaques was carried out by the replica method developed by Benton and Davis (12). A plaque plate was covered with a dry nitrocellulose membrane filter to

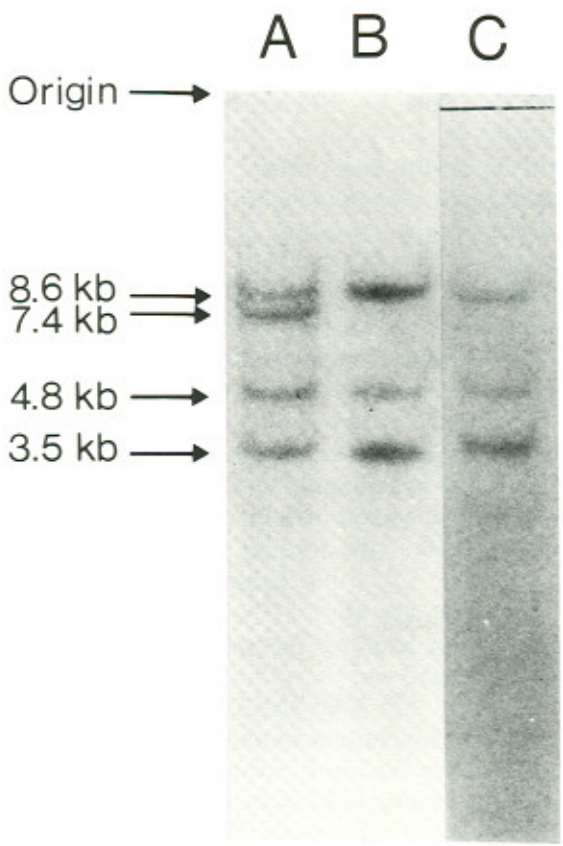

Fig. 2. λ_1 Gene sequence-containing DNA fragments in embryo and myeloma cells. High molecular weight DNA's extracted from 13-day-old BALB/c embryos (B), myelomas HOPC 2020 (a λ_1 chain producer) (A), and MOPC 321 (a κ chain producer) (C) were digested to completion with *Eco* RI, electrophoresed on a 0.9% agarose gel, transferred to nitrocellulose membrane filters, and hybridized with a nicktranslated *Hha* I fragment of the plasmid B1 DNA. This DNA fragment is 2.5 kb long and contains the full (or near full) λ_1 gene sequence (M. Hirama, G. Matthyssens, and S. Tonegawa, unpublished).

absorb the phage particles and unpackaged phage DNA in the plaques. The filter was briefly dipped in an alkali solution to disrupt the phage particles and also to denature DNA, neutralized in Tris buffer, and baked in a vacuum oven to fix the DNA. The filter was then

incubated under proper conditions with a radiolabeled hybridization probe (as the probe we used ^{125}I-labeled mRNA or nicktranslated plasmid DNA's containing κ or λ gene sequences), washed extensively, and subjected to autoradiography. Phages in the plaques that gave positive autoradiographic spots were picked, purified, and propagated in an EK-2 host, *E. coli* DP50 SupF$^+$ (8).

The efficiency of the *in vitro* packaging method is at least one, and usually two, orders of magnitude greater than that of transfection (11). We now obtain over 10^5 plaques by packaging 0.1 μg of mouse DNA fragments (about 10 kb in length, ligated with 0.3 μg of Charon 4A or λgt_{WES} DNA arms) in a standard 40 μl packaging mixture prepared from 10 ml each of the two heat-induced lysogen cultures. This scale of packaging is sufficient for the cloning of a unique mammalian gene, if the DNA preparation used is ten- to twentyfold enriched for the particular gene sequence. The simple preparative agarose gel electrophoresis, combined with Southern gel blotting using a nicktranslated cDNA clone probe, easily provides the necessary enrichment. The plaque screening method developed by Benton and Davis is easy, fast, and reliable. With a little care for controlling various factors that affect the plaque size (time of incubation, humidity of the agar plates), one can obtain as many as 10^5 distinct plaques on a single 20 × 20 cm plate, and thereby process as many plaques for *in situ* hybridization on a single sheet of membrane filter (19 × 19 cm). In summary, the combined use of preparative agarose gel electrophoresis, *in vitro* packaging, and the Benton and Davis plaque screening technique facilitates cloning of essentially any unique mammalian gene, for which hybridization probes are available, by the handling of a few micrograms of total cellular DNA, less than 100 μl of the packaging mix, and one or two large agar plate(s). Because the entire experiment is done on a relatively small scale, it is easy to contain and therefore reduce the chances of accidental escape of hypothetically hazardous clones.

