Characterization of a mouse DNA clone containing an immunoglobulin variable region gene

Nobumichi Hozumi, Christine Brack, Vincenzo Pirrotta* Rita Lenhard-Schuller and Susumu Tonegawa

Basel Institute for Immunology, Grenzacherstrasse 487, Postfach, CH-4005 Basel 5, and *Department of Microbiology, Biozentrum of the University of Basel, Klingelbergstrasse 70, 4056 Basel, Switzerland

ABSTRACT

A 4.8 kilobase mouse embryo DNA fragment was inserted into a phage λ genome and was subsequently characterized by electron microscopy, restriction enzyme mapping and partial DNA sequencing. This fragment contains a 400 base sequence which is homologous to that of an immunoglobulin light λ chain mRNA which spans 3.3 to 3.7 kilobases from one end of the fragment. Restriction enzyme mapping as well as partial nucleotide sequencing of the 3' terminal of the homology region confirm the previous conclusion [Tonegawa, Brack, Hozumi and Schuller, Proc. Natl. Acad. Sci. USA. 74, 3518-3522 (1977)] that the cloned DNA fragment contains a V_λ gene sequence which is separate from any C_λ sequence.

INTRODUCTION

We recently reported on the construction and isolation of a phage λ (λgt_{WES}-Ig 13) that carries a mouse embryo DNA insert containing an immunoglobulin gene (1). Partial characterization of this phage DNA by hybridization and electronmicroscopy indicated that this mouse gene contains the V region of a λ type light chain. The same study suggested that no corresponding C gene is present immediately adjacent to the V gene, thereby directly confirming the concept: two separate DNA segments code for V and C regions of an immunoglobulin chain in embryonic cells (2). Before a firm conclusion is drawn, however, complications arising from the presence of two subtypes of mouse λ chains should be considered. According to the available amino acid sequence data of mouse myeloma λ chains, λ_I chains differ from λ_{II} chains by only ten to thirteen residues in the V region, which is about 110 residues long, whereas the two types of chains differ by 27 residues in the C region of about the same length (3). Based on this comparison of amino acid sequence differences, λ_I and λ_{II} type chains would be expected to cross-hybridize less extensively in the C region than in the V region. Since our previous hybridization

studies on the λgt_{WES}-Ig 13 DNA were carried out with a λ_I type mRNA, we might not have detected a C gene of the λ_{II} type, which may lie immediately adjacent to the V_λ gene (1). Lack of a λ_{II} chain mRNA preparation of sufficient purity prevents us from answering the question by straight-forward hybridization studies. In order to clarify the matter, we have carried out more extensive characterization of the cloned mouse DNA fragment. The results presented here and elsewhere (Tonegawa, Maxam, Tizard, Bernard and Gilbert, Proc. Natl. Acad. Sci. U.S.A., in press) demonstrated that no C_λ-like sequence is contiguous to the V_λ gene sequence.

MATERIALS AND METHODS

   Enzymes Alkaline phosphatase was a generous gift of Dr M. Takanami (Kyoto University, Japan). T_4 polynucleotide kinase was donated by Dr M. Sugiura (National Institute of Genetics, Japan).

   Restriction endonucleases Eco RI (4), Mbo II (5), Bgl II (6) and Alu I (7) were prepared according to published methods. Hinf I was prepared by the unpublished procedure of R. Roberts. Pst I was obtained from Bethesda Research Laboratories. Hae III and Hind III were purchased from New England Biolabs and Miles Laboratories, respectively. Digestions for restriction endonucleases were performed at $37^{\circ}C$ on 0.2 µg of DNA in a volume of 20 µl for a period of 4 h. The reaction mixtures contained the following components: for Eco RI, 0.1 M Tris-HCl (pH 7.5), 0.05 M NaCl, 0.01 M $MgSO_4$, 0.001 M EDTA; for Bgl II, Mbo II, Alu I and Hinf I, 0.006 M Tris-HCl (pH 7.5), 0.006 M $MgCl_2$, 0.006 M 2-mercaptoethanol; for Pst I, 0.02 M Tris-HCl (pH 7.5), 0.01 M $MgCl_2$, 0.05 M $(NH_4)_2SO_4$; for Hae III, 0.006 M Tris-HCl (pH 7.4), 0.006 M NaCl, 0.006 M $MgCl_2$, 0.006 M 2-mercaptoethanol. After incubation, samples were heated at $65^{\circ}C$ for 3 min.

   Phage λgt_{WES}-Ig 13 was grown and purified as previously described (1). The experiments were carried out in a P3 facility in accordance with the NIH guide lines issued in June 1976.

   Preparation of the 4.8 kb cloned DNA The procedures were those previously described (8). Five mg of λgt_{WES}-Ig 13 DNA (100 µg/ml) were digested with Eco RI and concentrated tenfold with butanol-(2)(9). After dialysis against TA buffer (0.02 M Tris-acetate pH 8.0, 0.018 M

NaCl, 0.002 M EDTA, 0.02 M sodium acetate), the digested DNA was loaded in a slot (19 x 1 x 0.5 cm) of a 0.9% agarose gel in TA. Electrophoresis was carried out for 2 days at 1.5 mA/cm^2 of the gel cross section. The band containing the 4.8 kb DNA fragment was visualized on PEI thin layer chromatography sheet (Macherey-Nagel and Co., M N 300) under UV lamp (254 mµ) and cut with a scalpel. The agarose gel slice was melted in NaClO$_4$ and DNA was purified by a hydroxyapatite (Bio-Gel, DNA grade) column.

Formation of R-loop, Electronmicroscopy HOPC 2020 λ mRNA was purified by the method previously described (10). The mRNA was hybridized to the 4.8 kb mouse DNA fragment to form R-loops under the conditions described (1), following essentially the methods of Thomas et al. (11). The samples in 0.56 M NaCl were incubated at 50°C for 12 to 14 hr, alternatively samples in 0.2 M NaCl were incubated at 51°C for the same period of time. Both conditions gave similar results. The spreading solutions containing 70% formamide, 0.1 M Tris-HCl (pH 8.5), 0.01 M EDTA (pH 8.5), 0.1 mg/ml cytochrome C (CNBr treated) and 0.2 - 0.5 µg/ml DNA were spread on a hypophase containing 10-15% formamide, 0.01 M Tris (pH 8.5), and 0.001 M EDTA. Samples were picked up on collodion-coated copper grids, stained in 10^{-5} M uranylacetate in 90% ethanol, dried in isopropanol, and rotary-shadowed with Pt at an angle of 6°.

