Volume 8 Number 23 1980 Nucleic Acids Research

Organization of the sequences flanking immunoglobulin heavy chain genes and their role in class switching

Brett M.Tyler and Jerry M.Adams

Molecular Biology Laboratory, The Walter and Eliza Hall Institute of Medical Research, P.O. Royal Melbourne Hospital, Victoria 3050, Australia

Received 22 September 1980

SUMMARY

We have used heteroduplex analysis to investigate the sequences surrounding the germline Cγ1 and Cγ3 genes, and to compare them with those surrounding the Cμ gene. We detected an inverted pseudogene 5' to the Cγ3 gene and 50-65% homologous to it. A 400 bp region of the Cγ1 and Cγ3 3' flanking sequences was conserved as strongly as the genes (65-80%), suggesting it may have a specific function. The sequences 5' to the Cγ1 and Cγ3 genes and possibly also the Cμ gene are composed of tandem partially homologous repeats of a similar 250 bp unit, arranged in the case of the Cγ1 gene, in 2-5 kb blocks of alternating orientation. These repeats comprised over 13 kb of the spacer region separating the Cγ3 and Cγ1 genes. Recombination sites for heavy chain class switching fell within these repeated sequences, suggesting that recombination between partially homologous blocks of repeat sequences 5' to C_H genes generates the deletion responsible for class switching. This hypothesis was strongly supported by an examination of published nucleotide sequences around the recombination sites of rearranged Cγ1 and Cγ2b genes (1,2).

INTRODUCTION

Little is yet known about the organization or role of the sequences flanking eukaryotic genes. The sequences flanking immunoglobulin genes are of particular interest since both the light (L) and heavy (H) polypeptide chains which comprise immunoglobulin (Ig) molecules are encoded in the germline by separate genes for their N-terminal variable (V) and C-terminal constant (C) regions (3,4; for a review, see 5). In the mouse, the different classes of heavy chains (μ, δ, γ_1, γ_{2a}, γ_{2b}, γ_3, α and ε) are encoded by a cluster of C_H genes probably in the order C_μ, C_δ, $C_{\gamma 3}$, $C_{\gamma 1}$, $C_{\gamma 2b}$, $C_{\gamma 2a}$, C_α (5,6,7), and separate from the larger cluster of V_H genes (8). Very large spacer regions separate the $C_{\gamma 1}$ and $C_{\gamma 2b}$ genes (21,100 bp; 21.1 kb) (9) and also the $C_{\gamma 3}$ and $C_{\gamma 1}$ genes (>27 kb) (S. Cory, unpublished results).

The functional light and heavy chain genes in an antibody producing cell are formed by deletions which bring a V gene into close proximity with the C gene (7,10). Heavy chain gene expression requires an additional DNA rearrangement since, during the maturation of an Ig producing cell, a single V_H gene

may be expressed with a series of different C_H genes. Recent evidence suggests that this 'class switching' is achieved by a deletion which transfers the V_H gene from its initial point of joining, the C_μ gene, to a new C_H gene (6,7,11), the recombination site lying 4-5 kb 5' to the C_μ gene and 2-3 kb 5' to the C_H gene (2,9,11). The mechanism of class switching should therefore be reflected in the organization of the 5' flanking sequences of different C_H genes. The sequences flanking C_H genes should also reflect the evolution of the structural genes. In particular, any regions which have specific functions related to C_H gene expression should be conserved.

We have previously isolated clones of embryonic $C_{\gamma 1}$ and $C_{\gamma 3}$ genes (12; E. Webb, S. Cory, S. Gerondakis, J.M.A., manuscript in preparation) and of a plasmacytoma C_μ gene (13) (summarized in Table 1). Each clone contains extensive regions of flanking sequences. In this paper, we analyze, by electron microscopy, the structures of the sequences flanking the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes and compare them with each other and with those flanking the C_μ gene. Our most significant finding is that partially homologous blocks of tandemly repeated

Table 1. Summary of clones.

Clone	Gene	Extent of Flanking Sequences (kb) 5'	3'	Orientation of Insert[d]
G1.1[a]	$C_{\gamma 1}$	12.9[c]	0.3	L-5'-3'-R
G1.2[a]	$C_{\gamma 1}$	9.3[c]	3.0	R-5'-3'-L
G3.1[a]	$C_{\gamma 3}$	8.5	6.6	R-5'-3'-L
G3.2[a]	$C_{\gamma 3}$	8.5	6.6	L-5'-3'-R
μ1[b]	C_μ	3.8	3.8	R-5'-3'-L

[a] Derived from embryo DNA (E. Webb, S. Cory, S. Gerondakis, J.M.A., manuscript in preparation).

[b] Derived from plasmacytoma HPC76 (13).

[c] G1.1 and G1.2 terminate on the 5' side of the gene at same point within the genome. However G1.1 has suffered a 2.7 kb deletion and G1.2, a 6.1 kb deletion during cloning.

[d] L and R refer to the left and right vector arms respectively. Hence L-5'-3'-R indicates that the 5' terminus of the insert is joined to the left vector arm.

sequences are located 5' to the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes and possibly also the C_{μ} gene and that these blocks contain the recombination sites for class switching. We propose therefore that homologous recombination between these blocks of repeated sequences generates the deletion responsible for class switching.

MATERIALS AND METHODS

Construction of heteroduplexes. For most experiments, the two DNA species (0.5-1.1 µg/ml each) were mixed in 60% recrystallized formamide (Merck)/80 mM 1,4 piperazine diethanesulfonic acid (Pipes) pH 7.9/5 mM Tris.HCl pH 7.9/0.42 mM NaCl/3 mM EDTA, denatured at 80°C for 3 min, then annealed at 20°C for 15-60 min depending on the experiment. For G3.1/G1.2 and G3.2/G1.1 heteroduplexes (Figs. 2 and 3), in order to favour interstrand hybridization over intrastrand hybridization, the DNAs were incubated at 50°C for 30 min, then diluted 5 fold into hyperphase solution without cytochrome C and allowed to anneal for a further 15 min at 20°C before spreading.

To form snapbacks with single stranded DNA molecules (Fig. 1B), the DNA (2.2 µg/ml) was denatured, then immediately diluted 5 fold into the hyperphase solution without cytochrome C and allowed to anneal for 30 min at 20°C before spreading.