DNA CLONES ISOLATED FROM EMBRYO AND MYELOMA CELLS

We have applied the gene cloning procedures described in the last section to the various DNA components from BALB/c embryos (12 days old) and from both κ-chain secreting (MOPC 321) and λ-chain secreting (HOPC 2020 and J 558) myelomas. In the present paper we

TABLE I
List of λ Chain Gene Clones

Clones	DNA source	DNA pre-enrichment steps	Approximate[a] pre-enrichment factors	Screening procedures and probes	Approximate number of plaques screened	λ Gene[b] sequences contained	References
Ig 99λ	Embryo 3.5 kb	Agarose gel R-looping (one cycle)	300	Benton and Davis (12) with nick-translated cDNA plasmid	3,000	$V_{\lambda I}$	This paper
Ig 25λ	Embryo 8.6 kb	Agarose gel	15		80,000	$C_{\lambda I}$	This paper
Ig 303λ	HOPC 2020 7.4 kb	Agarose gel	15		70,000	$V_{\lambda I} + C_{\lambda I}$	Brack and Tonegawa (16)
Ig 13λ	Embryo 4.8 kb	Agarose gel R-looping (two cycles)	360	Kramer *et al.* (15) with [^{125}I]λ_I mRNA	4,000	$V_{\lambda II}$	Tonegawa *et al.* (7)

[a] See Tonegawa *et al.* (14) for the definition of the enrichment factor.
[b] This column lists the λ gene-sequences assigned to the cloned DNA's. See text for more details.

shall restrict ourselves to description and discussion of λ gene clones. Four types of λ gene clones were isolated. Three clones, Ig 25λ, Ig 13λ and Ig 99λ, are from the three DNA components visualized by Southern gel blotting of embryo DNA, while the fourth clone is from the DNA component (7.4 kb) that is present only in the myeloma (Fig. 2 and Table I).

TYPE OF λ GENE SEQUENCES CONTAINED IN THE DNA CLONES

In order to find out what kinds of λ gene sequences (V_λ, $C_{\lambda I}$, or V_λ plus $C_{\lambda I}$) are contained in the isolated clones, we carried out gel blotting (Southern) experiments with *Eco*RI digested, cloned DNA, using three different λ gene sequences probes. The first probe was the plasmid clone B1 that contains essentially the whole sequence of a λ_I gene. The second probe was a 470 base pair DNA fragment that was excised from the $V_{\lambda II}$-carrying Ig 13λ DNA by restriction endonucleases *Hae*III and *Mbo*II (13). Since $V_{\lambda I}$ and $V_{\lambda II}$ gene sequences are extensively homologous, this DNA fragment serves as a probe for both $V_{\lambda I}$ and $V_{\lambda II}$ gene sequences. The third probe was an approximately 400 base long cDNA that was synthesized on a purified HOPC 2020 λ_I mRNA using the oligo (dT_{12-18}) primer and was isolated by acrylamide gel electrophoresis in 98% formamide. Because of the specificity of the priming activity of the oligo(dT), and the size of the cDNA, this probe mostly contains $C_{\lambda I}$ gene sequences.

When digested with *Eco*RI, the Ig 303λ, Ig 25λ, Ig 99λ, and Ig 13λ DNA generated, in addition to the left (21.5 kb) and right (14 kb) arms of the phage DNA, fragments of 7.4, 8.6, 3.5, and 4.8 kb, respectively. The sizes of these DNA fragments are in good agreement with those assigned to the respective DNA fragments that were visualized by the gel blotting of the total cellular DNA. These DNA fragments all hybridized with the plasmid B1 probe. The 8.6 kb, Ig 25λ fragment hybridized with the $C_{\lambda I}$ probe but not with the V_λ probe. Conversely, the 3.5 kb, Ig 99λ fragments hybridized with the V_λ probe but not with the $C_{\lambda I}$ probe. The 7.4 kb Ig 303λ fragment hybridized with both the V_λ and the $C_{\lambda I}$ probes, while the 4.8 kb, Ig 13λ fragment hybridized with the V_λ probe but not with the $C_{\lambda I}$ probe. In addition, our current nucleotide sequencing studies demonstrated that the V_λ sequence contained in Ig 99λ and Ig 303λ are of the λ_I type (N. Hozumi, O. Bernard,

and S. Tonegawa, unpublished observations). The assigned λ gene sequences are listed in Table I.