Hybridization with cDNA cDNA synthesized from MOPC 104E λ mRNA with AMV reverse transcriptase was kindly provided by Dr. G. Matthyssens (to be published elsewhere). Full length cDNA was hybridized to the Ig|3 4.8 kb fragment as follows: Ig 13 DNA was denatured with a solution containing 0.1 M NaOH, 0.002 M EDTA, 0.02 M Tris; after addition of cDNA the solution was neutralized with HCl, containing NaCl. The hybridization mixture finally contained Ig 13 DNA 4 µg/ml, cDNA 3 µg/ml, NaCl 0.12 M, EDTA 0.002 M, Tris 0.02 M, and was incubated at 50°C for 2-3 hr. The samples were then spread for electronmicroscopy as described for R-loops.

Analytical agarose gel electrophoresis Electrophoresis was carried out in a vertical slab gel (24 x 15 x 0.15 cm) of 0.7% or 1.4% agarose. After completion of electrophoresis, the gels were stained with ethidium bromide (1 µg/ml in TA buffer). The gels were photographed onto Polaroid film (Polaroid 4 x 5, Land film type 55) through a red filter (B+W 42,

B+W Filter Fabrik) under ultraviolet illumination (254 mµ).

Analytical acrylamide gel electrophoresis Gel electrophoresis was carried out on a vertical slab gel (35 x 15 x 0.15 cm) containing 0.09 M Tris-borate (pH 8.3), 0.0025 M EDTA. The same buffer was used in electrophoresis reservoirs. 3.5% acrylamide gels were made as described by T. Maniatis et al. (12). At completion of the run, the gels were stained with 1 µg/ml of ethidium bromide in the electrophoresis buffer and photographed as described above.

Autoradiography for restriction site mapping After completion of electrophoresis, the acrylamide gel was transferred onto two sheets of Whatman 3 MM paper, covered with a sheet of Saranwrap, and dried under vacuum. The dried gel was subjected to autoradiography at -70°C using an Ilford tungstate intensifying screen.

In situ hybridization of agarose gel Gels were immersed in 0.2 N NaOH, 0.6 M NaCl for 45 min at room temperature. After washing with water, the gels were neutralized in 1.0 M Tris-HCl (pH 7.5), 0.6 M NaCl for 45 min at room temperature. The DNA within the gel was transferred onto a sheet of nitrocellulose filter (Schleicher and Schüll) essentially by the method described by Southern (13). After blotting, the nitrocellulose filters containing DNA were washed with 2 x SSC and baked in vacuo at 80°C for 2 h. They were then incubated at 50°C for 8 h in a hybridization mixture which contains, in a total volume of 3 ml, 50% formamide, 0.75 M NaCl, 0.075 M sodium citrate, 0.1 M pipes (pH 7.5), 2 x 10^6 cpm ^{125}I-HOPC 2020 λ mRNA (5 x 10^7 cpm/µg). Following hybridization, the filter was washed with 2 x SSC, treated with RNase A (20 µg/ml) and RNase T_1 (2 U/ml) at 37°C for 20 min, and subjected to autoradiography using Ilford tungstate intensifying screens at -70°C.

5'-^{32}P labeling of DNA fragments The 4.8 kb Eco RI fragments (5 µg) or the 1.52 kb Hae III fragment (2 µg) were treated with bacterial alkaline phosphatase. The termini were labeled using γ-^{32}P-ATP and T_4 polynucleotide kinase. The reaction mixture for phosphatase contained in 50 µl: the DNA fragments, 0.1 mM $MgCl_2$, 0.1 M Tris-HCl (pH 8) and 0.4 U of bacterial alkaline phosphatase. Incubation was for 1 h at 37°C. This reaction mixture was extracted with water-saturated phenol three times, followed by ether extraction and ethanol precipitation. The labeling mixture for kinase contained in 100 µl: 0.05 M Tris-HCl (pH 7.5), 0.01 M $MgCl_2$, 0.005 M dithiothreitol, 0.085 mM spermidine, 2 µM

γ-^{32}P-ATP (2260 Ci/mmole, New England Nuclear), the phosphatase-treated
DNA fragments, and 1.5 U of T_4 polynucleotide kinase. The reaction was
carried out at 37°C for 30 min. The terminally labeled 1.52 kb Hae III
fragment was digested with Hind III and electrophoresed on a 3.5% acrylamide
gel for 14 h at 2.5 V/cm. The end labeled 0.62 kb Hind III fragment was
extracted from the gel as described below. The labeled 4.8 kb Eco RI
fragment was digested with Hae III and the digest was fractionated by
electrophoresis through 1.4% agarose gel for 12 h at 2 V/cm. The terminally-
labeled 1.52 kb Hae III fragment was isolated from the agarose gel by
the NaClO$_4$-hydroxyapatite method (see above).

Extraction of DNA from acrylamide gel slices Bands were cut out
from gels and chopped in small slivers with a scalpel. The DNA was
extracted by shaking for several hours with three portions of buffer
containing 0.05 M Tris HCl (pH 7.6), 0.1% SDS, and 0.3 M ammonium acetate
saturated with phenol. The pooled volume was extracted once with phenol,
then with ether, and loaded on a 0.1 ml DEAE cellulose column in a
Pasteur pipette. The column was washed with 3 ml of 0.1 M KCl. The DNA
was eluted with 30% triethylammonium carbonate and finally precipitated
with ethanol.

DNA sequence determination The purified Hinf I 330 base pair
fragment was treated with 3 μg of bacterial alkaline phosphatase for 1 h
at 37°C in 100 μl containing 0.05 M Tris-HCl (pH 7.8), 0.1 M KCl, 0.01 M
MgCl$_2$ and 0.01 M dithiothreitol. The DNA was then extracted 4 times
with phenol, then with ether and then precipitated with 3 volumes of
ethanol. The dried precipitate was resuspended in the same buffer
containing 10 pmoles γ-^{32}P-ATP, 1000 mCi/μmole; 8 units of polynucleotide
kinase were added and the reaction was incubated at 37°C for 40 min.
After ethanol precipitation, the labeled DNA was resuspended in 50 μl of
a buffer containing 0.01 M Tris (pH 7.6), 0.01 M MgCl$_2$ and 0.01 M β-
mercaptoethanol and digested overnight with Mbo II. The labeled fragments
were separated by gel electrophoresis and extracted from the gel. After
purification they were then subjected to the Maxam-Gilbert sequencing
procedure (14). After methylation with dimethyl sulfate, a sample was
depurinated by heating at pH 7 followed by heat and alkali to give the
G>A cleavage, or by acid treatment at 0°C followed by heat and alkali to
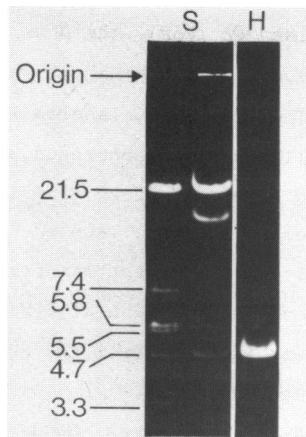give the A>G cleavage. Pyrimidine cleavage was done with 15 M hydrazine

FIGURE 1: Gel electrophoresis pattern of restriction endonuclease Eco RI digestion products of λgt$_{WES}$-Ig 13 DNA. λgt$_{WES}$-Ig 13 (1 μg) was digested with Eco RI and fractionated on a 0.7% agarose gel for 14 h at a potential of 2 V/cm. The products were stained with ethidium bromide. The stained DNA was transferred onto a sheet of nitrocellulose filter and hybridized to ^{125}I-labeled HOPC 2020 λ chain mRNA. λ phage DNA digested with Eco RI was used as a migration marker. The molecular sizes of the marker are given in kilobase. S: stained DNA. H: hybrids. The autoradiogram shown here was developed for 20 h.

