A Complete Immunoglobulin Gene Is Created by Somatic Recombination

Christine Brack, Minoru Hirama, Rita Lenhard-Schuller and Susumu Tonegawa
Basel Institute for Immunology
Grenzacherstrasse 487
Postfach, 4005 Basel 5
Switzerland
High molecular weight DNAs extracted from 13 day old BALB/c embryos (B), myelomas H 2020 (αλ chain producer) (A) and MOPC 321 (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 nick-translated RsaI fragment of the plasmid B1 DNA (for details, see Experimental Procedures).

FIG. 1. Agarose gel electrophoresis profiles of an EcoRI digest of total HOPC 2020 myeloma DNA. DNA (4 mg) was fractionated in 0.9% agarose (1 cm thick, 20 cm wide). One-fifth aliquot each of DNA eluted from 5-mm gel slices was hybridized with 125I-labeled (5 x 10^7 cpm/μg) whole λ chain mRNA from HOPC 2020 myeloma (●) or with its 3' end half (O---O) to C0.5 = 10,000 mol sec/liter (see ref. 4 for definition of C0.5). Inputs were 1300 cpm and 620 cpm for the whole and half RNA probes, respectively. Other procedures have been described (3). Numbers at the top indicate size (in kb) of duplex DNA markers. Fractions indicated by the brackets were pooled and used in ligation reaction.
An Immunoglobulin Heavy Chain Variable Region Gene Is Generated from Three Segments of DNA: $V_H$, $D$ and $J_H$

P. Early, H. Huang, M. Davis, K. Calame and L. Hood
Division of Biology
California Institute of Technology
Pasadena, California 91125

Introduction

The immunoglobulin molecule is a complex entity with two major functions—recognition of foreign substances (antigen binding) and the elimination or destruction of these foreign substances (effector func-
<table>
<thead>
<tr>
<th>Clone</th>
<th>DNA Source</th>
<th>DNA Pre-enrichment Steps</th>
<th>Approximate Pre-enrichment Factors</th>
<th>Screening Procedures and Probes</th>
<th>Approximate Number of Plaques Screened</th>
<th>λ Gene (^a) Sequences Contained</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ig 99(\lambda)</td>
<td>Embryo 3.5 kb</td>
<td>Agarose gel R looping (one cycle)</td>
<td>300</td>
<td>Benton and Davis (1977) with nick-translated cDNA plasmid</td>
<td>3,000</td>
<td>(V_{\alpha})</td>
<td>This paper</td>
</tr>
<tr>
<td>Ig 25(\lambda)</td>
<td>Embryo 8.6 kb</td>
<td>Agarose gel</td>
<td>15</td>
<td>Benton and Davis (1977) with nick-translated cDNA plasmid</td>
<td>80,000</td>
<td>(C_{\alpha})</td>
<td>This paper</td>
</tr>
<tr>
<td>Ig 303(\lambda)</td>
<td>H2020 7.4 kb</td>
<td>Agarose gel</td>
<td>15</td>
<td>Kramer, Cameron and Davis (1976) with (^{125}\text{I}-\lambda_i) mRNA</td>
<td>70,000</td>
<td>(V_{\alpha} + C_{\alpha})</td>
<td>Brack and Tonegawa (1977)</td>
</tr>
<tr>
<td>Ig 13(\lambda)</td>
<td>Embryo 4.8 kb</td>
<td>Agarose gel R looping (two cycles)</td>
<td>360</td>
<td>Kramer, Cameron and Davis (1976) with (^{125}\text{I}-\lambda_i) mRNA</td>
<td>4,000</td>
<td>(V_{\alpha})</td>
<td>Tonegawa et al. (1977); Hozumi et al. (1978)</td>
</tr>
</tbody>
</table>

\(^a\) See Tonegawa et al. (1976) for the definition of the enrichment factor.
\(^b\) This column lists the \(\lambda\) gene sequences assigned to the cloned DNAs. See the text for additional details.
sequences, respectively. Our current nucleotide sequencing studies have demonstrated that the \( V \) sequence contained in \( \lambda 99 \) and \( \lambda 303 \) is of the \( \kappa \) type (N. Hosumi, O. Bernard and S. Tonegawa, unpublished observations).

**Location of the \( \lambda \) Chain Gene Sequence in the DNA Clones**

The position of \( \lambda \) chain gene sequences in the \( \lambda 25a \) and \( \lambda 99 \) DNA clones was determined by \( R \) loop mapping using \( \lambda \) mRNA purified from H 200 myeloma (White and Hogness, 1977).

\( R \) loop molecules formed with the 8.6 kb \( \lambda 25a \) fragment displayed a double loop structure composed of a 410 nucleotide \( R \) loop located 3.9 kb from one end and a 1.2 kb double-stranded DNA loop (Figure 3a and Table 2). This structure closely resembles the triple loop formed by \( \lambda 303a \) (Brack and Tonegawa, 1977), except that the \( \lambda 25a \) hybrids have a long RNA tail (~260 bases) instead of the second, smaller \( R \) loop (Figure 4 and Table 2).