LOCATION OF THE λ CHAIN GENE SEQUENCE IN THE DNA CLONES

The position of λ chain gene sequences in the DNA clones was determined by R-loop mapping (6). The clones Ig 303λ DNA was incubated with HOPC 2020 λ chain mRNA under the conditions for R-loop formation. Upon examination in the electron microscope, more than 50% of the molecules displayed DNA–RNA hybrid regions. However, we did not observe the single R-loop of about 1000-nucleotide length that would have been expected if the mRNA had hybridized with a stretch of DNA corresponding to contiguous V and C gene sequences. Instead, two small R-loops (460 and 380 base long) separated by a double-stranded DNA loop (1.2 kb) were observed (Fig. 3). The interpretation of such a hybrid structure is that one RNA molecule is annealed to two stretches of DNA that are separated by a duplex DNA region. In many cases the hybrid segment or R-loop generated by the longer homology had a short whisker (50–100 nucleotides) at its left end. The whisker is probably the poly(A) at the 3′ end of the mRNA. This suggests that the longer homology is composed of the C gene sequence and that the shorter one is composed of the V gene sequence. The validity of this assumption was confirmed by analysis of the heteroduplexes formed between the Ig 303λ DNA and a V_λ gene-carrying DNA fragment (16).

R-loop molecules formed with the 8.6 kb, Ig 25λ fragment displayed a double loop structure composed of a 410 nucleotide R-loop at 3.9 kb from one and, and a 1.2 kb double-stranded DNA loop (Fig. 3). This structure closely resembles the triple loop of Ig 303λ, except that the Ig 25λ hybrids have a long RNA tail (~260 bases) instead of the second, smaller R-loop. Because the Ig 25λ fragment showed homology only with $C_{\lambda I}$, but not with V_λ sequences (see above), we conclude that the 410 base pair R-loop contains the $C_{\lambda I}$ gene sequence, and that the long RNA tail corresponds to the 5′-end or V-coding part of the mRNA. A second, short RNA tail (~100 bases) observed sometimes at the other end of the R-loop would correspond to the poly(A) sequence at the 3′-end of the mRNA. The presence of the 1.2 kb DNA loop indicates that Ig 25λ DNA contains a short homology region that hybridizes to a region near the V–C junction of the mRNA molecule. This second homology, which we call the J sequence, is separated from the $C_{\lambda I}$

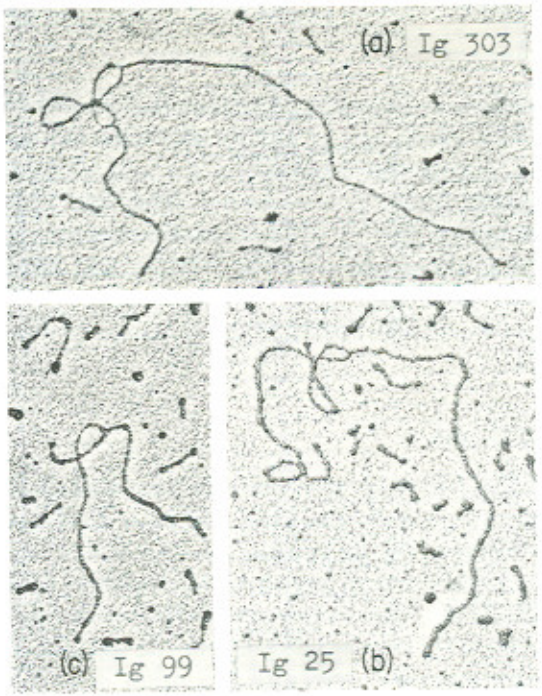

Fig. 3. R-loop molecules obtained by hybridizing HOPC 2020 λ_1 mRNA with the *Eco* RI fragments of the DNA clones. (a) Ig 303λ DNA showed two R-loops corresponding to the V and C genes that are separated by the double-stranded DNA loop of about 1.2 kb. The short tail at the end of one loop is the 3′ poly(A). (b) Ig 25λ DNA displays one R-loop corresponding to the C gene, the double-stranded DNA loop, and a long RNA tail corresponding to V sequences. The short tail observed in some molecules is the 3′ poly (A) tail. (c) Ig 99λ DNA has one R-loop corresponding to V sequences, and a long RNA tail that is composed of the C gene sequences plus poly(A) tail.

sequence by 1200 base pairs. It is too short to be visualized as a separate R-loop but is strong enough to hold the double loop structure together.