(C+T) or 15 M hydrazine + 2.5 M NaCl (C>T), followed by heating with 0.5 M piperidine. The products were run on a 20% acrylamide gel 40 cm long containing 7 M urea, at 800 - 1000 volts, with two loadings.

RESULTS

Location and length of homology   When λgt$_{WES}$-Ig 13 DNA was digested with Eco RI, three fragments were generated (Figure 1). Two of these fragments corresponded to the left and the right arms of the parental phage λgt$_{WES}$-λB (21 kb and 14 kb). The third fragment of 4.8 kb contained a sequence homologous to ^{125}I-labeled, purified λ chain mRNA isolated from HOPC 2020 λ myeloma.

In order to determine the position and length of the homology region, R-loops were formed between HOPC 2020 λ chain mRNA and the 4.8 kb Eco RI fragment which had been purified by agarose gel electrophoresis. We obtained about 50% hybrid molecules. About half of these were normal R-loops with an RNA tail at one end of the loop structure (Figure 2B). Other hybrids showed, instead of a loop, a short stretch of thick, probably "triple-stranded" region, with one or two single stranded tails at the end(s) (Figure 2A). In the histogram shown in Figure 3, the
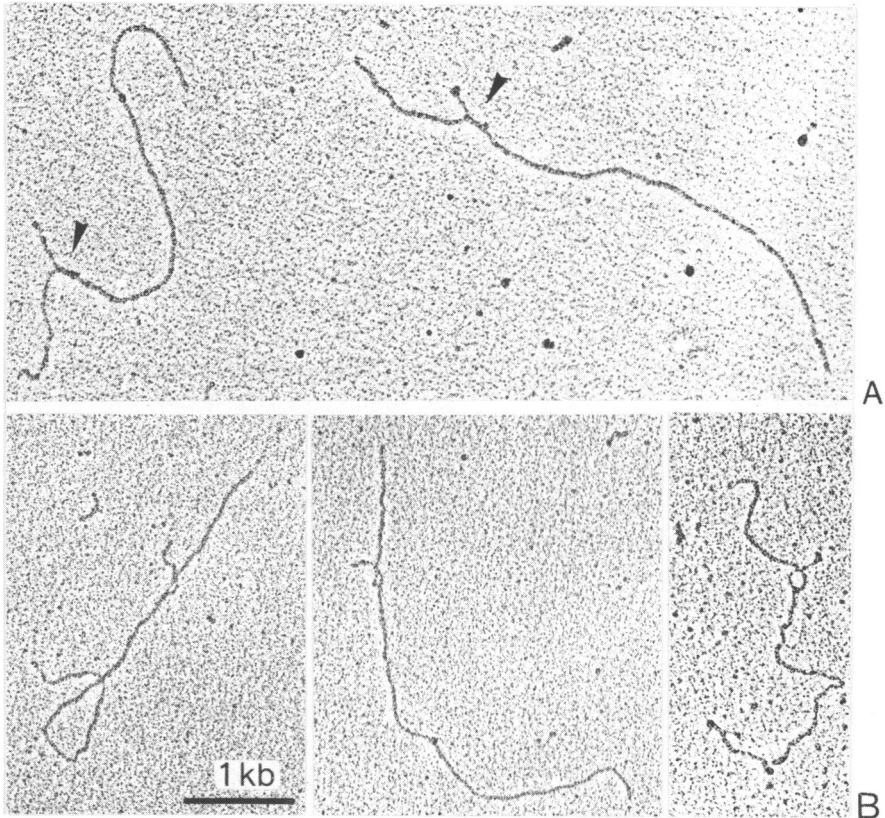
Figure 2: Electron micrographs of R-loop molecules formed between the 4.8 kb Ig 13 DNA and HOPC 2020 mRNA (see Materials and Methods). The thick "triple-strand" regions in A (→) are probably collapsed R-loops.

solid line shows the position and extension of the R-loops (32 molecules). The position at which the RNA tail extends from the triple strand (42 molecules) corresponded exactly to the tail-carrying end of the R-loop (dotted line, histogram, Figure 3). We therefore conclude that the "triple-stranded" structures are collapsed R-loops. Since the 3'-end half of the mRNA does not hybridize to the mouse DNA insert, (see ref. 1), the tail should correspond to the 3'-end of the mRNA or to C gene sequences.

In R-loop preparations the RNA molecules or RNA tails are usually not fully extended and it is therefore difficult to measure the length of the non-hybridized part of the mRNA extending from the R-loop. In order to obtain more accurate measurements, full length cDNA (synthesized on M104 λ mRNA with the AMV reverse transcriptase) was hybridized to
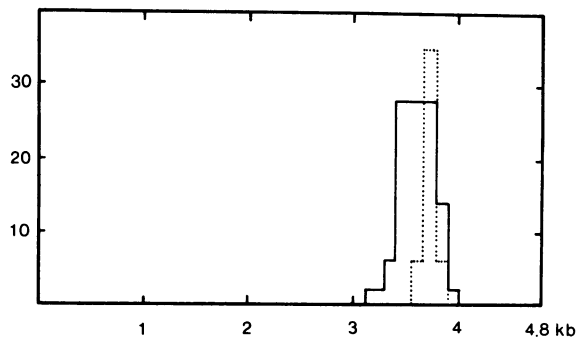
Figure 3: Histogram showing position and extension of R-loops measured on 32 hybrid molecules (solid line). The dotted line represents the position of the RNA tail extruding from "triple strand" structures, as well as from R-loops; it allows for the orientation of the molecule.

denatured Ig 13 DNA. The different parts of the hybrid molecules were measured (see Fig. 4) on 35 molecules: a = 3.29 $\pm$ 0.15 kb, b = 1.06 $\pm$ 0.22 kb, homology region (double-stranded) (h) = 0.398 $\pm$ 0.103 kb, tail (t) = 0.373 $\pm$ 0.10 kb.