Electron microscopy of heteroduplexes. Three sets of conditions were used for preparing heteroduplexes for electron microscopy. For spreading at low stringency, the heteroduplex mixture was diluted 5 fold to yield a hyperphase containing 50% recrystallized formamide (Merck)/0.2 M NaCl/100 mM Tris.HCl pH 8.5/10 mM EDTA/50 µg/ml cytochrome C (Sigma or Calbiochem) which was spread onto a hypophase containing 15% or 20% unpurified formamide (Fluka or Merck)/20 mM Tris.HCl pH 8.5/2 mM EDTA. For moderate stringency the final hyperphase was 60% recrystallized formamide/100 mM NaCl/100 mM Tris.HCl pH 8.5 /10 mM EDTA/50 µg/ml cytochrome C and the hypophase was 30% unpurified formamide/ 10 mM Tris.HCl pH 8.5/1 mM EDTA. For high stringency, the DNA was diluted 7.5 fold to give a hyperphase of 75% formamide/ 75 mM NaCl/50 mM Tris.HCl pH 8.5/5 mM EDTA/50 µg/ml cytochrome C. The hypophase was 40% formamide in twice distilled water.

To estimate the stringencies of the hyperphases and hypophases we used the formula $Tm = 81.5 + 0.41\ (\%\ G+C) + 16.6 \log_{10} (X^+) - 0.7\ (\%\ HCONH_2)$ (14). This formula neglects any effect of the cytochrome C, but it accurately predicted the Tm of T_7 DNA even in its presence (14). Using the formula, the conditions described above are all isodenaturing and, assuming a 50% G+C content, and a spreading temperature of 20°C, correspond to Tm -35°C at low

stringency, Tm -25°C at moderate stringency and Tm -13°C at high stringency.

After spreading, the heteroduplexes were picked up on parlodion coated grids, stained with 0.1 mM uranyl acetate, rinsed once in 90% ethanol, blotted dry, rotary shadowed with platinum/palladium (80:20), and viewed using a Phillips EM300. Photographed molecules were measured from prints using a Numonics Graphics Calculator 240.

Estimation of sequence homology from isodenaturing spreads of heteroduplexes. To calculate the homology between two sequences which hybridized under given conditions, we used the formula: 1°C depression of Tm = 1.5% mismatch (15). Hence sequences which hybridize at Tm -35°C but not at Tm -25°C (e.g. the $C_{\gamma 3}$ gene and its inverted repeat (Fig. 1)) must be between (100 - 1.5 x 35) % ≅50% and (100 - 1.5 x 25) % ≅65% homologous if their G+C content is 50%. Sequences which hybridize in only a proportion of heteroduplex molecules must be close to their Tm; hence their homology can be estimated from the temperature of spreading. For example, 450 bp of the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes hybridized in only 23% of G3.1/ G1.2 heteroduplexes spread at low stringeny (approx. Tm -35°C) (Fig. 2). Hence the mismatch must be approximately 1.5 x 35 = 52.5% and so the homology must be 45-50%, or, for a more variable and heterogeneous set of hybrids (e.g. Fig. 3), 40-55%.

RESULTS

A partially homologous inverted repeat of the $C_{\gamma 3}$ gene. When we examined the sequences surrounding the $C_{\gamma 3}$, $C_{\gamma 1}$ and C_{μ} genes for inverted repeat sequences using heteroduplex analysis, we detected two pairs of such sequences in the $C_{\gamma 3}$ clones under conditions of low stringency (approx. Tm -35°C: see Materials and Methods) (Fig. 1A). As shown in Fig. 1C, these two pairs of inverted repeats (a and a', b and b') are arranged one pair within the other. Repeats a and a' did not hybridize under conditions of moderate stringency (approx. Tm -25°C), indicating that these two sequences were only 50-65% homologous (see Materials and Methods), while b and b' remained hybridized even at high stringency (approx. Tm -13°C), indicating that they were >80% homologous. When the snapbacks were aligned with the restriction map (Fig. 1C), sequence a' coincided with the $C_{\gamma 3}$ structural gene. We confirmed our alignment by examining snapbacks formed by the 5' terminal EcoRI/Bam fragment of the mouse DNA insert. Fig. 1B shows that, as expected from the map, this fragment displayed only hybridization between b and b' (arrowed). Since the Bam site at one end of the fragment falls just inside the $C_{\gamma 3}$ gene, the distance from the snapback to this Bam site (the short arm in molecules like that in

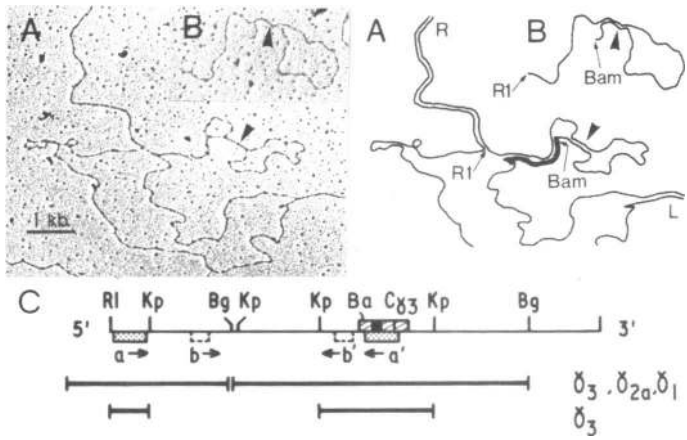

Figure 1. Heteroduplexes of G3.1 showing inverted repeat sequences. (A) Heteroduplex between G3.1 (upper strand) and the vector CH4A formed at low stringency. L and R refer to the left and right vector arms respectively and the thickened region, the Cγ3 gene. (B) Snapback formed by the 5' EcoRI/Bam fragment of G3.1 (see C). (C) Location within G3.1 of the inverted repeat sequences and of the restriction fragments to which cloned Cγ probes hybridized. In B and C, RI, Kp, Ba and Bg indicate EcoRI, Kpn 1, Bam HI and Bgl 2 cleavage sites. Arrows indicate orientations. Black regions within the gene (the hatched box) indicate intervening sequences. The bars indicate the DNA fragments to which the various probes (indicated on the right) hybridized. Hybridization was at 50°C in 0.75 M NaCl, 0.075 M Na_3 citrate.

Fig. 1B), gives the distance from the snapback to the 5' end of the gene. This was 370 ± 30 bp (standard error: s.e.). Since in the heteroduplexes of Fig. 1A, the distance from the b.b' snapback to sequence a' was 420 ± 100 bp (s.e.), we concluded that, as shown in Fig. 1C, sequence a' did correspond to the $C_{\gamma 3}$ gene and therefore that sequence a was a partially homologous copy of the gene. The coding and intervening sequences must have been equally well conserved within the inverted repeat since the snapbacks between the gene and its repeat included the large intervening sequence in the gene which separates the CH1 and CH2 domain coding segments.

The presence of the repeated gene sequence was independently confirmed by filter hybridization of cloned $C_{\gamma 3}$, $C_{\gamma 2a}$ and $C_{\gamma 1}$ probes to restriction digests of G3.1 under conditions of very low stringency. As summarized in Fig. 1C, these probes hybridized not only to the Bgl 2 and Kpn 1 fragments containing the genes, but also to the fragments predicted by electron microscopy to contain the inverted repeat of the gene.