The 3.5 kb, Ig 99λ fragment formed a single R-loop very similar to the one observed in Ig 13λ (7). It is 380 nucleotides long, lies in the middle of the DNA fragment, i.e., 1.65–1.66 kb from either end, and carriers a ~340 nucleotide RNA tail at one end. Since this DNA fragment contains a $V_{\lambda I}$ sequence (see above), the R-loop should contain the $V_{\lambda I}$ sequence and the RNA tail should correspond to the $C_{\lambda I}$ sequence plus the poly(A).

SEQUENCE HOMOLOGY BETWEEN THE CLONED DNA FRAGMENTS

Analysis of sequence homology between the cloned DNA fragments, both within the λ chain genes and in the adjacent regions, may be helpful in discovering the mechanism by which somatic rearrangement of immunoglobulin genes occurs. The four cloned fragments were therefore hybridized in various combinations and the heteroduplex molecules analyzed by electron microscopy. We shall describe below the observations made with various combinations of the three λ_I DNA clones Ig 99λ, Ig 25λ, and Ig 303λ. A summary of all the results is given in Fig. 4.

Ig 25λ VERSUS Ig 303λ

Ig 25λ hybridized with Ig 303λ DNA to form Y-shaped heteroduplex molecules with two single-stranded and one double-stranded arms (Fig. 5). The lengths of the three arms are: 2.87 kb (long single strand), 1.98 kb (short single strand), and 5.47 kb (double strand). The measurements indicate that the shorter single-stranded arm and one strand of the double-stranded arm correspond to the Ig 303λ DNA, whereas the longer single-stranded arm and the other strand of the double-stranded arm derive from the Ig 25λ DNA. In the entire 5.5 kb homology region we observed no local mismatchings that would indicate partial nonhomology between the two strands. R-loop mapping had

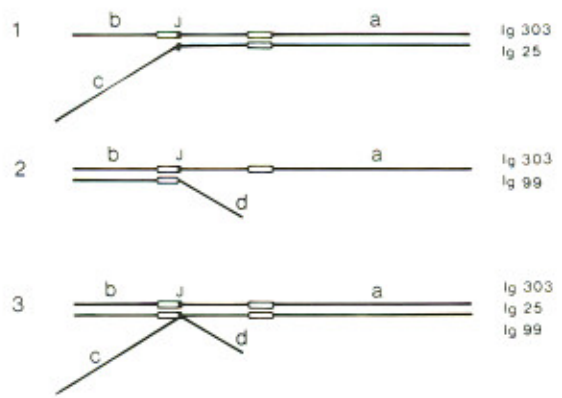

Fig. 4. Interpretation of the heteroduplex molecules. The position of V and C gene sequences (white boxes) was deduced from R-loop molecules.