From these results and the R-loop mapping we can conclude that the homology region is about 400 nucleotides long, and is localized between 3290 and 3690 ($\pm$20) nucleotide pairs from one end of the mouse DNA which lies to the left in the conventional map of a λ phage genome (1).
Figure 5 shows the map of the 4.8 kb mouse DNA fragment as obtained from electronmicroscopy.

Mapping of restriction enzyme cleavage sites

In order to characterize the mouse DNA insert further, the cleavage sites of several restriction enzymes were determined. Two kinds of experiments were carried out. In one series of experiments, the 4.8 kb Eco RI fragment was digested by several restriction enzymes which cleave DNA relatively infrequently. The digests were electrophoresed in 1.4% agarose gel, and the DNA fragments containing sequences homologous to HOPC 2020 λ chain mRNA were identified by the blotting technique developed by Southern (13). Since the electronmicroscopy study described in the last section indicated that the homology region lies between 3.29 and 3.69 kb from one end of the 4.8 kb fragment, this was used as a guide in ordering the fragments whenever possible.

Pst I cleaved the 4.8 kb DNA into two fragments of 3.33 and 1.47 kb (Figure 6A). While both fragments hybridized to the H2020 λ chain mRNA, the level of hybridization of the larger fragment was only marginal.
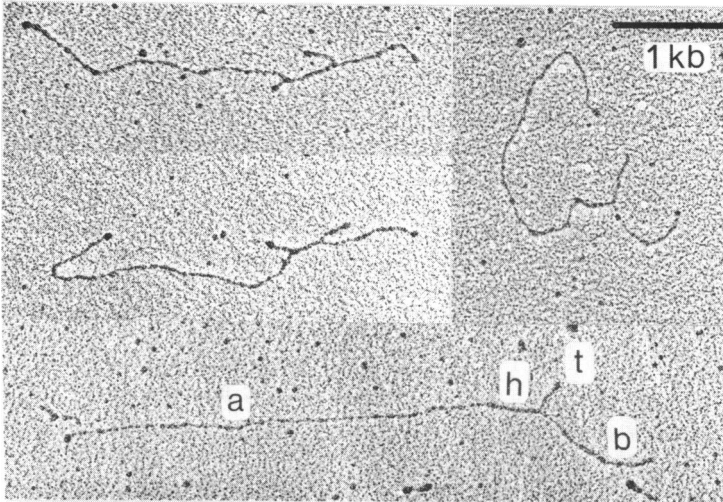
Figure 4: Hybrid molecules made by annealing cDNA to denatured Ig 13 DNA. The structure consists of two single-stranded arms (a,b), the double-stranded homology region (h) and a single-stranded tail (t). The different parts of the molecules were measured, using as internal standards DNAs of phages fd and PM2.

Since the homology region lies between 3.29 and 3.69 kb from the left end of the Eco RI fragment the smaller Pst I fragment (1.47 kb) which contained most of the homology region should contain the right Eco RI site.

Hae III generated three fragments of 2.53, 1.52 and 0.75 kb (Figure 6B). Hybridization was dominant in the 1.52 kb fragment, but longer exposure revealed weak hybridization in the 2.53 kb fragment. The order of the three Hae III fragments which is compatible with the homology region map shown in Figure 5, is 0.75, 2.53 and 1.52 kb fragments from left to right.

Hind III cleaved the Eco RI DNA piece into two fragments of 3.90 and 0.90 kb (Figure 6C). Only the 3.90 kb fragment showed hybridization with the H2020 λ chain mRNA. In order to determine the order of the two fragments, the Eco RI DNA piece was digested with Hind III and Pst I simultaneously (Figure 6C). The double digestion generated three fragments of 3.33, 0.90 and 0.57 kb. This result indicates that the Hind III site is within the Pst I, 1.47 kb fragment.

Mbo II generated many fragments, among which only the largest, 0.96 kb fragment hybridized with the mRNA (Figure 6A). Attempts were not
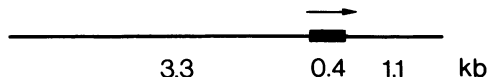
Figure 5: A map of Ig-13 DNA showing the position and orientation of
the region homologous to HOPC 2020 mRNA. The lengths are given in
kilobases (kb).

made to order these fragments (see below for Mbo II cleavage sites in
the Hae III, 1.52 kb fragment).

Digestion by Bgl II generated two bands at positions corresponding
to 1.64 and 1.51 kb (Figure 6B). The sum of the length of the two
fragments, 3.15 kb, is 1.65 kb shorter than it should be. Intensity of
the two DNA bands suggests that the 1.64 kb band contains two fragments
which did not resolve under the electrophoresis conditions used.
Hybridization was detected only in this band (Figure 6B). The order of
the other two Bgl II fragments was not determined. The results described
above are summarized in Figure 8A. In addition, Bgl I, Bam HI and Hha I
were found to have no cleavage site in the 4.8 kb Eco RI piece.

In the second series of experiments, cleavage sites of several
enzymes within the 1.52 kb Hae III fragment were determined by the
partial digestion method (15).   Two DNA fragments, each having a ^{32}P
label at an opposite end of the Hae III fragment were prepared. First,
the 4.8 kb Eco RI fragment was terminally labeled with γ-^{32}P-ATP by T$_4$
polynucleotide kinase. After digestion with Hae III, the 1.52 kb fragment
with ^{32}P label only on the right end (see Figure 8A) was isolated.
Second, the unlabeled 1.52 kb Hae III fragment was isolated and labeled.
Then, the 0.62 kb Hind III fragment labeled at the left end was obtained.