Homology between the DNA fragments containing the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes. To determine the amount of homology between the sequences surrounding the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes, we examined G1.1/G3.2 and G1.2/G3.1 heteroduplexes, in which the mouse DNA inserts have the same orientation. These molecules showed two reproducible regions of homology at low stringency as illustrated by the panel of molecules in Fig. 2. The first, HRI, was present in both G1.2/G3.1 (Fig. 2A) and G1.1/G3.1 (Fig. 2B) heteroduplexes. It presumably corresponds to the structural gene sequence, since in G1.1 it mapped 300 bp from the right vector arm, precisely the position of the $C_{\gamma 1}$ gene (Fig. 2B). HRI also coincided with the $C_{\gamma 1}$ gene in G1.2 and the $C_{\gamma 3}$ gene in G3.1 and G3.2. In two-thirds of the G1.2/G3.1 heteroduplexes (26/42) and in all G1.1/G3.2 heteroduplexes (15/15), HRI was interrupted by an approx. 470 bp region of non-hybridization, which corresponds closely in size and position to the intervening sequences within the $C_{\gamma 1}$ gene (16) and the $C_{\gamma 3}$ gene (E. Webb, S. Cory, S. Gerondakis, J.M.A., manuscript in preparation) which separate the CH1 and

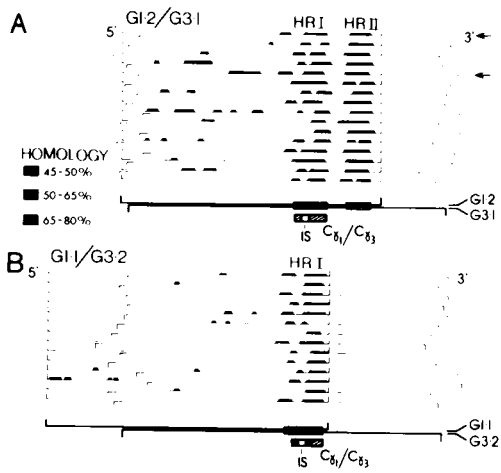

Figure 2. Heteroduplex analysis of homology between Cγ1 and Cγ3 clones. Insert lengths were normalized to 14.0 kb (G1.2), 14.7 kb (G1.1) or 16.7 kb (G3.1 and G3.2). In A and B, the Cγ1 strands are displayed, while the lines with down-turned tails indicate the positions of the ends of the Cγ3 insert in each molecule with respect to HRI and HRII. The filled blocks indicate hybridization with the G3.1 or G3.2 strand respectively. Summary diagrams below A and B give the mean positions of HRI and HRII within each clone calculated from 29 G1.2/G3.1 heteroduplexes (A) or 15 G1.1/G3.2 molecules (B), plus the degrees of homology within HRI and HRII. The diagonally hatched boxes indicate the positions of the genes within each clone and IS indicates an intervening sequence (16; E. Webb, S. Cory, S. Gerondakis and J.M.A., manuscript in preparation).

CH2 domain coding segments of each gene (Fig. 2). All of HRI except for the intervening sequence region was stable at moderate stringency (approx. Tm -25°C) but not at high stringency (approx. Tm -13°C), indicating that, as summarized in Fig. 2, the $C_{\gamma 1}$ and $C_{\gamma 3}$ coding sequences are 65-80% homologous (see Materials and Methods), but the intervening sequences only 45-50% homologous.

The second homology region (HRII) was located 1.0 kb to the 3' side of HRI and averaged 1.3 kb in length. In a quarter of the molecules (12/53), it was interrupted by a short (~300 bp) region of non-hybridization, indicating that the central region was poorly conserved. The 400 bp gene proximal region of HRII is as strongly conserved as the gene sequences, being 65-80% homologous (Fig. 2A). The gene distal 700 bp showed 50-60% homology.

The sequences flanking the 5' side of the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes are composed of tandem partially homologous repeats of the same unit. Both G1.2/G3.1 and G1.1/G3.1 heteroduplexes displayed at low stringency various regions of apparent hybridization between sequences 5' to the genes (Fig. 2). The variability of the hybridization indicates that under the spreading conditions, the hybrids were close to their Tm, and hence that the hybridizing regions must be 40-55% homologous.

Several alternative patterns of hybridization were observed (Fig. 3). Because we could not readily distinguish the G1.2 and G3.1 strands in these heteroduplexes, we plotted both strands from the 5' terminus of the insert to the midpoint between HRI and HRII. Fig. 3A shows two heteroduplexes in which the hybridization is in the same register; that is, the distance from the hybridizing region to the genes is the same along both strands. In B, hybridization has occurred to essentially the same regions as in A, but the distances from the hybridizing region to the genes differ by about 1 kb along the two strands. In C, this difference of register is ~3 kb. In D, as indicated by the oblique lines, the molecules show changes of register within otherwise continuous regions of hybridization. In E, multiple shifts of register occur within a single molecule.

These variable patterns of hybridization and changes of register imply that the hybridizing sequences are repeated several times on at least one of the strands. In fact, the sequences must be repeated on both the G1.2 and G3.1 strands since some molecules, e.g. those arrowed in Fig. 3, exhibited shifts of register markedly shorter than the hybridizing regions. These short shifts can only occur if the hybridizing regions are composed of multiple copies of a short repeating unit, the lengths of the shifts of register being

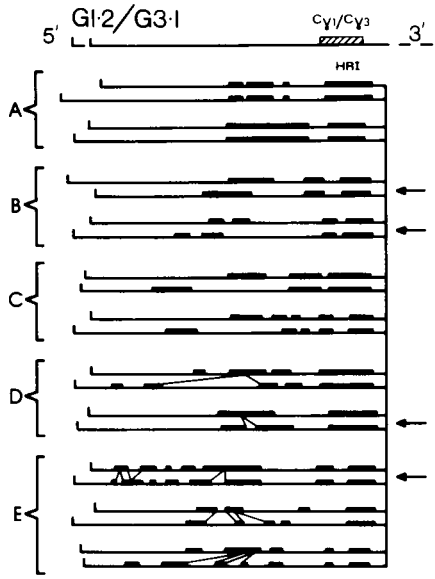

Figure 3. Alternative patterns of hybridization between sequences 5' to the Cγ1 and Cγ3 genes in G1.2/G3.1 heteroduplexes. Both strands of each heteroduplex are shown plotted from the 5' end of the insert to the midpoint between HRI and HRII without distinguishing between the G3.1 and G1.1 strands. In each case, the sum of the lengths of the two strands was normalized to 22.6 kb (11.6 for G1.2 + 11.0 for G3.1). The filled blocks indicate regions of hybridization.

equal to the repeat length or to integral multiples of it.