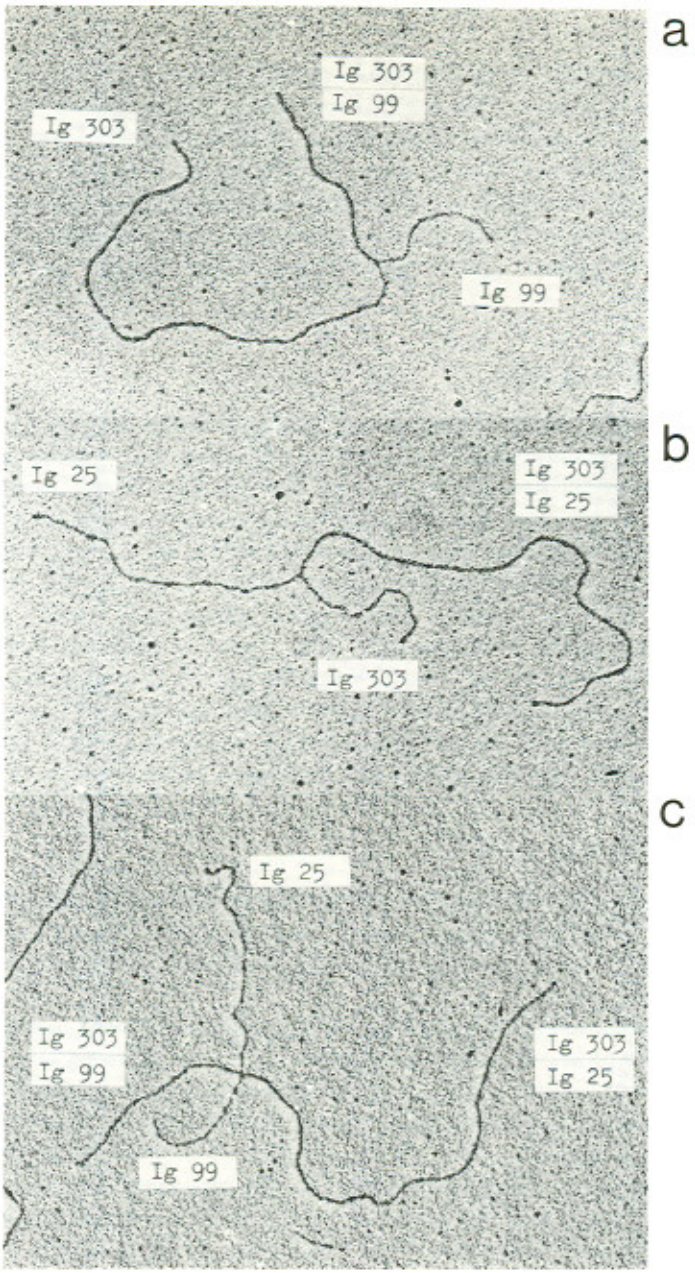

Fig. 5. Heteroduplex molecules formed by various combinations of the three λ_I sequence-containing cloned DNA fragments. (a) Myeloma DNA Ig 303λ versus embryo DNA Ig 99λ; (b) myeloma DNA Ig 303λ versus embryo DNA Ig 25λ; (c) combination of all three clones gave double heteroduplex structures.

shown that in both DNA fragments the $C_{\lambda I}$ gene lies between 3.8 and 4.2 kb from one end. The 5.5 kb homology region therefore contains the $C_{\lambda I}$ gene. In addition, most if not all of the 1.2 kb DNA segment separating the V and C sequences in the Ig 303λ fragment is highly homologous to the DNA segment of similar length that separates the J and C sequences in the Ig 25λ fragment. The measurements indicated that the J sequence lies very near or at the branch point of the heteroduplex.

Ig 99λ VERSUS Ig 303λ

These two fragments also formed Y-shaped heteroduplex molecules with two single-stranded (5.48 and 1.53 kb) and one double-stranded (1.98 kb) arms (Fig. 5). The sum of the lengths of the long single-stranded arm and the double-stranded arm corresponds to the length of the Ig 303λ fragment, whereas the short single strand and the second strand of the double-stranded region belong to the Ig 99λ fragment. Again, we did not observe any local nonhomology in the double-stranded part of the heteroduplex. These results indicate that on one end of each fragment, the two DNA's are highly homologous. The V_{λ} sequence lies between 1.66 and 2.0 kb from one end of both Ig 303λ and Ig 99λ DNA's. Therefore the V gene is within the 2.0 kb homology region and lies near the fork of the Y-shaped heteroduplex (Fig. 5).

Ig 303, Ig 25, AND Ig 99

The results described in the last two sections strongly suggest that a large part of the Ig 303 fragment (5.5 kb) is homologous to the Ig 25 DNA whereas the rest of the molecule (2 kb) is homologous to the Ig 99 DNA.

In order to confirm this, the three DNA fragments were mixed, denatured, and annealed. A low proportion of the molecules displayed a double heteroduplex structure or cruciform structure composed of the double-stranded and two single-stranded arms (Fig. 5). Measurements of the arms allowed us to assign each part of the hybrid to the three DNA fragments as illustrated in Fig. 4. The observed structure can be formed only if each of the two ends of the Ig 303 fragment is homologous to only one of the two DNA fragments Ig 99 or Ig 25. The two homology regions meet at a point which corresponds to the J sequence in Ig 25 and lies near the junction between the V sequence and the 1.2 kb intron of Ig 303 DNA. The myeloma DNA fragment, Ig 303, thus

seems to be entirely composed of DNA segments that are homologous to parts of the two embryonic fragments.

EVIDENCE FOR SOMATIC REARRANGEMENT OF IMMUNOGLOBULIN GENES

The heteroduplex analysis of the three λ_I DNA clones, combined with the gel blotting analysis of the total cellular DNA's demonstrated beyond doubt the occurrence of somatic rearrangements of immunoglobulin gene sequences. The double heteroduplex structure generated by co-annealing the three cloned λ_I DNA's is incompatible with the trivial alternative interpretation (3) of the results obtained by restriction enzyme mapping of total cellular DNA. The observed heteroduplex structures can not be artifacts of DNA cloning. The length of each of the cloned DNA fragments coincides well with that of the corresponding cellular DNA fragments visualized by the gel blotting technique. Furthermore, certain restriction enzyme sites identified in the cloned DNA's are also present at corresponding positions in cellular DNA (5).