Partial digestion of the 1.52 kb Hae III fragment with Hinf I
generated six fragments of 350, 770, 870, 1200, 1360 and 1520 base pairs
(Figure 7A). Partial digestion of the 0.62 kb Hind III fragment with
Hinf I generated four fragments of 150, 160, 320 and 620 base pairs
(Figure 7B). These results led to identification of Hinf I sites which
are located within the 1.52 kb Hae III fragment at the intervals of 150,
10, 160, 330, 100, 420 and 350 base pairs from left to right. The two
sites flanking the 10 base pair fragments were not identified as such in
the partial digest of the 1.52 kb Hae III fragment, because separation
of the two fragments with and without the 10 base pairs were beyond
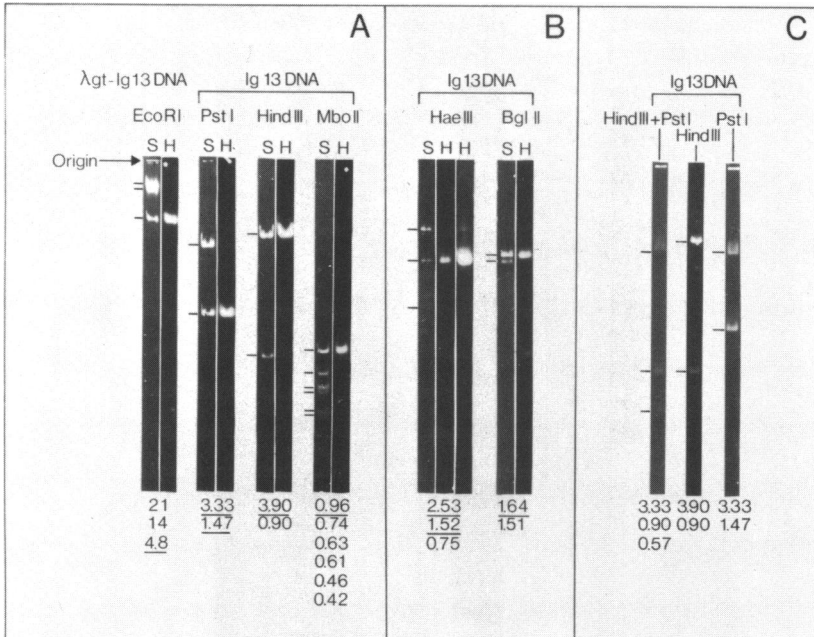resolution of the electrophoresis. Complete digestion of the 1.52 kb

Figure 6: Gel electrophoresis patterns of λgt_WES-Ig 13 DNA or its mouse DNA insert (Ig 13 DNA) digested by various restriction endonucleases. (A) and (B): digestion products were electrophoresed on a 1.4% agarose gel for 12 h at a potential of 1.7 V/cm. The gels were stained with ethidium bromide. The stained DNAs were transferred to nitrocellulose filters and hybridized to ^{125}I-HOPC 2020 mRNA. (C): DNA fragments were electrophoresed and visualized as described in (A) and (B). In (A), (B) and (C), Eco RI digestion products of λ phage DNA and SV40 DNAs digested with Hind III or Hae III were used as migration markers (data not shown). The lines beside the lanes indicate DNA bands. The sizes of various fragments (in kilobase pairs) are shown at the bottom of the gel slots. The underlined numbers correspond to DNA fragments containing sequences hybridizable to ^{125}I-HOPC 2020 λ mRNA. S: stained DNA. H: hybrids. The autoradiograms shown were allowed to develop for 24 h except for the right slot of Hae III (H) which was developed for 3 days (see text).

Hae III fragment with Hinf I generated fragments of 425, 350, 325, 165, 155 and 100 base pairs (data not shown). The result is in good agreement with the partial digestion experiments. The 10 base pair fragment was too small to be observed by ethidium bromide staining. Similarly, partial digestion of the 1.52 kb Hae III fragment with Alu I led to identification of eight cleavage sites, from left to right: 300, 250, 70, 150, 180, 515, 5 and 50 base pairs (Figure 7C). Complete digestion of the 1.52 kb Hae III fragments with Alu I generated fragments of 540, 305, 245, 175, 160 and 70 base pairs (data not shown). The two small
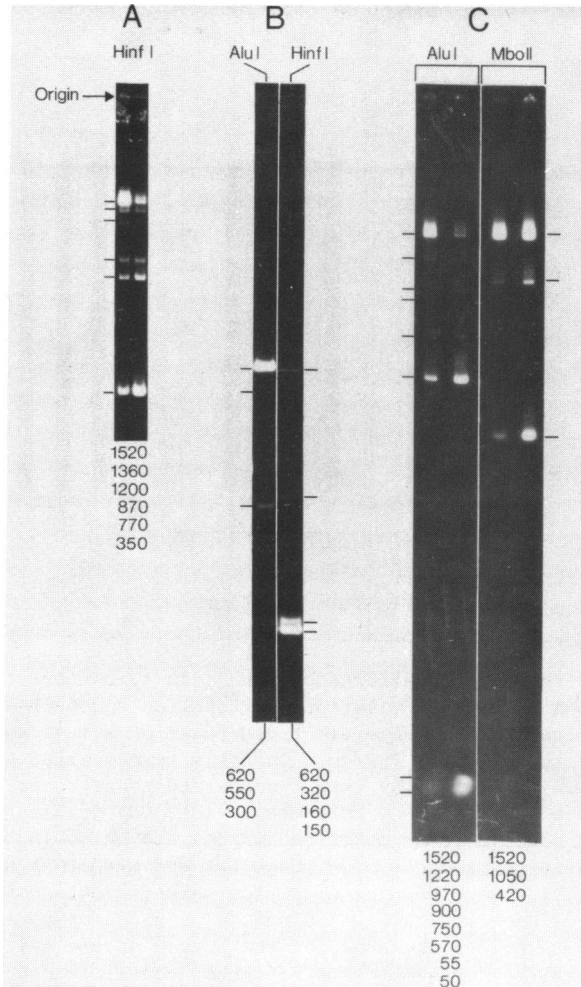
Figure 7: Autoradiogram of 1.52 kb Hae III B fragments and 0.62 kb Hind III fragments partially digested with various restriction endonucleases. (A): the end-labeled 1.52 kb Hae III DNA fragment (760 cpm) was partially digested with Hinf I and fractionated by electrophoresis through a 3.5% acrylamide gel for 11 h at a potential of 3 V/cm. (B): the terminally - labeled 0.62 kb Hind III fragment (4700 cpm) was partially digested with the restriction enzymes indicated and electrophoresed as described in Figure 6A. The autoradiogram shown was allowed to develop for 12 h. (C): the end-labeled 1.52 kb Hae III fragment was partially digested with the restriction enzymes indicated and analyzed as described in the legend to Figure 6A. SV40 DNA digested with Hind III was terminally-labeled with polynucleotide kinase and used as a migration marker (data not shown). The autoradiogram shown was developed for 3 days. In (A), (B) and (C), the sizes of fragments (in base pairs) are shown at the bottom. The lines beside the lanes indicate DNA bands.
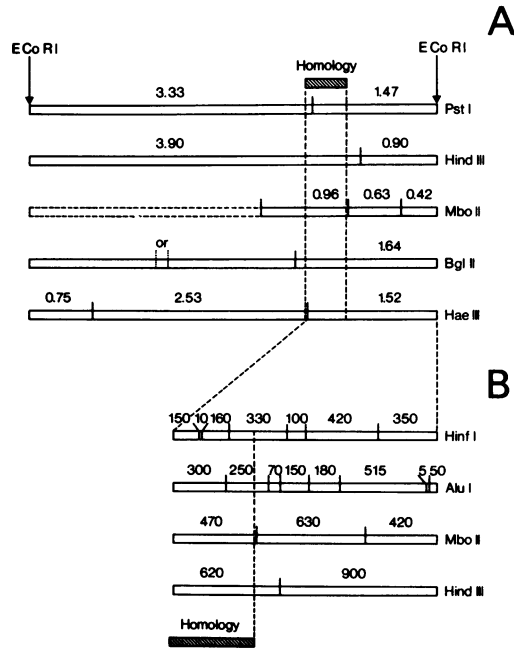
Figure 8: The restriction map of the mouse DNA insert. (A): the 4.8 kb
Eco RI fragment. (B) : the 1.52 kb Hae III fragment. Fragment sizes
are given in kilobase pairs for (A) and in base pairs for (B). The
shaded boxes indicate the region homologous to HOPC 2020 λ mRNA. The
exact position of the Bgl II cleavage site at the left side was not
determined. Mbo II map is incomplete (column with dotted lines). See
text for further explanation.