To determine the length of the repeat unit, we measured 53 changes of register on 35 molecules. As shown in Table 2, the sizes of the smallest 39 changes fell into discrete groups corresponding to 1, 2 or 3 copies of a unit approximately 240 bp long. To estimate the repeat length more precisely, we

Table 2. Calculation of repeat unit lengths from sizes of changes of register.

Group[a]	Number of Measurements	Mean Length[b] (kb) ± s.d.	Differences Between Groups (kb)	Number of Repeat Units[c]
A	21	0.24 ± 0.06		1
B	9	0.48 ± 0.07	0.24	2
C	9	0.74 ± 0.07	0.26	3

[a] Groups of similarly sized shifts of register.

[b] s.d. is standard deviation.

[c] Since the differences between the group means are approximately equal to each other and to the smallest group mean (A), these differences, and also mean A, must represent one repeat unit length. Hence, means B and C must represent two and three repeat units respectively.

determined the slope of a plot of the lengths of individual shifts of register against their estimated repeat unit number (from Table 2), using regression analysis. The length was 245 ± 10 bp (s.e.).

To determine the distribution of the repeated sequences in G1.2 and G3.1, we examined G1.2/G3.1 heteroduplexes in which the G3.1 strand could be distinguished by its characteristic snapback. Hybridization indicative of repeated sequences was observed in G3.1 over the entire region lying between the inner pair of inverted repeats (b and b' in Fig. 1C) and in two molecules, also to the 1.6 kb sequence between b and a. Thus as summarized in Fig. 8C, apart from the inverted repeats described above, it is likely that the entire sequence 5' to the $C_{\gamma 3}$ gene in G3.1 is composed of multiple copies of a 250 bp repeat unit. In G1.2, the repeated sequences probably comprise all of the 8.6 kb region from 1.0 kb 5' to the $C_{\gamma 1}$ gene to 3.3 kb from the 5' end of the insert (Fig. 8A). Two molecules showed hybridization to within 1.6 kb of the end of the insert, indicative of further repeats within this region.

We obtained further evidence for repetitive sequences 5' to the $C_{\gamma 1}$ gene and for the repeat length from heteroduplexes of the 5' terminal Hind III/ EcoRI fragments of G1.1 and G1.2, which had been subcloned into pBR322. Since G1.2 has a 3.4 kb deletion within this segment with respect to G1.1 (E. Webb, S. Cory, S. Gerondakis, J.M.A., manuscript in preparation), the majority of heteroduplexes displayed a single large deletion loop (Fig. 4A). However, in about 20% of molecules this was replaced by two or three smaller loops of varying sizes. As illustrated in Fig. 4F, these alternative structures imply the presence of tandemly repeated sequences within and immediately adjacent to the region of G1.1 deleted from G1.2. The multiple small loops result from hybridization between a sequence in pG1.2 adjacent to the 3' end of the deletion and different repeated homologous sequences within the deletion loop of pG1.1. The sizes of the smallest loops formed a clear progression corresponding to two copies (e.g. Fig. 4B & E), three, four (e.g. D & E) or five copies (e.g. C) of a unit whose length was 255 ± 10 bp (s.e.), as determined by regression analysis. This length is in close agreement ($p>0.5$) with that obtained from G1.2/G3.1 heteroduplexes (245 ± 10 bp).

The heteroduplexes of Fig. 4 were formed and spread at moderate stringency. Hence the homology between the repeat units must be at least 65%. In fact, it must be considerably greater for hybrids between non-identical repeat units (e.g. b-e) to be resistant to displacement by the fully homologous sequence.

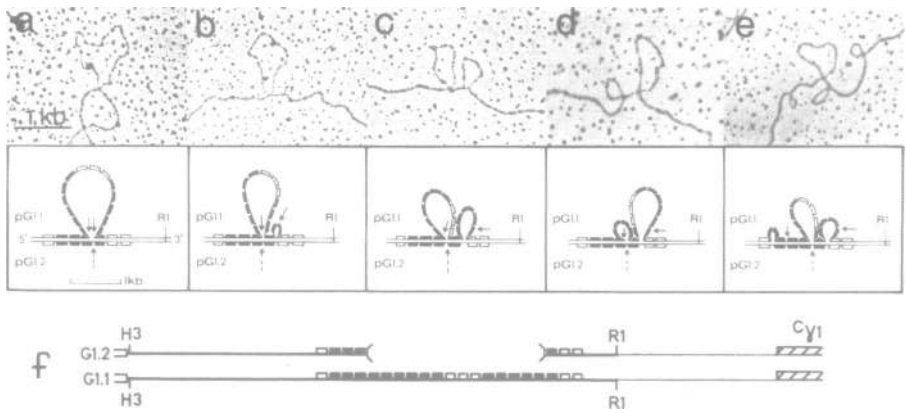

Figure 4. Repeated sequences 5' to the Cγ1 gene revealed by pG1.1/pG1.2 heteroduplexes. In each heteroduplex, the arm extending left to the cleaved Hind III site was distinguished by length from the one extending to the right across the EcoRI site and into the plasmid sequences (4.8 kb cf 5.7 kb respectively). In the schematic representations and in f, filled blocks represent repeat units inferred directly from particular classes of heteroduplexes (including those shown); open blocks represent those inferred by extrapolation. Continuous arrows mark the end points of the deleted region in pG1.1 and broken arrows the position of the deletion in pG1.2. RI and H3 indicate EcoRI and Hind III cleavage sites. In f, the bold lines indicate the Hind III/EcoRI fragments of G1.2 and G1.1 that were subcloned in pBR322 to form pG1.2 and pG1.1.

The repeat units 5' to the $C_{\gamma 1}$ gene are arranged in 2-5 kb blocks of alternating orientation. When we examined G1.1/Ch4A heteroduplexes for inverted repeat sequences, we frequently observed snapback structures at low stringency (see Materials and Methods). Since these did not have a reproducible position or size and were usually very short (<300 bp), we plotted 48 of the apparent snapbacks on a two-dimensional map of the clone by plotting the location of the gene-proximal strand of the snapback stem against that of the gene distal strand. Specific hybridization should appear as tight clusters on such a map, whereas artefacts should be randomly scattered.