Since the \( \lambda 25a \) fragment showed homology only with \( C_\kappa \) and not with \( V \) sequences (see above), we conclude that the 410 bp \( R \) loop contains the \( C_\kappa \) 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) sometimes observed 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 \( \lambda 25a \) DNA contains a short homology region that hybridizes to a region near the V-C junction of the mRNA molecule. This second homology region, which we call the \( J \) sequence, is separated from the \( C_\kappa \) sequence by 1200 bp. 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 \( \lambda 99 \) fragment formed a single \( R \) loop very similar to the one observed in \( \lambda 13a \) (Tonegawa et al., 1977). It is 380 nucleotides long, lies in the middle of the DNA fragment (that is, 1.65-1.86 kb from either end) and carries a ~340
Figure 3. R Loop Molecules Obtained by Hybridizing HOPC 2020 λ, mRNA with the Eco RI Fragments of the DNA Clones

(a) 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.

(b) 99A 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.
<table>
<thead>
<tr>
<th>DNA Clone</th>
<th>RNA/DNA*</th>
<th>F†</th>
<th>N‡</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ig 13λ</td>
<td>8</td>
<td>27.2</td>
<td>34</td>
<td>3.29 ± 0.15</td>
<td>1.06 ± 0.22</td>
<td>0.388 ± 0.103</td>
<td>0.400 ± 0.200§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ig 99λ</td>
<td>5</td>
<td>81.2</td>
<td>40</td>
<td>1.65 ± 0.07</td>
<td>1.66 ± 0.11</td>
<td>0.375 ± 0.050</td>
<td>0.344 ± 0.099§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ig 25λ</td>
<td>10</td>
<td>84.0</td>
<td>36</td>
<td>3.09 ± 0.14</td>
<td>3.89 ± 0.15</td>
<td>1.22 ± 0.08</td>
<td>0.255 ± 0.065§</td>
<td>0.412 ± 0.048</td>
<td>0.10 ± 0.05</td>
</tr>
<tr>
<td>Ig 303λ</td>
<td>66</td>
<td>50.4</td>
<td>53</td>
<td>1.66 ± 0.08</td>
<td>3.75 ± 0.20</td>
<td>1.20 ± 0.09</td>
<td>0.380 ± 0.030</td>
<td>0.440 ± 0.040</td>
<td>0.09 ± 0.04</td>
</tr>
</tbody>
</table>

* Molar ratio of mRNA over DNA.
† (F) frequency of R loop molecules (in %), calculated by screening 250 molecules per sample.
‡ (N) number of molecules measured.
§ RNA tails do not always fully spread out and, therefore, these values cannot be compared directly with those of corresponding R loops. For identification of the various segments, see Figure 4. The values of mean lengths and standard deviations are given in kb.
Figure 4. Schematic Representation of R Loops on All Four Cloned Mouse Ig DNA Fragments
The relation between the four clones is shown. See Table 2 for the length measurements of various parts of the R loops.
<table>
<thead>
<tr>
<th>Heteroduplex</th>
<th>N</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Ig 303λ × Ig 25λ</td>
<td>51</td>
<td>5.47 ± 0.25</td>
<td>1.98 ± 0.22</td>
<td>2.87 ± 0.20</td>
<td>—</td>
</tr>
<tr>
<td>(2) Ig 303λ × Ig 99λ</td>
<td>48</td>
<td>5.48 ± 0.48</td>
<td>1.98 ± 0.52</td>
<td>—</td>
<td>1.53 ± 0.13</td>
</tr>
<tr>
<td>(3) Ig 303λ × Ig 99λ × Ig 25λ</td>
<td>27</td>
<td>5.54 ± 0.30</td>
<td>1.87 ± 0.19</td>
<td>2.85 ± 0.27</td>
<td>1.45 ± 0.15</td>
</tr>
</tbody>
</table>

For identification of the various molecules (1–3) and the different segments a–d, see Figure 6. All lengths (mean length and standard deviation) are given in kb. (N) number of measured molecules.
Figure 9. Heteroduplex Molecules Formed by Various Combinations of the Three s, Sequences-Containing Closed DNA Fragments
(a) Mammalian DNA (30°S) versus avian erythrocyte DNA (30°S).
(b) Mammalian DNA (30°S) versus avian erythrocyte DNA (15°).
(c) Combination of all three strains gives double heteroduplex structure.
Figure 6. Interpretation of the Heteroduplex Molecules

See Table 3 for the lengths of the various segments. The positions of V and C gene sequences (white boxes) were deduced from R loop molecules. See the Discussion for the exact position of the J sequence.
Figure 8. Arrangement of Mouse λ Gene Sequences in Embryos and λ Chain-Producing Plasma Cells

In embryo DNA, a full λ gene sequence consists of two parts that lie on two separate Eco RI fragments. On one of these fragments, 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 intervening sequence (I). On the second Eco RI fragment, the coding sequence is also split into two parts by a 1250 base long intervening sequence (I2). The two parts are for the constant region peptides (C) and approximately 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 chain-producing myeloma (H 2020), the λ gene sequence is rearranged as a result of one (or more) recombination(s) that involves 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 into direct contact. The limits of the corresponding sequences in the embryo and the myeloma DNAs 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. Neither is it intended to imply that the embryo and myeloma V sequences are identical. Additional short intervening sequences may be present in the C sequences. Arrows with numbers indicate Eco RI sites.
An Immunoglobulin Heavy Chain Variable Region Gene Is Generated from Three Segments of DNA: \( V_H, D \) and \( J_H \)

P. Early, H. Huang, M. Davis, K. Calame and L. Hood
Division of Biology
California Institute of Technology
Pasadena, California 91125

Introduction
The immunoglobulin molecule is a complex entity with two major functions—recognition of foreign substances (antigen binding) and the elimination or destruction of these foreign substances (effector func-