Volume 10 Number 11 1982 Nucleic Acids Research

Simple DNA sequences in homologous flanking regions near immunoglobulin V_H genes: a role in gene interaction?

Justus B.Cohen, Kim Effron, Gideon Rechavi, Yinon Ben-Neriah, Rina Zakut and David Givol

Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel

Received 17 March 1982; Revised and Accepted 11 May 1982

ABSTRACT

Five closely related immunoglobulin V_H genes (subgroup II) were compared by sequencing of several kb of DNA. In three of the genes homology greater than 75% was found along an area of 4 kb that includes the coding region. The homology in flanking regions is only slightly lower than that in the coding sequences. Two other genes, which are located on the same EcoRI fragment, show high homology to the first three genes in the coding and immediately flanking regions. In more distant flanking regions no homology is found with the first three genes. This indicates that their evolutionary history differs from that of the other three genes. A region of simple DNA sequence composed of repetitive TCC and TCA elements was found at a distance of ∿380 bp upstream from the initiator ATG of these V_H genes. This region is the site where the two sets of genes abruptly start to diverge. The structure of the simple DNA sequence in the various V_H genes suggests that it may be involved in gene interaction. We propose that both simple DNA sequences and homology in flanking regions serve a function in the correction of V_H genes, which seem to be rather free to diverge and drift into pseudogenes. A correction mechanism may help this gene family to maintain its two major features, multiplicity and diversity.

INTRODUCTION

Both multiplicity and diversity of immunoglobulin variable regions are essential features for keeping the immune system in a state of maximum efficiency. It is well documented now that V region multiplicity is due in part to the presence of a large number of V genes in the germ-line (1,2). Diversity has been shown to be partly created by somatic processes such as combinatorial V-D-J joining (3), flexibility in the joining site (3-7), and somatic mutations (4,8-12). However, as we recently showed, the basic diversity of V genes is present in the germ-line and is generated throughout evolution (13). This diversity is mainly centered in the hypervariable regions (HV), which are essential for determining the antibody specificity, whereas the framework sequences are more conserved.

We and others have shown that V genes can accumulate mutations which

0305-1048/82/1011-3353$ 2.00/0

inevitably include termination codons, changes of residues that are critical for the protein structure, small deletions, etc. (10,11,13,14). This indicates that the need for V genes to diverge apparently permits detrimental mutations to occur even though this may lead to the inactivation of genes (11,13). Since selection pressure does not operate continuously on every V gene it is likely that a significant portion of the immune repertoire may drift to pseudogenes, thereby diminishing its multiplicity. This problem may be overcome if mechanisms for correction or use of these pseudogenes would exist, such as to prevent the family from losing its multiplicity and counterbalance the drift to pseudogenes (15).

Mainly two mechanisms have been proposed to function in gene correction (16-18): unequal crossing-over (17,19) and gene conversion (16,18,20-22). In both mechanisms homologous flanking regions would increase the chance of gene recombination (1,18,23-25) which eventually may lead to gene correction. In addition signals, such as simple repetitive DNA, may affect gene correction by functioning as hot spots for recombination (21,22,26).

In this paper we present the coding sequence of a mouse V_H gene and give a comparison of a major part of the flanking regions of five V_H genes. The flanking sequences of three of the five genes are homologous throughout nearly 4 kb of sequences that have been compared. The two other genes, however, show a completely unrelated sequence in their 5' flanking region, but are homologous to the first three genes throughout the coding region and up to a simple DNA sequence, which is present in different forms in analogous positions in all five genes. This simple sequence which is composed of repetitive elements TCC and TCA is present in the vicinity of the coding region, approximately 380 bp upstream from the initiator ATG. The main features of this simple sequence region are: a) It differs in length in various genes and shows a tendency towards deletion and/or duplication; b) It shows the repetitive elements either in a linear form, like $(TCC)_n$ $(TCA)_n$, or in a mixed arrangement, and c) It is the border of the region where all five genes have homologous sequences. These features are compatible with possible recombination events which started in this simple sequence region. We propose that this region may be involved in a mechanism of gene interaction in the V_H gene family, and may serve a role in the concerted evolution of this family. Gene conversion may be one of the consequences of this gene interaction.