The results reported here identified a single recombination site on each of the two embryonic DNA clones. These sites were visualized as the branch point of the Y-shaped heteroduplex molecules formed between the cloned myeloma DNA (Ig 303λ) and either of the two cloned embryonic DNA's (Ig 99λ and Ig 25λ). That the branch points correspond to a single site on the Ig 303λ DNA was suggested by the double heteroduplex structure, in which two single-stranded tails, one of Ig 99λ and the other of Ig 25λ, extend from a single site on the Ig 303λ DNA. We conclude that the three cloned λ_I DNA's are related by a single recombination event as illustrated in Fig. 6. Embryonic DNA recombines at the right end of the V sequence and the left end of the J sequence to generate the sequence arrangement present in myeloma DNA. The entire 1.2 kb intron in the Ig 303λ DNA originates from the Ig 25λ DNA. Measurements of various parts of the R-loops and heteroduplex structures described in the present work are entirely consistent with this model. Further support has come from our more recent nucleotide sequencing studies (O. Bernard, N. Hozumi, and S. Tonegawa, unpublished results). These studies revealed that the Ig 303 V DNA segment codes for the polypeptide chain consecutively for a length corresponding to a complete V-region as defined by amino acid sequence studies, whereas the Ig 99λ V DNA segment ceases to

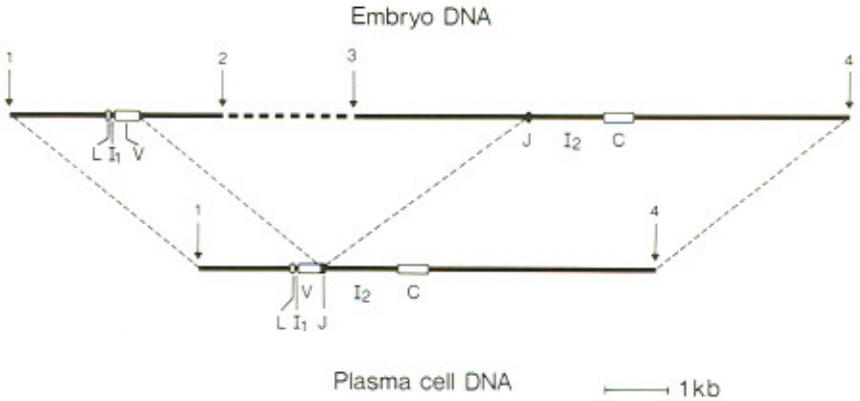

Fig. 6. Arrangement of mouse λ_I gene sequences in embryos and λ_I chain-producing plasma cells. In embryo DNA a full κ_I gene sequence is split into two parts that lie separately on two *Eco* RI fragments. On one, the coding sequence is further split into two parts, one for most of the leader peptides (L) and the other for the rest of the leader peptides plus the variable region peptides (V). The two coding sequences are separated by a 93 nucleotide long intron (I_1) (N. Hozumi, O. Bernard, and S. Tonegawa, unpublished observation). On the second *Eco* RI fragment the coding sequence is also split into two parts by a 1250 base long intron (I_2). The two parts are for the constant region peptides (C) and about 13 residue peptides near the junction of the variable and constant regions (J). The relative orientation of and the distance between the two *Eco* RI fragments are unknown. In the DNA of a λ_I chain-producing myeloma (HOPC 2020) the λ_I gene sequence is rearranged as a result of one or more recombination(s) that involve sequences in the two embryonic *Eco* RI fragments. One recombination takes place at the ends of the V and the J sequences and brings the two sequences directly in contact. The limits of the corresponding sequences in the embryo and the myeloma DNA's are indicated by thin dotted lines. The Figure is not intended to imply that the recombination results in deletion or looping-out of the embryonic DNA sequences that lie between the V and the *Eco* RI site 2, or between the *Eco* RI site 3 and the J. It is not intended to imply that the embryo and myeloma V sequences are identical. Additional short introns may be present in the C sequences. Arrows with numbers indicate *Eco* RI sites.

code at residue 97. This suggests that the nucleotides necessary to code for the extra amino acids at the end of the Ig 303λ V region are contributed by the J sequence on the Ig 25λ DNA. Our correct sequencing study of the J region confirmed this contention.