The snapbacks fell into four broad clusters (Fig. 5A). Their statistical significance was determined using David's Empty-Cell Test (17): when the map was divided into 50 equal areas, 28 contained no lollipop stems indicating that the snapbacks resulted from specific hybridization ($p<0.001$). However, from the broadness of the clusters, we inferred that the hybridization was between blocks of repeated sequences arranged in opposite orientations rather than unique inverted repeats. Fig. 6 illustrates the variable hybridization

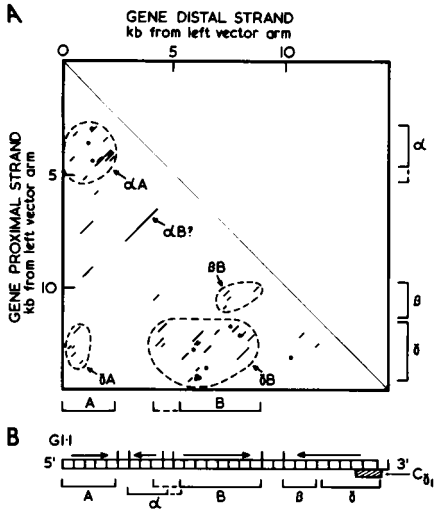

Figure 5. Blocks of repeated sequences arranged in alternating orientations revealed by G1.1 snapbacks. (A) Two-dimensional map of G1.1 snapbacks. The locations and sizes of all potential snapbacks (54), including simple crossovers, were determined in 18 G1.1/Ch4A and 10 G1.1/G3.1 heteroduplex molecules and plotted as described in the text. Lines represent apparent double-stranded regions, and dots, crossovers. (B) Locations and orientations of the blocks of repeat sequences within G1.1. For clarity, the repeat units are drawn as twice the length of the actual repeat unit. Arrows indicate relative orientations.

which can occur between two poorly homologous blocks of repeated sequences and which would result in the dispersed type of clusters seen in Fig. 5A.

From the positions of the clusters in Fig. 5A, we infer that, as shown in Fig. 5B, the G1.1 snapbacks result from variable hybridization between five essentially non-overlapping blocks of repeat units designated α, β and γ (one orientation) and A and B (opposite orientation). These blocks comprise almost all of the $C_{\gamma 1}$ 5' flanking sequences. The uneven distribution of snapback hybridization between the different blocks presumably reflects different degrees of homology. A portion of region γ overlaps the $C_{\gamma 1}$ structural gene. This may be due to measuring errors or to artefactual 'snapbacks', but it is also possible that the sequences 5' to the $C_{\gamma 1}$ gene are related to sequences within the gene as observed in the Drosophila rRNA genes (18), and the chicken conalbumin gene (19).

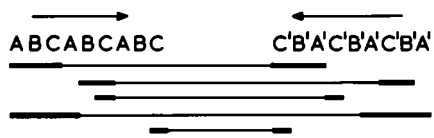

Figure 6. Snapback hybridization between two hypothetical blocks of repeated sequences arranged in opposite orientations. Five possible snapbacks are shown, the heavy blocks indicating the regions of hybridization. The arrows indicate the relative orientations of the blocks.

As shown previously, much of the $C_{\gamma 3}$ 5' flanking sequences is composed of multiple copies of a repeat unit homologous to that found 5' to the $C_{\gamma 1}$ gene (see Fig. 3). Examination of individual G1.2/G3.1 heteroduplexes showed that as summarized in Fig. 8C, probably all of the units 3' to inverted repeat b have the orientation of blocks α, β and γ of G1.1, while those 5' to inverted repeat b have the opposite orientation.

Extragenic sequences conserved between C_H genes. Since DNA regions important in the expression of C_H genes may have been conserved, we examined heteroduplexes between the C_μ and $C_{\gamma 1}$, and C_μ and $C_{\gamma 3}$ clones for regions of homology. Fig. 7A shows histograms of hybridization at low stringency between the $C_{\gamma 3}$ and C_μ clones, G3.1 and μ1. A single short region of homology (<100 bp) was observed 0.9 kb 3' to the C_μ gene and 3.6 kb 3' to the $C_{\gamma 3}$ gene. This was not present at high stringency (~Tm -13°C). At low stringency, the C_μ and $C_{\gamma 1}$ clones μ1 and G1.2 hybridized in three places (Fig. 7B). As shown by the vertical lines, the region located 3' to the two genes coincided with the region of the C_μ clone homologous to the $C_{\gamma 3}$ clone ($p>0.5$), suggesting a sequence common to all three C_H clones. The existence of a common sequence was further indicated by hybridization (not shown) of the three C_H clones to a V_H clone, M48 (8) which also appeared to contain the sequence (Figs. 8A, C & E).

The conservation of a sequence 3' to all three C_H genes examined is analogous to the conservation of a short sequence 3' to the mouse α and β^{maj}

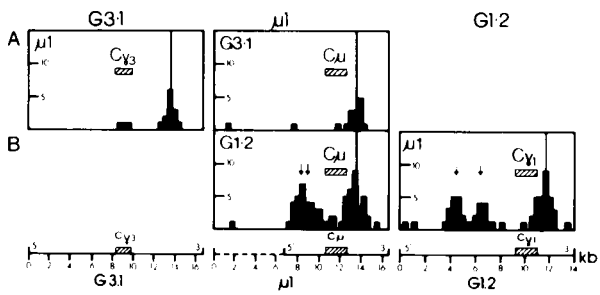

Figure 7. Histogram analysis of homology between three C_H clones. A shows hybridization between G3.1 and μ1 and B between μ1 and G1.2. The number of molecules displaying strand contact at low stringency were plotted for each position along both strands. In a particular heteroduplex experiment, the different strands were plotted in separate panels, the positions of the gene within a strand being indicated by a hatched box. The 5' terminal 6.8 kb EcoRI fragment of the μ1 insert, indicated by the broken line, is derived from the vector (13).

globin genes (20) and suggests that the sequence may have a specific function. Alternatively, it may correspond to the interspersed moderately repeated class of sequences (21).

The two other regions of homology between μ1 and G1.2 (arrowed in Fig. 7B) mapped 5' to both genes and within the region of G1.2 composed of tandemly repeated sequences (although the μ1 strand shows only a single histogram peak, individual molecules clearly showed two separate hybridizing regions). This suggests that part of the C_μ 5' flanking sequences may also be tandemly repetitious. Although no similar hybridization was observed between the $C_{\gamma 3}$ and C_μ 5' regions, the relevant sequences may have been deleted from our C_μ clone (Fig. 8E) (13).

DISCUSSION

Fig. 8 summarizes the organization of the sequences surrounding the $C_{\gamma 1}$, $C_{\gamma 3}$ and C_μ genes. The sequences 5' to the $C_{\gamma 1}$ genes (Fig. 8A and B) are composed of tandem partially homologous repeats of a 250 bp unit (the open boxes), arranged in four blocks with alternating orientations: A, α, B and β-γ. In clone G1.2, most of block B has been lost through a cloning deletion. The sequences 5' to the $C_{\gamma 3}$ gene (8C) are also primarily composed of tandem repeats of a 250 bp unit 40-55% homologous to that comprising the $C_{\gamma 1}$ 5' region. The $C_{\gamma 3}$ 5' region contains, in addition, a partially homologous (50-65%) inverted repeat of the $C_{\gamma 3}$ gene (sequence a in Fig. 8C and D) and a pair of inverted repeats (the stippled boxes b and b') which are >80% homologous to one another. The strongest regions of homology between the $C_{\gamma 1}$ and $C_{\gamma 3}$ clones (Figs. 2 and 8A-D) are HRI, which corresponds to the two genes, and HRII, located 1 kb 3' to the genes. A short sequence is also conserved between 3' flanking regions of the C_μ, $C_{\gamma 1}$ and $C_{\gamma 3}$ genes (the short filled boxes in A, C and E) while the $C_{\gamma 1}$ and C_μ clones share two short regions of homology 5' to the genes located within the repeated sequences 5' to the $C_{\gamma 1}$ gene. The significance of these various structural features is discussed below.