MATERIALS AND METHODS

Materials

Restriction endonucleases and T_4 ligase were form New England Biolabs (Beverly, Massachusetts). E. coli polymerase I (n. Klenow)and calf intestine alkaline phosphatase were obtained from Boehringer (Mannheim, FRG). [α-^{32}P] deoxynucleoside triphosphates (2000-3000 Ci/mmole) were purchased from Amersham.

Isolation of recombinant clones

The Charon 4A recombinant bacteriophage library containing partially EcoRI digested BALB/c embryo DNA has been described (27). This library was screened with the nick-translated recombinant plasmid pV(11)[2] which contains MPC11 V_H cDNA (28). The screening was done according to the method of Benton and Davis (29).

Analysis of V_H containing Charon 4A clones

DNA from V_H containing recombinant clones was analysed by single and double digestion with the restriction endonucleases EcoRI, HindIII, BamHI, BglII, SacI, and KpnI. The digests were electrophoresed on 0.7% horizontal agarose gels. The gels were blotted onto nitrocellulose filter paper according to Southern (30). Gels were blotted in 10 X SSC for one hour to obtain a single blot, or for 20 minutes followed by 50 minutes to obtain two blots from the same gel. Filters were baked for 2 hours at 80°C. Prehybridization was for 2 hrs in hybridization buffer containing 6 X SSC, 5 X Denhardt solution, 0.5% SDS and 25 µg/ml denatured E. coli DNA. After prehybridization denatured ^{32}P-labelled probe was added to the buffer and hybridization was performed for 14 hrs at 65°C. After hybridization filters were washed three times at 65°C for 30 min in 3 X SSC, 0.1% $NaPP_i$ and 0.1% SDS. The dried filters were exposed to pre-flashed Curix RP-2 X-ray film with intensifying screens.

Subcloning of V_H containing restriction endonuclease fragments

Subcloning of V_H containing restriction endonuclease fragments was performed essentially as described by Bolivar et al. (31). V_H containing recombinant phages were digested with BamHI or BamHI and HindIII (Ch108), HindIII (Ch104), HindIII and EcoRI (Ch105) or EcoRI (Ch111). The digests were analysed by agarose gel electrophoresis. pBR322 was digested with the same one or two restriction endonucleases and treated with alkaline phosphatase. Recombinant phage fragments and cleaved plasmid were ligated with T_4 DNA ligase and the ligated mixture was used to transform $CaCl_2$-treated E.coli strain HB101. Bacterial clones containing V_H genes were identified using the

method of Grunstein and Hogness (32). The V_H containing PstI fragment of MPC11 cDNA plasmid pV(11)[2] was used as a probe (28).

Analysis of V_H containing subclones

V_H containing subclones were digested by restriction endonucleases which do not cut or cut infrequently in pBR322 and subjected to agarose gel electrophoresis, gel blotting and hybridization as described for the analysis of Charon 4A recombinant clones. Fine mapping was performed on end-labelled fragments using the procedure of Smith and Birnstiel (33). The fragment of interest was labelled at one end as described below. 10-25 ng (∿ 8000 cpm) was partially digested at 25°C with 0.1 unit of different enzymes in a reaction mixture of 25 µl. At several time points 5 µl aliquots were taken from the mixture and immediately pooled in a tube containing 5 µl 0.2M EDTA on ice. At the end of the reaction the pooled digest was heated to 65°C for 5 min, cooled on ice for 5 min and applied to a 5% acrylamide gel. After the run, the gel was covered with Saran Wrap and exposed to X-ray film.

End-labelling of restriction fragments

1-5 µg of DNA from a pBR subclone or an isolated fragment was digested with the appropriate enzyme. End-labelling of 5' protruding ends was performed using one or two [α-^{32}P] deoxynucleoside triphosphates (20 µCi each) and 5 units of reverse transcriptase in a reaction mixture of 20 µl. After 30 min incubation at 37°C, the reaction was stopped by heating to 70°C. If necessary, this was followed by digestion with a second enzyme. End-labelling of 3' protruding (SacI) or blunt ends (HaeIII) was done in a 20 µl reaction mixture containing 20 µCi [α-^{32}P]dGTP and the Klenow fragment of E. coli DNA polymerase I. Incubation was for 15 min at 37°.