GENE IN PIECES

The R-loop mapping demonstrated the presence of a 1.2 kb intron both on the Ig 303λ and the Ig 25λ DNA. The resolution of cytochrome spreadings will not allow detection of introns shorter than about 100

nucleotides. Indeed, the 93 base long intron near the region corresponding to the amino terminal of the Ig 13λ DNA was revealed only by nucleotide sequence determination (13). Our recent nucleotide sequencing studies (O. Bernard, N. Hozumi, and S. Tonegawa) established that both Ig 303λ and Ig 99λ DNA contain an intron equivalent to that of Ig 13λ both in length and position. These findings are incorporated into Fig. 6. It should be added that additional short introns may be revealed in the C DNA sequence by the nucleotide sequencing now in progress in this laboratory. In any case, in each of the three λ_I DNA clones the protein-coding sequences are arranged in discrete pieces. For instance, the somatically rearranged, complete λ_I gene in the myeloma (Ig 303λ) consists of at least three DNA segments, one coding for the leader, one for the V region, and one for the C region. The introns separating the three coding segments are present in the original mouse DNA and are not introduced during the cloning procedure. The unique restriction enzyme cleavage sites that have been identified within the introns of the cloned DNA were also demonstrated in the uncloned cellular DNA (5).

Recent studies on other genes of eukaryotes (17–22), as well as their viruses (23–26) have revealed several cases of this unexpected gene structure: informational DNA interspersed with introns (silent sequences). Introns are probably transcribed together with the informational DNA and subsequently excised during maturation of pre-mRNA. One recent experiment concerning the mouse globin β chain gene seems to support this hypothesis (27). These observations have led us to propose an additional evolutionary pathway for creation of genes in higher organisms (13). By this pathway, a new gene can be created from two or more separate DNA segments upon emergence, at the boundary of the DNA segments, of mutations that generate signals for RNA splicing. If the new polypeptide chain coded by the spliced RNA has survival value, such mutations may be fixed in evolution. As the splicing does not have to be 100% efficient, creation of the new gene need not destroy the old: usually a disadvantageous event.

We assume that RNA splicing is an intramolecular reaction. This will restrict the operation of the gene creation mechanism described above to the space of a single transcription unit. One way to enlarge the effectiveness of this gene creation mechanism is to shuffle DNA segments by introducing them into transcription units. DNA sequences thus newly introduced into a transcription unit are then available for splicing with pre-existing sequences. Actual evolutionary use of such DNA sequences depends on the emergence of mutations lead-

ing to new splicing signals at the proper positions in the transcription unit. We hypothesize that many genes in higher organisms have arisen through such evolutionary processes.

SOMATIC REARRANGEMENT AS A MECHANISM FOR GENE CONTROL IN CELL DIFFERENTIATION

Does a higher organism utilize such a gene creation mechanism in the normal process of cell differentiation? We consider that the immunoglobulin genes are the perfect example. Here, the J sequence seems to play a key role by providing a "bridge" between DNA recombination and RNA splicing. Its left half most probably contains a nucleotide sequence for a site-specific recombination with the V DNA segment, while the sequence in the right half would almost certainly be involved in the RNA splicing event that connects the V and C sequences.

When a gene is created in this manner during ontogeny only in a particular subpopulation of cells composing an organism, the recombination itself can provide a novel mechanism for gene control in cell differentiation. The somatic rearrangement involving particular immunoglobulin gene sequences seem to be restricted to a small subpopulation of cells. Arrangement of a κ chain sequence in DNA's of several nonlymphatic adult tissues was identical to that of embryo DNA when analyzed by the electrophoresis–hybridization assay (28; also M. Hirama, unpublished observation). Analogous experiments carried out using DNA from a λ_I chain-producing myeloma and a κ sequence probe and *vice versa* indicated that there is a mutual exclusion in rearrangement of κ and λ chain DNA sequences (28). The results shown in Fig. 2 gave an additional example, namely that the 7.4 kb *Eco*RI fragment carrying the rearranged full λ_I gene sequence is absent in the DNA of κ-producing MOPC 321. It should be added that presence of the 7.4 kb fragment is not peculiar to HOPC 2020 myeloma DNA. DNA's from other λ_I-producing myelomas studied (MOPC 104E and J 558) also contained this fragment. [The $V_{\lambda I}$ regions synthesized by these myelomas are identical and differ from HOPC 2020 $V_{\lambda I}$ region by two residues (29), but all $V_{\lambda I}$ regions are believed to share an identical germ line V gene (30).] Conversely, absence of the 7.4 kb fragment is not peculiar to MOPC 321 DNA. DNA's from other κ chain-producing myelomas studied (TEPC 124 and MOPC 21) gave no such fragment (M. Hirama, unpublished observations).