fragments of 50 and 5 base pairs which were deduced to be present from

partial digestion, were not detected in the completely digested DNA

mixture. Hind III and Alu I recognize and cleave the sequence AAGCTT

and AGCT, respectively. Partial digestion of the 620 base pair Hind III

fragment (labeled at the left end) should give 300, 550 and 620 base

pair fragments. As shown in Figure 6B, this prediction was confirmed.

Partial digestion of the terminally (left end) labeled 1.52 kb Hae

III fragment with Mbo II lead to identification and ordering of the DNA

fragments of 470, 630 and 420 base pairs, from left to right (Figure

7C). Complete digestion confirmed this deduction, yielding 630, 470 and

415 base pair fragments (data not shown). These results are summarized

in Figure 8B.

Nucleotide sequences DNA sequencing gave direct evidence that the

mouse DNA insert contains λ chain gene sequences. As Figure 8B shows,

Figure 9: Sequencing gels of the Hinf 330 base pair fragment. The fragment was end-labeled with polynucleotide kinase, cleaved with Mbo II and the resulting fragments, A (left) and B (right) were separated by gel electrophoresis. The two fragments were treated according to Maxam and Gilbert (14) in four separate reactions cleaving the sequence at G>A, A>G, C>T and C+T. The gels contained 20% acrylamide and 7 M urea and were run at 800-1000 volts. The positions of the marker dyes, xylene cyanol and bromphenol blue, are indicated. Panel A shows a sequencing gel of the left fragment, Mbo II, containing sequences homologous to λ chain mRNA. Panel B shows a sequencing gel of the right end of the Hinf fragment. Note that in this gel the bands in the pyrimidine cleavage reactions are slightly delayed relative to the purine cleavage bands, probably due to incomplete elimination of the piperidine from the samples applied to the gel.

```
                 Hinf I
                      ↓
OBSERVED     5' GA  NTC  ___  GGC  TCC  CTG  ATT  GGA  GAC  AAG  GCT  GCC  CTC  ACC  ATC  ACA


                      65                     70                     75
PREDICTED    5' CGX  UUY  UCX  GGX  UCX  CUX  AUY  GGX  AAY  AAR  GCX  GCX  CUX  ACX  AUY  ACX
               AGR        AGY       AGY  UUR  AUA                      UUR       AUA



OBSERVED     5' GGG  GCA  CAG  ACT  GAG  GAT  GAT  GCA  ATG  TAT  TTC  TGT  GCT  CTA  TGG  TAC  AGC


                      80                     85                     90                     95
                                             GAY  AUG                                     UUY  CGX
PREDICTED    5' GGX  GCX  CAR  ACX  GAR  GAY  GAR  GCX  AUY  UAY  UUY  UGY  GCX  CUX  UGG  UAY  AGY
                                                       AUA                      UUR            UCX
```

Figure 10: Partial nucleotide sequence of the left half of the Hinf 330 base pair fragment. The observed nucleotide sequence (Figure 9A) is shown in comparison with that predicted from the amino acid sequence of a $V_{\lambda II}$ region (MOPC 315 λ chain), which is identical with that of $V_{\lambda I}$ (MOPC 104E λ chain) except for positions 85, 87, 94 and 95. Nucleotide sequences predicted from MOPC 315 λ chain at these positions are shown in italics. Positions 94 and 95 are within a hypervariable region. X,N: any one of the four bases. R: purine. Y: pyrimidine.

the Hinf I 330 base pair fragment spans the boundary of the homology

region and therefore should contain both V and C homologous sequences

and the region immediately adjacent to them. This fragment was labeled

at the 5'-ends with polynucleotide kinase, cut with Mbo II and the

nucleotide sequence of the two halves was determined by the method of

Maxam and Gilbert (14). The corresponding autoradiographs are shown in

Figure 9A, B. The resulting sequences are shown in Figures 10 and 11

for the left and right half, respectively. To identify the region of

homology in the amino acid sequence of the λ chain, we aligned with the

observed DNA sequence of the left half all the possible Hinf I sites in

the nucleotide sequence predicted from the amino acid sequence. Figure

10 shows that the possible Hinf I site at amino acid positions 63 and 64

of the V_λ region gives an excellent alignment with the sequence predicted.

The exceptions to the fit are interesting. At positions 85 and 87, the

DNA sequence corresponds to the predicted sequences of MOPC 315 $V_{\lambda II}$

instead of MOPC 104E $V_{\lambda I}$. At two more positions, 94 and 95, the sequence

corresponds to $V_{\lambda I}$. Does this mean that embryonic DNA contains a single

V_λ gene which is a hybrid of the $V_{\lambda I}$ and $V_{\lambda II}$ types? We believe that a

more likely explanation is that there are separate germ line genes for

$V_{\lambda I}$ and $V_{\lambda II}$, and that the gene contained in the present clone is the