An inverted C_γ pseudogene. The inverted sequence homologous to the $C_{\gamma 3}$ gene ('a' in Figs. 8C and D) is probably a non-functional 'pseudogene' and does not encode a novel class of γ chains since its orientation is incompatible with the deletion mechanism responsible for formation of an expressed heavy chain gene (6,7). Moreover, the sequence hybridizes equally well to both the coding and intervening sequences of the $C_{\gamma 3}$ gene, in contrast to the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes where the intervening sequences rarely hybridized. The

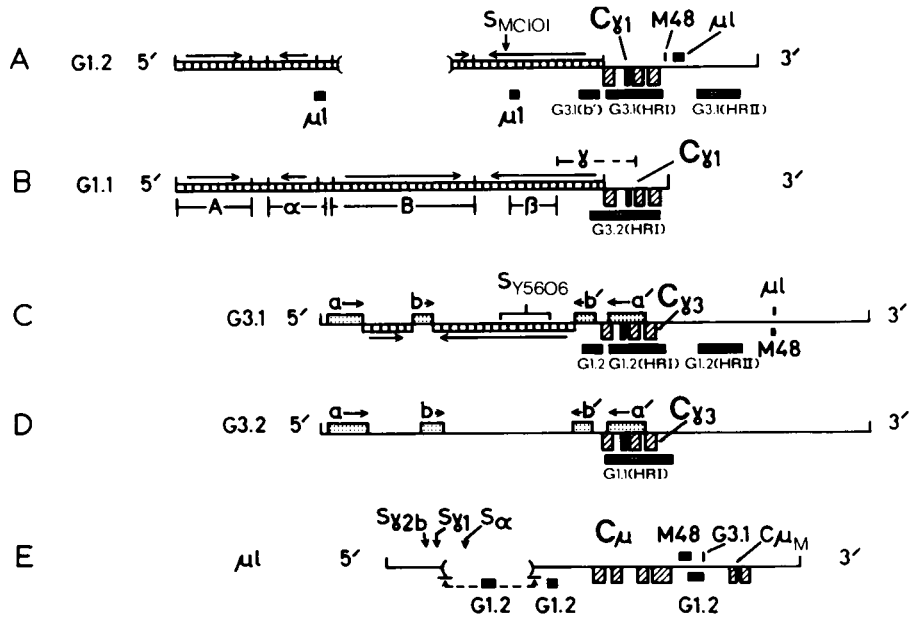

Figure 8. Structure of sequences surrounding the Cγ1, Cγ3 and Cμ genes. Filled boxes represent regions of homology; open boxes, repeat units; and stippled boxes, inverted repeats. G3.2 was not analyzed independently for the presence of repeated sequences. S_{MC101} and S_{Y5606} indicate the switch recombination sites of the plasmacytomas MC101 (1) and Y5606 (S. Cory, E. Webb, J. Gough and J.M.A., submitted) respectively. Sγ1, Sγ2b and Sα indicate the positions within the Cμ 5' flanking sequences of recombination sites for switching to Cγ1, Cγ2b and Cα. CμM is the segment of the Cμ gene encoding the C-terminus of μ chains in membrane bound IgM (35).

regions of the sequence corresponding to the $C_{\gamma 3}$ coding sequences have thus been no more strongly conserved than the region corresponding to the intervening sequences. An inverted pseudogene has been described for a Drosophila heat shock protein (22) while directly repeated pseudogenes exist of the 5S rRNA genes of Xenopus (23), and of several globin genes (24,25,26).

Conservation of the C_{γ} genes and their flanking sequences. The homology between the coding sequences of the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes is about 65-80%. By contrast, that between the large intervening sequences of each gene is only 45-50%. This compares with 75% homology between the coding regions of the $C_{\gamma 1}$ and $C_{\gamma 2b}$ genes and 51% between the large intervening sequences (27). Poor conservation of intervening sequences has also been observed for globin genes (29,30), and the two rat preproinsulin genes (31). These observations argue

strongly that the function, if any, of intervening sequences is not dependent upon their nucleotide sequence.

The gene proximal 400 bp of HRII was the only flanking region to be as highly conserved as the coding sequences of the genes. This suggests it has a specific function. For instance it may represent the regions encoding the C-terminal sequences of the γ1 and γ3 chains of membrane bound IgG (32,33) like the region 3' to the C_μ gene which encodes the C-termini of membrane bound μ chains (34,35).

Spacer regions between C_H genes contain tandem repeats. We have shown that multiple partially homologous repeats of a similar 250 bp unit comprise the sequences 5' to both the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes (Figs. 8A-C). In agreement with this, we could identify a 240-245 bp repeat within the published nucleotide sequence of 600 bp of the $C_{\gamma 1}$ 5' flanking region (1) (Fig. 10). Since the $C_{\gamma 2a}$ and $C_{\gamma 2b}$ genes are more closely related to both the $C_{\gamma 1}$ and the $C_{\gamma 3}$ genes than the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes are to each other (28), we would expect that the $C_{\gamma 2a}$ and $C_{\gamma 2b}$ 5' flanking sequences are also tandemly repetitive.

The C_μ 5' flanking sequences are probably also repetitive since a region 1.5-2 kb 5' to the C_μ gene is partially homologous to the repetitive sequences 5' to the $C_{\gamma 1}$ gene (Figs. 7A and 8A and E) and since most of the embryonic C_μ clones described (1,9,36) contain cloning deletions 1-2 kb 5' to the gene, a characteristic of clones bearing tandemly repeated sequences (37). Thus tandem repeats may comprise the regions 5' to all C_H genes.

The spacer between the $C_{\gamma 3}$ and $C_{\gamma 1}$ genes is at least 27 kb long, the $C_{\gamma 3}$ gene lying 5' to the $C_{\gamma 1}$ gene (S. Cory, unpublished data). The 13 kb of repeated sequences present in G1.1 thus accounts for a substantial proportion of this distance. The spacer regions between the 5S and the 40S rRNA genes of Xenopus also consist largely of tandemly repeated sequences (38,39). On the other hand, those between the rabbit β-like globin genes (40) and between the histone genes of two sea urchin species (41,42) do not.