DNA sequencing

DNA sequencing was performed by the method of Maxam and Gilbert (34), as modified by Smith and Calvo (35). The reaction products were loaded on either short (0.3 mm thick x 20 cm wide x 40 cm long) or long (0.3 mm thick x 40 cm wide x 80 cm long) 6% urea-acrylamide gels (36). Usually three loadings of sequenced fragments were applied on the same gel. DNA sequences from the short gels could be accurately read up to 200 bp, while the sequences from the large gels could be read up to 500 bp.

RESULTS

A simple DNA sequence is present 5' to five V_H genes

We isolated several cross-hybridizing clones from a BALB/c mouse embryo DNA library by screening with the plasmid pV(11)[2], which contains V_H cDNA

derived from myeloma MPC11 (28). Fragments that contain the V_H coding region were subcloned in pBR322. Previously we reported the V_H coding sequences of four of these clones (13). In this paper the sequence of large parts of flanking DNA from these clones, and the nearly complete sequence of a fifth clone, pCh105, including its V_H coding region, are given.

The sequences were determined according to the strategy shown in Fig. 1. Homologous portions of the sequences of these five clones were compared. As indicated in Fig. 2, we compare the sequences in three regions. The first includes the coding region, and ∿ 950 bp 5' to codon nr. 1 and ∿ 570 bp 3' to the last codon of V_H. The major conclusions from this comparison of approximately 1800 bp of sequence from five different V_H genes are as follows (Fig.3): Within this subgroup the size and position of the various segments, leader (15 codons), intron (81-83 bp), V_H segment (102 codons) and recombination signals are very similar in all genes. As we showed before, variability in the coding

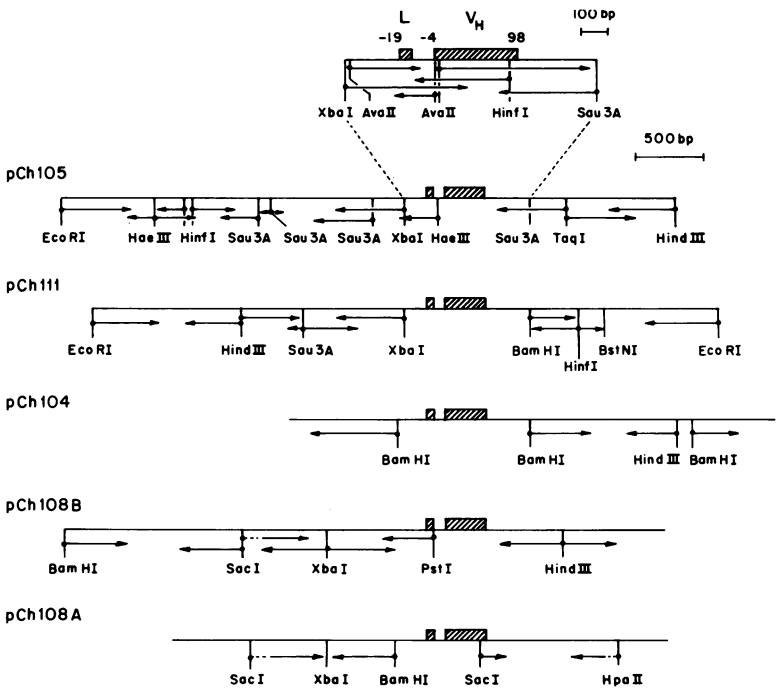

Fig. 1. Strategy used to determine the nucleotide sequence of the V_H gene of the subclone pCh105, and of a portion of the flanking sequences of pCh105, pCh111, pCh104, pCh108B, and pCh108A. Hatched areas show the coding regions of the leader (L) and V_H.