On the molecular level one can conceive many variations of a gene control mechanism operated by somatic DNA rearrangement. For in-

stance, in the case of immunoglobulin genes, a V sequence-carrying DNA segment that has been transcriptionally silent may be excised and inserted into a constitutively active transcription unit containing the C sequence. Alternatively, the act of V DNA insertion itself may create a new promoter at the very site of insertion. As we previously discussed (16), a gene control mechanism directly dependent on somatic sequence rearrangement in DNA seems to fulfill most easily the "one lymphocyte clone–one light chain" rule of the immune system.

ACKNOWLEDGMENTS

We are very grateful to Dr. B. Hohn for invaluable help in the packaging experiments. We thank Dr. W. Arber, Dr. J. Carbon, Dr. R. Curtis, Dr. P. Leder, and Drs. K. and N. E. Murray for bacteria and phage strains. We also thank Dr. J. Beard, Dr. N. Hozumi, Dr. C. P. Kung, and Dr. J. Summers for enzymes, and Dr. G. Matthyssens who participated in the construction of the Bl plasmid. We are grateful to Mr. G. Dastoornikoo, Mr. A. Traunecker, and Mrs. R. Hiestand for their expert technical assistance.

REFERENCES

1. Dryer, W. J., and Bennett, J. C. (1965) *Proc. Natl. Acad. Sci. U.S.A.* **54**, 864–879.
2. Gally, J. A., and Edelman, G. M. (1970) *Nature (London)* **227**, 341–346.
3. Hozumi, N., and Tonegawa, S. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 3628–3232.
4. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–518.
5. Brack, C., Hirama, M., and Lenhard-Schuller, R. (1978) *Cell* **15**, 1–14.
6. White, R. L., and Hogness, D. S. (1977) *Cell* **10**, 177–192.
7. Tonegawa, S., Brack, C., Hozumi, N., and Schuller, R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3518–3522.
8. Leder, P., Tiemeier, D., and Enquist, L. (1977) *Science* **196**, 175–177.
9. Blattner, F., Williams, B. G., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Furlong, L., Grunwald, D. J., Kiefer, D. O., Moore, D. D., Schumm, J. W., Sheldon, E. L., and Smithies, O. (1977) *Science* **196**, 161–169.
10. Mandel, M., and Hoga, A. (1970) *J. Mol. Biol.* **78**, 453–471.
11. Hohn, B., and Murray, K. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3259–3263.
12. Benton, W. D., and Davis, R. W. (1977) *Science* **196**, 180–182.
13. Tonegawa, S., Maxam, A. M., Tizard, R., Bernard, O., and Gilbert, W. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1488–1489.
14. Tonegawa, S., Hozumi, N., Matthyssens, G., and Schuller, R. (1976) *Cold Spring Harbor Symp. Quant. Biol.* **41**, 877–889.
15. Kramer, R. A., Cameron, J. R., and Davis, R. W. (1976) *Cell* **8**, 227–232.
16. Brack, C., and Tonegawa, S. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5652–5656.
17. Glover, D. M., and Hogness, D. S. (1977) *Cell* **10**, 167–176.
18. Wellauer, P. K., and David, I. B. (1977) *Cell* **10**, 193–212.

19. Breathnach, R., Mandel, J. L., and Chambon, P. (1977) *Nature (London)* **270**, 314–319.
20. Jeffreys, A., and Flavell, R. (1977) *Cell* **12**, 1097–1107.
21. Goodman, H. M., Olson, M. V., and Hall, B. D. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5453–5457.
22. Tilghman, S. M., Tiemeier, D. C., Seidman, J. G., Peterlin, B. M., Sullivan, M., Maizel, J. V., and Leder, P. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 725–729.
23. Berget, S. M., Moore, C., and Sharp, P. A. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3171–3175.
24. Chow, L. T., Gelinas, R. E., Broker, T. R., and Robert, R. J. (1977) *Cell* **12**, 1–8.
25. Klessig, D. F. (1977) *Cell* **12**, 9–21.
26. Aloni, Y., Dhar, R., Laub, O., Horowitz, M., and Khoury, G. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3686–3690.
27. Tilghman, S. M., Curtis, P. J., Tiemeier, D. C., Leder, P., and Weissmann, C. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 1309–1313.
28. Tonegawa, S., Brack, C., Hozumi, N., and Pirrotta, V. (1977) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 921–931.
29. Weigert, M. G., Cesari, I. M., Yonkovich, S. J., and Cohn, M. (1970) *Nature (London)* **228**, 1045–1049.
30. Tonegawa, S. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 203–207.