$V_{\lambda II}$ gene. The basis for this explanation is that positions 94 and 95

```
                          Alu I  Hind III                    Hinf I
                           ↓ ↓                                 ↓
3'...AAGTTGTTTAACATATTATTAATACGGAACTACTGTTCGAAACAAATAGTTGAACCGTCTTTTG___CTNAG   5'
5'...TTCAACAAAITGIATAATAATTATGCCTTGATGACAAGCTTTGTTTATCAACTTGGCAGAAAAC___GANTC   3'

      PheAsnLysLeuTyrAsnAsnTyrAlaLeuMetThrSerPheValTyrGlnLeuGlyArgLys

        SerThrAsnCysIleIleIleMetPro→*←→*←GlnAlaTrpPheIleAsnLeuAlaGluAsn

        GlnGlnIleVal→*←→*←LeuCysLeuAspAspLysLeuCysLeuSerThrTrpGlnLys
```
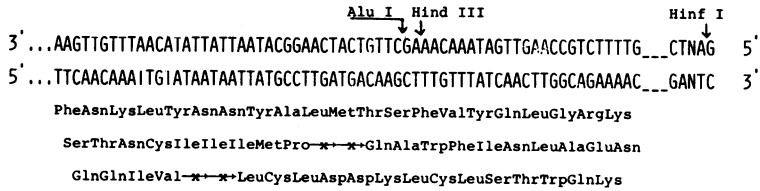
Figure 11: Partial nucleotide sequence of the right half of the Hinf 330 base pair fragment. The three amino acid sequences are shown which are in accordance with the experimentally determined nucleotide sequence. Alu I and Hind III sites are indicated by arrows. →*←→ indicates a termination triplet.

are in the hypervariable region, while positions 85 and 87 are framework residues (16). We suppose that the two germ line genes code for common amino acids at positions 94 and 95 and that the amino acids observed in MOPC 315 λ_{II} at these positions are the result of somatic changes (see Discussion).

Beyond amino acid 98 the observed DNA sequence diverges from the expected sequence, indicating the end of the homology region. Is there then no C gene immediately adjacent to the V gene? Since λ_I type mRNA was used·in the hybridization experiments, a $C_{\lambda II}$ sequence might not have been detected because of the relatively large difference in the amino acid sequences of the two C regions (29 out of 102 amino acids) (3). However, the sequence of the right half of the Hinf fragment shows that the region adjacent to the V gene bears no noticeable similarity to the sequences predicted from either $C_{\lambda I}$ or $C_{\lambda II}$. Figure 11 shows the observed DNA sequence, its complementary sequence, and the corresponding amino acid sequence obtained by reading the complementary sequence in each of the three frames. Reading in two of the frames runs into double termination codons, while the third frame yields an amino acid sequence which bears no similarity to either the $C_{\lambda I}$ or the $C_{\lambda II}$ region. Whether or not some residual sequence homology exists between this region and C_λ genes remains to be determined. We conclude that the V_λ gene in this clone does not have a contiguous C_λ gene. More extensive DNA sequencing in this and adjacent regions confirms these results (17).

DISCUSSION

We reported here detailed characterization of λgt_{WES}-Ig 13 DNA by electronmicroscopy, restriction enzyme mapping and partial DNA sequencing.

Electronmicroscopic studies of hybridization with cDNA and R-loops
formed between the isolated Eco RI fragment of the Ig 13 DNA clone and a
purified λ chain mRNA from HOPC 2020 myeloma not only confirmed that the
cloned DNA contains a 400 base long sequence which is highly homologous
to one end of the mRNA molecule, but also demonstrated that the entire
homology region lies internally, spanning 3.3 to 3.7 kb from one end of
the 4.8 kb mouse DNA fragment. As we reported earlier, this end lies to
the left in the conventional map of a λ phage genome (1). Since the
Ig 13 clone was prepared from the agarose gel DNA fraction which contains
a V_λ gene sequence, it is very likely that the homology region represents
a V_λ gene sequence. This conjecture was confirmed by our earlier experi-
ments which showed that while the cloned DNA hybridizes with the whole λ
chain mRNA, it did not hybridize with the 3'-end half of the same mRNA
which contains the C_λ gene sequence but no V_λ gene sequence (1).

Presence of internally localized V_λ gene sequence in the cloned
mouse DNA strongly suggests that in the embryonic DNA, V and C gene
sequences are not contiguous. In the case of κ chain mRNA, the two gene
sequences are known to be contiguous and there is no reason to believe
that a λ chain mRNA is different from a κ chain mRNA in this respect
(18). Thus, if the V and C gene sequences are indeed separate in embryo
DNA, a novel mechanism must be provoked by which coding information in
two separate DNA sequences is integrated during genesis of mature mRNA
molecules (8,19,20).

In order to investigate the complications arising from the presence
of λ subtypes (see Introduction), we constructed an extensive map of
restriction enzyme cleavage sites in the homology region and regions
flanking it. Although the ultimate purpose of constructing such a map
was to determine the nucleotide sequences of the relevant regions,
information obtained from the map itself was useful in answering the
above questions, at least partially. Among the six cleavage sites
discovered within the homology region, four located to the right-hand
side were in good agreement with the amino acid sequence of either $V_{\lambda I}$
or $V_{\lambda II}$ regions. The positions of the other two sites relative to these
four sites suggested that the former are in the 5'-end untranslated
region, for which we have no way of testing the fit. Eight cleavage
sites were discovered in the DNA segment which occupies a region immediate-
ly adjacent to, and to the right of the homology region. This segment

should correspond to a C_λ gene sequence, should it exist in contiguiuty with the V_λ gene sequence. Some of the six cleavage sites are clearly inconsistent with the assumption. For instance, the Mbo II and Hind III sites separated by 150 base pairs should not exist in either a $C_{\lambda I}$ or $C_{\lambda II}$ gene sequence.

The issues raised were more directly analysed by determining the nucleotide sequence of the 330 base long Hinf I fragment. This DNA fragment covers the rightward boundary of the homology region, and hence the nucleotide sequence at the left end of the fragment should correspond to a V_λ region, whereas the sequence at the right end should correspond to a C_λ region, should a C_λ gene lie in contiguity with a V_λ gene. The 92 base long sequence at the left end fits excellently with the predicted sequence of the MOPC 104E $V_{\lambda I}$ or MOPC 315 $V_{\lambda II}$ region. Exceptions are at positions 85 and 87, where the determined nucleotide sequences fit only with the $V_{\lambda II}$ region, and at positions 94 and 95 where there is fitting only with the $V_{\lambda I}$ region. From these results we would like to tentatively conclude that the cloned sequence is of a germ line $V_{\lambda II}$ gene. The reason is as follows. From RNA-DNA hybridization kinetics, we previously concluded that both $V_{\lambda I}$ and $V_{\lambda II}$ genes are virtually unique, namely, there exists only one copy of each, per haploid genome (19,21). A large number of different $V_{\lambda I}$ or $V_{\lambda II}$ regions are generated from the $V_{\lambda I}$ or $V_{\lambda II}$ gene carried in germ cells by some somatic mechanism. Whatever these mechanisms are, variation thus generated is highly enriched in the subregions called "hypervariable" regions. Most, if not all, of the variation outside the "hypervariable" regions (i.e. in "framework" residues) seems to be the result of evolutionary processes. It is these framework residues which characterize a germ line V gene (16). The determined nucleotide sequence of the Ig 13 clone corresponds exactly to the $V_{\lambda II}$ region in framework residues, while it differs from the predicted sequence of the $V_{\lambda I}$ region in at least two framework residues. It is for these reasons that we think the Ig 13 V gene is a $V_{\lambda II}$ gene. Lack of fit at positions 94 and 95, both in the hypervariable region, could be explained as follows. Since the Ig 13 clone was prepared from the DNA of early embryonic cells, the $V_{\lambda II}$ gene contained is probably the germ line gene sequence. On the other hand, the MOPC 315 $V_{\lambda II}$ region might be considered to have been somatically altered; hence, the presence of two unfitting triplets in the hypervariable region. Fit of the

determined nucleotide sequence, the putative germ line $V_{\lambda II}$ gene, to the MOPC 104E $V_{\lambda I}$ region in these triplets is not surprising. This is because a close evolutionary relationship between the two germ line V genes, $V_{\lambda I}$ and $V_{\lambda II}$, is evident from the presence of extensive amino acid sequence homology in the two types of V regions (22).