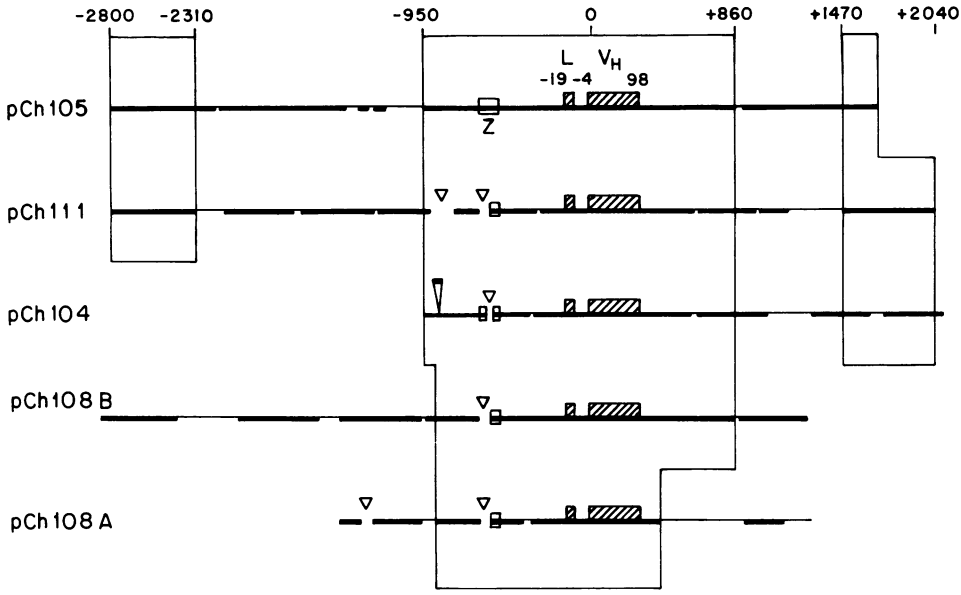

Fig. 2. Schematic representation of the extent of sequence determination in five V_H containing subclones. Hatched boxes show the coding regions. Dark horizontal bars indicate the sequenced regions (see Fig. 1). Open space and ∇ within the horizontal lines indicate deletions, as compared with pCh105. The large open boxes show the location of the sequences given in Figs. 3, 5 and 6: central box, Fig. 3; left box, Fig. 5; right box, Fig. 6. 0, codon Nr 1 of V_H. Z, sequence referred to as zipper (see text).

regions is most pronounced in the hypervariable regions (13). The DNA sequences of all five genes upstream from the initiation ATG are highly (75-85%) homologous, and do not show significant deviations from one another. Starting at position -524 all genes show a region of simple repetitive DNA sequence composed of TCC and TCA triplets (Fig. 3, italics). This sequence is most extensive in pCh105, where it consists of more than 100 bp. In the other clones this region shows a similar structure, but parts of it have been deleted. In comparison with pCh105, pCh104 has lost more than half of the $(TCA)_n$ stretch; in pCh108B and pCh108A the complete $(TCC)_n$ sequence, and part of the $(TCA)_n$ stretch are deleted, while pCh111 shows a mosaic-like arrangement of the TCC and TCA elements with loss of more than half of the sequence. An autoradiograph, showing the sequence of three clones in this region, is given in Fig. 4.

The basic features of this region: the mixed versus linear arrangement of the repetitive trinucleotide elements, and the size heterogeneity indicate