Homologous recombination between repetitive sequences may generate heavy chain class switching. The recombination sites for class switching for several rearranged C_H genes in Fig. 8 are indicated by 'S'. In the case of the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes, these sites clearly fall within repetitive sequences: in $C_{\gamma 1}/C_{\gamma 3}$ heteroduplexes, the regions containing the sites hybridized in several different registers (see Fig. 3) and uninterrupted by any visible region of non-homology. The switch sites 5' to the C_μ gene ($S_{\gamma 2b}$, $S_{\gamma 1}$ and S_α) also fall within regions which may be repetitive (see above). These observations suggest that, as illustrated in Fig. 9, the deletion responsible for

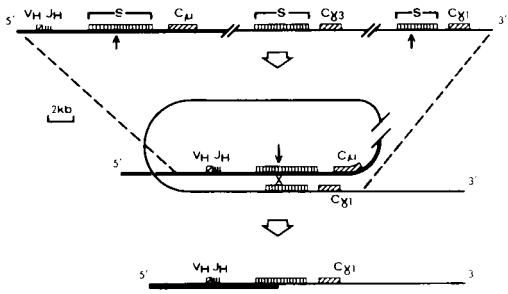

Figure 9. Generation of the class-switching deletion by homologous recombination. 'S' indicates the blocks of repeated sequences involved in the class switching recombination event. V_H and J_H indicate heavy chain variable and joining region genes. The small arrows indicate the point of recombination.

heavy chain class switching is generated by homologous recombination between partially homologous blocks of repeated sequences 5' to each of the C_H genes.

This hypothesis gained further support from nucleotide sequences published during the preparation of this manuscript, of short regions of 5' sequences surrounding the recombination sites of rearranged $C_{\gamma 1}$ and $C_{\gamma 2b}$ genes (1,2,43,44). Honjo and co-workers (1,2) sequenced three regions of DNA which comprised the rearranged $C_{\gamma 1}$ gene of MC101. Two were derived from the C_μ and $C_{\gamma 1}$ 5' regions (A and C in Fig. 10), while the third, a 492 bp sequence (B)

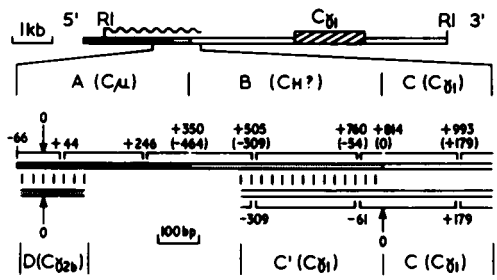

Figure 10. Structure of the sequences surrounding Cγ1 and Cγ2b switch recombination sites. The solid bar indicates sequences in the rearranged Cγ1 gene derived from Cμ flanking sequences and the open bar, from embryonic Cγ1 sequences while the stippled bar indicates a region interposed between the Cγ1-derived and Cμ-derived sequences (1,2). The cross-hatched bar indicates embryonic Cγ2b 5' sequences. The horizontal square brackets indicate repeating units while the numbers refer to the numbering of the nucleotide sequence in Fig. 7 of ref. 2 (above the line, unbracketed) and in Fig. 3 of ref. 1 (above the line in brackets and below the line). The arrows on the left indicate the Cμ-Cγ2b switch sites in MOPC 141 while that on the right indicates the site within the embryonic Cγ1 5' sequences of the Cμ-Cγ1 site in MC101 (2). The wavy line at the top of the figure indicates the sequenced region (1,2). The region 5' to nucleotide -66 in the μ sequence was sequenced, but we could identify no homologies or repeating units within it.

interposed between these two, was presumably derived from the 5' region of another C_H gene (2). We observed from the sequences presented that, as indicated by the horizontal square brackets in Fig. 10, all three regions were repetitive: each could be divided into a 210-250 bp repeat unit plus a portion of a second unit. In addition, within a 600 bp region of germline $C_{\gamma 1}$ 5' sequences containing the MC101 recombination site (C' + C), we could identify two 240 bp repeats and part of a third. The repetitive C_μ 5' region (A) also contained the single recombination site in the rearranged $C_{\gamma 2b}$ gene of MOPC141 (2,43). Dunnick and co-workers (44) sequenced part of the $C_{\gamma 1}$ switch region in the MOPC21 mutant IF2 and identified tandem 20 bp and 49 bp repeats, leading them to a similar model for C_H switching. These short repeats may represent components of the higher order 250 bp repeat observed here.

When we compared the regions 5' to the C_μ, $C_{\gamma 3}$, $C_{\gamma 1}$, $C_{\gamma 2b}$ genes and to a presumptive fifth C_H gene (the interposed region), we found that as summarized in Table 3, all were 50-55% homologous to one another. This homology, more extensive than that noted (2), comprises the whole of the sequenced $C_{\gamma 1}$, $C_{\gamma 2b}$ and interposed regions. In the case of the $C_{\gamma 1}$ and $C_{\gamma 3}$ genes, and possibly the C_μ gene and the unidentified C_H gene, these homologous regions are composed of tandem repeats of a 250 bp unit. All contain recombination sites for

Table 3. Mutual homology between heavy chain switch regions.

Region[a] 5' to	A C_μ	B (C_H)	C $C_{\gamma 1}$	D $C_{\gamma 2b}$	$C_{\gamma 3}$
A	100	50[d]	50	54	n.d.[b]
B		100	52	50	n.d.
C			100	51	40-55[c]
D				100	n.d.

[a] Regions A, B, C and D refer to the regions 5' to the Cμ, C_H (see text), Cγ1 and Cγ2b genes respectively, and are defined in Fig. 10.

[b] n.d. = not determined.

[c] From electron microscopy (this paper).

[d] Percentage homology was calculated using the total length of the paired sequences including gratuitous insertions. The length of such insertions introduced to maximize homology was no more than 20% of the length of either sequence; except for B and C where it was 30%. Large differences in length (>20 bp) were resolved primarily as a single deletion and excluded from the calculation.

C_H switching and, as indicated by the vertical bars in Fig. 10, for a rearranged $C_{\gamma 1}$ and $C_{\gamma 2b}$ gene these sites (arrowed) occur precisely in register with the nucleotide sequence homology between the recombining regions (A and D for $C_{\gamma 2b}$, B and C' for $C_{\gamma 1}$). This is strongly indicative of homologous recombination.

A key prediction of this model is that, for a given C_H gene, the site of recombination will vary in different plasma cells and plasmacytomas both with respect to the C_μ gene and with respect to the other C_H gene. Such variation with respect to the C_H gene has been reported for three rearranged $C_{\gamma 1}$ and two rearranged $C_{\gamma 2b}$ genes (1) and with respect to the C_μ gene for two rearranged $C_{\gamma 2a}$ and two rearranged C_α genes (S. Cory, E. Webb, J. Gough, J.M.A., submitted).