The amino acid sequences predicted from the 61 nucleotides determined from the right end of the Hinf 330 base pairs fragment has no resemblance to either the $C_{\lambda I}$ or $C_{\lambda II}$ regions. Two out of the three amino acid sequences each generated by reading the nucleotide sequence in three different phases contain gaps due to two successive termination codons. The third sequence has no similarity to either the $C_{\lambda I}$ or $C_{\lambda II}$ regions. Combining these results with the fact that the 61 nucleotide long DNA segment is located well within the region that would be occupied by a C gene, if it existed in contiguity with a V gene, we conclude that $V_{\lambda II}$ and $C_{\lambda II}$ genes are separate in embryonic DNA.

Identity of the homology region to a $V_{\lambda II}$ gene was confirmed by more extensive nucleotide sequencing. We have recently completed sequencing of a segment about 1,000 nucleotides long, which contains the entire homology region and the adjacent regions (17). In all but one of the framework residues (11 out of 12) where $V_{\lambda I}$ (MOPC 104E) and $V_{\lambda II}$ (MOPC 315) differ, the determined nucleotide sequence fits to the $V_{\lambda II}$ region. The significance of the one exception is not clear at this time. It might reflect "leakiness" of the mechanism by which somatic change is restricted to the "hypervariable" region. Other, more trivial explanations such as the occurrence of a fortuitous mutation, or error in protein sequencing, are not impossible. Absence of a $C_{\lambda II}$ gene adjacent to the homology region was also confirmed by more extensive sequencing (17). It should, however, be mentioned that neither the present work nor the work cited here eliminates the possiblity that the Ig 13 DNA contains a $C_{\lambda II}$ gene sequence in the region where no DNA sequence is available. This, however, does not alter the conclusion that $V_{\lambda II}$ and $C_{\lambda II}$ genes are separate in embryonic DNA. The possibility that the two gene sequences do lie contiguously in embryonic DNA, but were separated fortuitously during gene cloning procedures, is not formally eliminated. We think that this possibility is highly unlikely in the light of our studies on λ_I type genes. We recently isolated from embryonic cells two DNA fragments that carry a $V_{\lambda I}$ gene and a $C_{\lambda I}$ gene

separately (Tonegawa, Brack and Lenhard-Schuller, unpublished). We also isolated from HOPC 2020 myeloma a third DNA fragment that carries both $V_{\lambda I}$ and $C_{\lambda I}$ genes (20). Heteroduplex and R-loop studies of these DNA clones demonstrated directly that the two genes are distant in embryo DNA and are brought closer together in myeloma DNA. Analysis of total DNA from embryos and the myeloma by the gel blotting technique gave results entirely consistent with the conclusion drawn from the studies of isolated DNA clones (Hirama and Tonegawa, unpublished).

REFERENCES
1.   Tonegawa, S., Brack, C., Hozumi, N. and Schuller, R. (1977) Proc. Natl. Acad. Sci. USA 74, 3518-3522.
2.   Dreyer, W.J. and Bennett, J.C. (1965) Proc. Natl. Acad. Sci. USA. 54, 864-869.
3.   Dugan, E.S., Bradshaw, R.A., Simms, E.S. and Eisen, H.N. (1973) Biochemistry 12, 5400-5416.
4.   Greene, P.J., Betlach, M.C. and Boyer, H.W. (1974) In Methods in Molecular Biology (ed. R.B. Wickner) Vol. 7, p. 87, Marcel Dekker, New York.
5.   Gelinas, R.E., Myers, P.A. and Roberts, R.J. (1977) J. Mol. Biol. 114, 169-179.
6.   Bickle, T.A., Pirrotta, V. and Imber, R. (1977) Nucleic Acids Research 4, 2561-2572.
7.   Roberts, R.J., Myers, P.A., Morrison, A. and Murray, K. (1976) J. Mol. Biol. 102, 157-165.
8.   Hozumi, N. and Tonegawa, S. (1976) Proc. Natl. Acad. Sci. USA, 73, 3628-3632.
9.   Stafford, D.W. and Bieber, D. (1975) Biochim. Biophys. Acta 378, 18-21.
10. Tonegawa, S., Brack, C., Hozumi, N., Matthyssens, G. and Schuller, R. (1977) Immunological Rev. 36, 73-94.
11. Thomas, M., White, R.L. and Davis, R.W. (1976) Proc. Natl. Acad. Sci. USA, 73, 2294-2298.
12. Maniatis, T., Jeffrey, A. and van de Sande, H. (1975) Biochemistry 14, 3787-3793.
13. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
14. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA, 74, 560-564.
15. Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Research 3, 2387-2398.
16. Cohn, M., Blomberg, B., Geckler, W., Raschke, W., Riblet, R. and Weigert, M. (1974) In The Immune System, genes, receptors, signals. (ed. E.E. Sercarz et al.) p. 89 Academic Press, New York.

17. Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O. and Gilbert, W. (1978) Proc. Natl. Acad. Sci. USA, in press.
18. Milstein, C., Brownlee, G.G., Cartwright, E.M., Jarvis, J.M. and Proudfoot, N.J. (1974) Nature 252, 354-359.
19. Tonegawa, S., Hozumi, N., Matthyssens, G. and Schuller, R. (1976) Cold Spring Harbor Symp. Quant. Biol. 41, 877-899.
20. Brack, C. and Tonegawa, S. (1978) Proc. Natl. Acad. Sci. USA, 74, 5652-5656.
21. Tonegawa, S. (1976) Proc. Natl. Acad. Sci. USA 73, 203-207.
22. Kabat, E.A., Wu, T.T. and Bilofsky, H. In Variable Regions of Immunoglobulin Chains, National Institutes of Health, Bethesda, MD 20014.