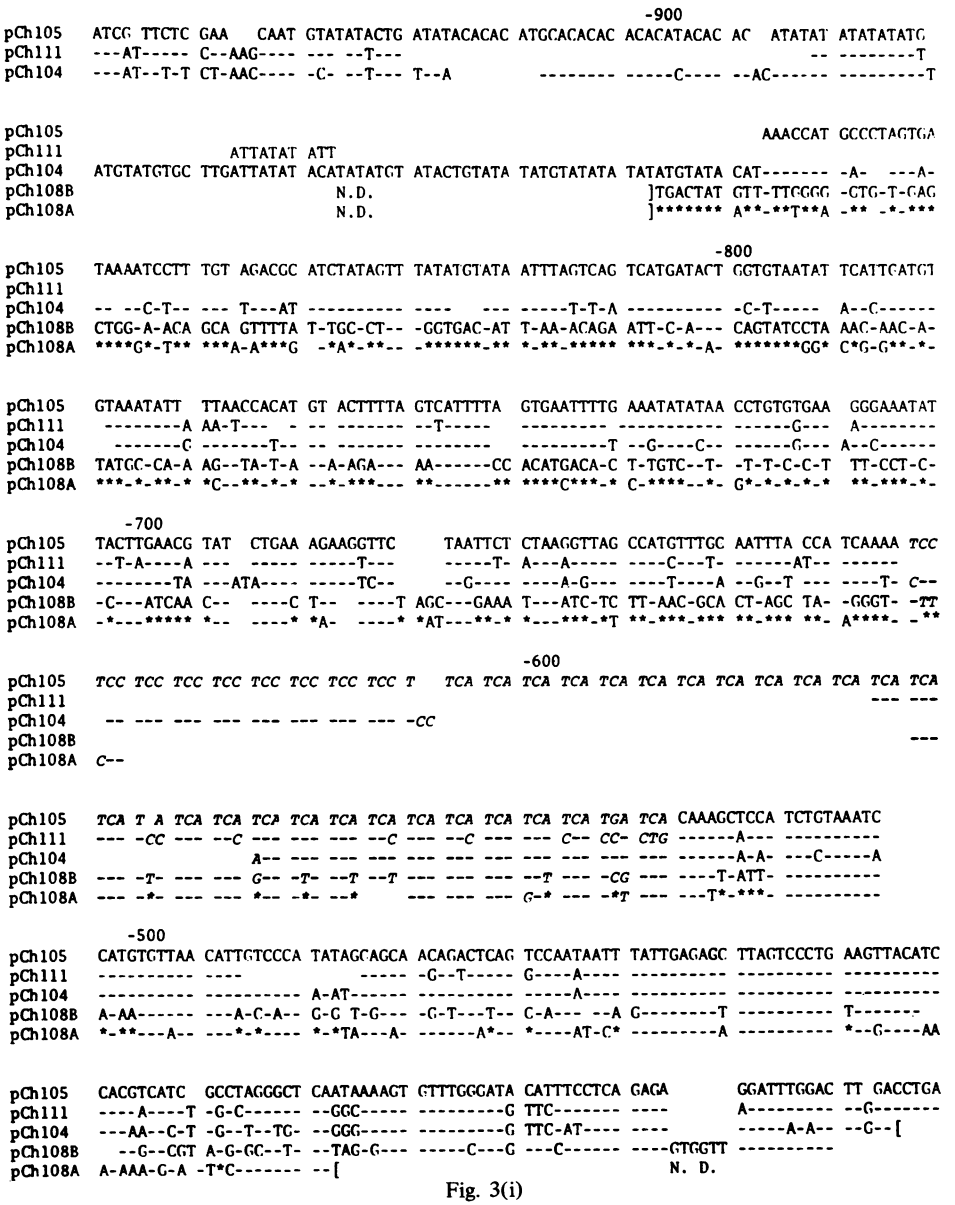

Fig. 3(i)

that this region may have been involved in unequal DNA exchanges. For reasons discussed by Kedes (26) for an analogous case, we will refer to this region of simple DNA sequence as "zipper" or "Z".