The mechanism proposed here for class switching differs markedly from the highly site specific joining of V and J genes (10,45). The differing mechanisms presumably reflect the requirement in V-J joining for a high degree of precision to maintain a functional polypeptide as opposed to C_H switching where recombination occurs within the J_H-C_H intervening sequence.

An unresolved aspect of this model is why 'switching' events do not occur at random during B lymphocyte ontogeny (32). Perhaps the homology between the switch regions is so low (40-60%) that a specialized switching recombinase is required. Alternatively, recombination in B lymphocytes and their progenitors might be repressed until class switching was required, e.g. until after contact with antigen.

ACKNOWLEDGMENTS

We thank E. Webb, S. Gerondakis, S. Cory, D. Kemp and W. Peacock for valuable discussions and A. Abbot and T. Mandel for patient advice on electron microscopy. This work was supported by the National Cancer Institute (RO1-CA12421), the American Heart Association and the National Health and Medical Research Council (Canberra).

REFERENCES

1. Kataoka, T., Kawakami, T., Takahashi, N. and Honjo, T. (1980) Proc. Nat. Acad. Sci. U.S.A. 77, 919-923.
2. Takahashi, N., Kataoka, T. and Honjo, T. (1980) Gene, In press.
3. Dreyer, W.J. and Bennett, J.L. (1965) Proc. Nat. Acad. Sci. U.S.A. 54, 864-869.
4. Tonegawa, S., Hozumi, N., Matthyssens, G. and Schuller, R. (1976) Cold Spring Harbor Symp. Quant. Biol. 41, 877-889.

5. Adams, J.M. (1980) Immunology Today 1, 10-17.
6. Honjo, T. and Kataoka, T. (1978) Proc. Nat. Acad. Sci. U.S.A. 75, 2140-2144.
7. Cory, S. and Adams, J.M. (1980) Cell 19, 37-51.
8. Kemp, D.J., Cory, S. and Adams, J.M. (1979) Proc. Nat. Acad. Sci. U.S.A. 76, 4627-4631.
9. Maki, R., Traunecker, A., Sakano, H., Roeder, W. and Tonegawa, S. (1980) Proc. Nat. Acad. Sci. U.S.A. 77, 2138-2142.
10. Brack, C., Hirama, M., Lenhard-Schuller, R. and Tonegawa, S. (1978) Cell 15, 1-14.
11. Davis, M.M., Calame, K., Early, P.W., Livant, D.L., Joho, R., Weissman, I.L. and Hood, L. (1980) Nature 283, 733-739.
12. Tyler, B.M. and Adams, J.M. (1980) Gene 10, 147-155.
13. Gough, N.M., Kemp, D.J., Tyler, B.M., Adams, J.M. and Cory, S. (1980a) Proc. Nat. Acad. Sci. U.S.A. 77, 554-558.
14. Davis, R.W. and Hyman, R.W. (1971) J. Mol. Biol. 62, 287-301.
15. Laird, C.D., McConaughy, B.L. and McCarthy, B.J. (1969) Nature 224, 149-154.
16. Sakano, H., Rogers, J.H., Hüppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. and Tonegawa, S. (1979a) Nature 277, 627-633.
17. Bradley, J.V. (1968) Distribution-Free Statistical Tests, pp. 304-310, New Jersey, Prentice Hall.
18. Kidd, S.J. and Glover, D.M. (1980) Cell 19, 103-119.
19. Cochet, M., Gannon, F., Hen, R., Maroteaux, L., Perrin, F. and Chambon, P. (1979) Nature 282, 567-574.
20. Leder, A., Miller, H.I., Hamer, D.H., Seidman, J.G., Norman, B., Sullivan, M. and Leder, P. (1978) Proc. Nat. Acad. Sci. U.S.A. 75, 6187-6191.
21. Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059.
22. Goldschmidt-Clermont, M. (1980) Nucl. Acids Res. 8, 235-252.
23. Miller, J.R., Cartwright, E.M., Brownlee, G.G., Fedoroff, N.V. and Brown, D.D (1978) Cell 13, 717-725.
24. Hardison, R.C., Butler, E.T. III., Lacy, E., Maniatis, T., Rosenthal, N. and Efstratiadis, A. (1979) Cell 18, 1285-1297.
25. Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1980) Cell 19, 959-972.
26. Nishioka, Y., Leder, A. and Leder, P. (1980) Proc. Nat. Acad. Sci. U.S.A. 77, 2806-2809.
27. Miyata, T., Yasunaga, T., Yamawaki-Kataoka, Y., Obata, M. and Honjo, T. (1980) Proc. Nat. Acad. Sci. U.S.A. 77, 2143-2147.
28. Yamawaki-Kataoka, Y., Sato, K., Shimizu, A., Kataoka, T., Mano, Y., Ono, M., Kawakami, M. and Honjo, T. (1979) Biochemistry 18, 490-494.
29. Van den Berg, J., Van Ooyen, A., Mantei, N., Schamböck, A., Grosveld, G., Flavell, R.A. and Weissman, C. (1978) Nature 276, 37-44.
30. Konkel, D.A., Maizel, J.V. Jr. and Leder, P. (1979) Cell 18, 865-873.
31. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979) Cell 18, 545-558.
32. Abney, E.R., Cooper, M.D., Kearney, J.F., Lawton, A.R. and Parkhouse, R.M.E. (1978) J. Immunol. 120, 2041-2049.
33. Oi, V.T., Bryan, V.M., Herzenberg, L.A. and Herzenberg, L.A. (1980) J. Exp. Med. 151, 1260-1274.
34. Parkhouse, R.M.E., Lifter, J. and Choi, Y.S. (1980) Nature 284, 280-281.
35. Early, P.W., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980b) Cell 20, 313-319.
36. Calame, K., Rogers, J., Early, P., Davis, M., Livant, D., Wall, R. and Hood, L. (1980) Nature 284, 452-455.
37. Brutlag, D., Fry, K., Nelson, T. and Hung, P. (1977) Cell 10, 509-519.
38. Wellauer, P.K., Dawid, I.B., Brown, D.D. and Reeder, R.H. (1976) J. Mol. Biol. 105, 461-486.

39. Fedoroff, N.V. and Brown, D.D. (1978) Cell 13, 701-716.
40. Shen, C-K.J. and Maniatis, T. (1980) Cell 19, 379-391.
41. Sures, I., Lowry, J. and Kedes, L.H. (1978) Cell 15, 1033-1044.
42. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978) Cell 14, 655-671.
43. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. and Tonegawa, S. (1980) Nature 286, 676-683.
44. Dunnick, W., Rabbitts, T.H. and Milstein, C. (1980) Nature 286, 669-675.
45. Early, P.W., Huang, H., Davis, M., Calame, K. and Hood, L. (1980a) Cell 19, 981-992.