Finally, as shown in Fig. 3, it is remarkable that 5' to the Z region

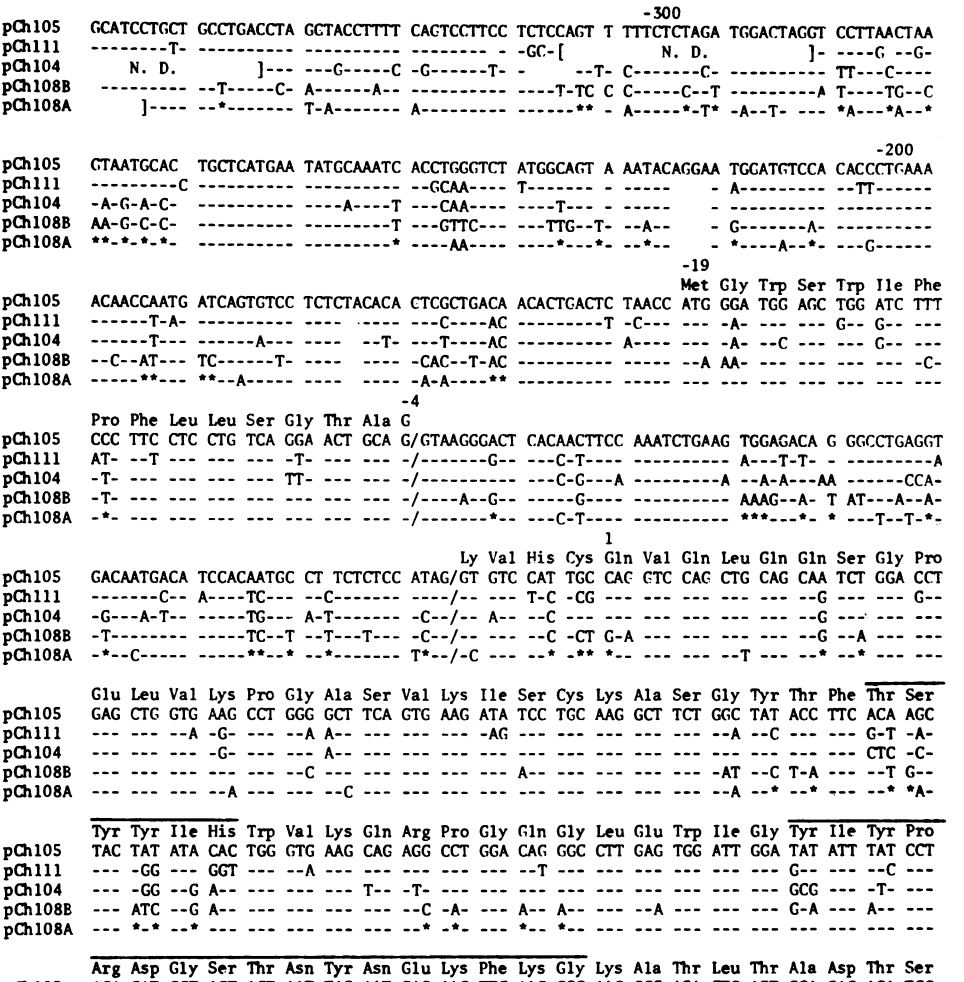

Fig. 3(ii)

the sequences of two of the clones, pCh108B and pCh108A, completely lost their homology to the other three genes. Beyond Z, they remain however homologous to each other, as the other three clones also keep close homology to

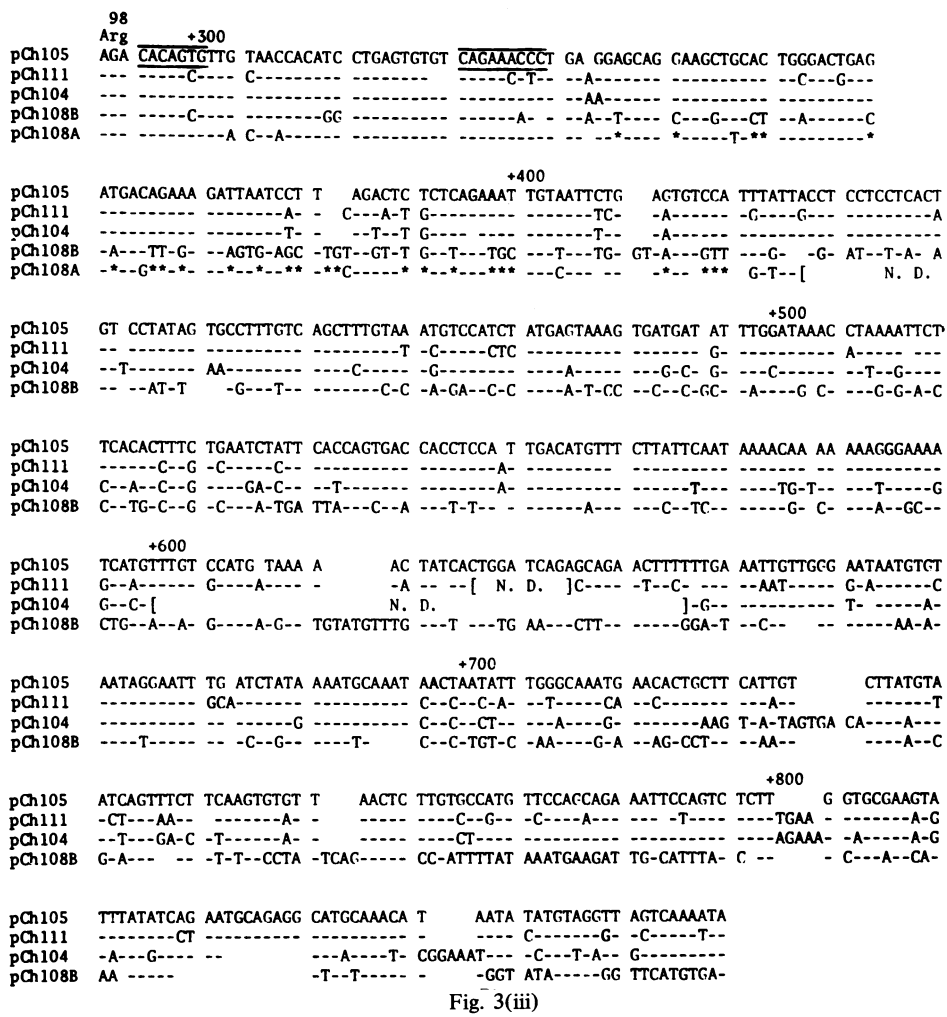

Fig. 3(iii)

Fig. 3. Nucleotide sequence of V_H coding and immediately flanking regions from five V_H genes. Amino acid residues predicted by the nucleotide sequence of pCh105 are given above the line of the pCh105 codons. Hypervariable regions 1 and 2 are overlined. Recognition signals for V_H-D joining are boxed. The simple DNA sequence (Z) in the 5' flanking regions of these genes is given in italics. Dashes indicate identity to pCh105. Stars show where pCh108A is identical to pCh108B but different from pCh105. Open spaces indicate deletions. Regions between parentheses were not determined (N.D.)

each other in this region.

Homology extends over four kb of V_H containing DNA

Because the immediately flanking regions of five V_H genes are homologous,

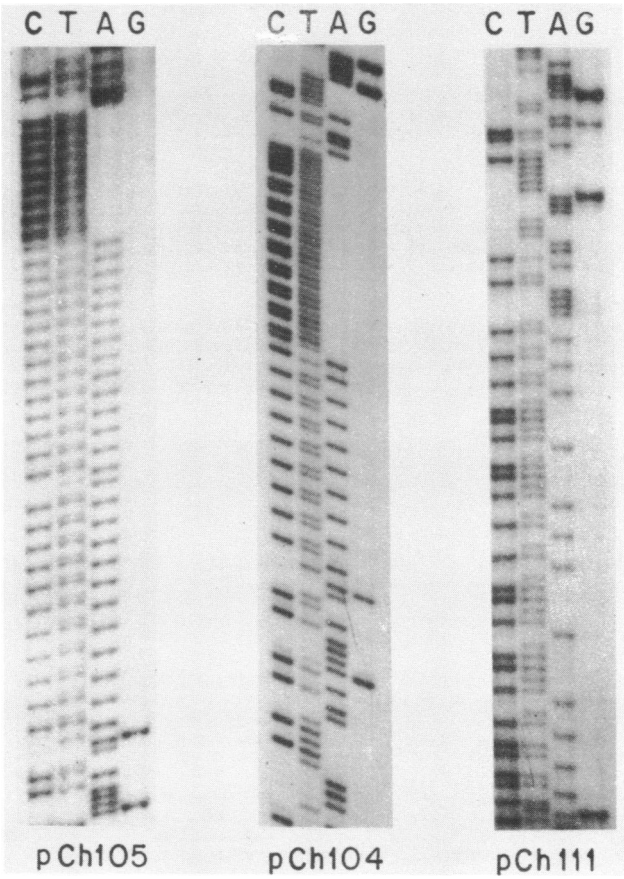

Fig. 4. Autoradiographs of sequencing gels showing the region of simple DNA sequence in the clones pCh105, pCh104, and pCh111. The sequences were determined by the procedure of Maxam and Gilbert (34) after 3' end-labelling of the XbaI site which is located approximately 230 bp downstream from the simple sequence in pCh105 and pCh111, or after 3' end-labelling of a BamHI site, which is located approximately 175 bp downstream from the simple sequence in pCh104.

as shown above, we were interested to determine the extent of homology at regions further away from the genes. For this purpose we determined in three of the clones (pCh105, pCh111 and pCh108B) the DNA sequence of several areas at the far 5' flanking regions according to the strategy shown in Fig. 1 and the map of Fig. 2. This sequence information was used to compare homology at various distances from the coding region. As an illustration, Fig. 5 gives the sequence of $\sim$ 500 bp from a region located between 2.3 and 2.8 kb 5' to

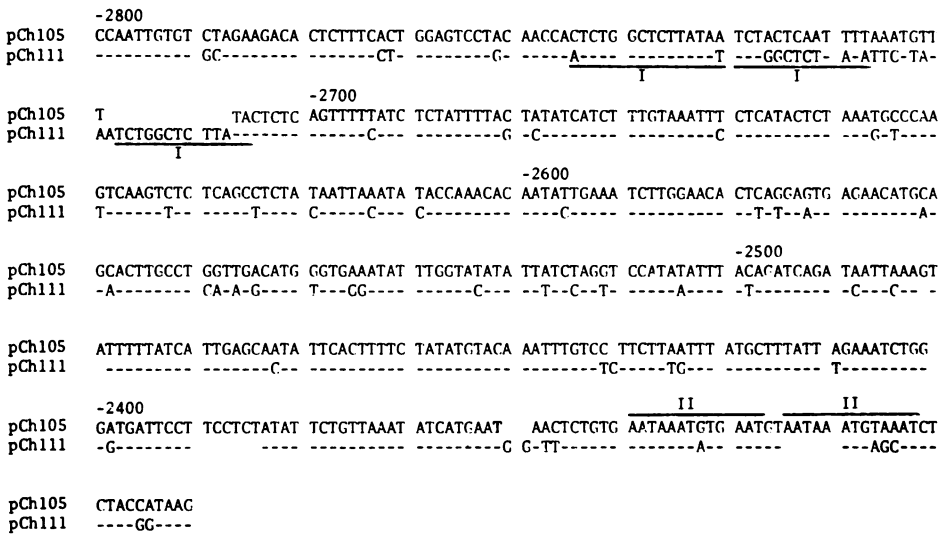

Fig. 5. Nucleotide sequence comparison of pCh105 and pCh111 in a region between ∿ 2300 and 2800 bp 5' to the first codon of V_H (see Fig. 2). Symbols as in Fig. 3.

the first V_H codon (see also Fig. 2). The homology in this region is ∿ 85%, if the insertions/deletions introduced to maximise homology are not considered. The sequence of pCh108B is not shown in this figure, as no homology with this region of pCh105 or pCh111 could be detected in the pCh108B sequence.

In a similar way sequences of three clones, pCh105, pCh111 and pCh104, from a region approximately 1500 bp 3' to the last codon of V_H, were compared (Fig. 6). Overall homology in this region is ∿ 90% (105 vs. 111), 85% (105 vs 104) and 82% (111 vs 104). These data, and more sequence information from regions between those given in Figs. 3,5 and 6, were used to construct the homology plots of Fig. 7. The location of the sequences used for this comparison is indicated by heavy horizontal bars in Fig. 2. The sequences were compared over consecutive stretches of 50 bp in the flanking regions and 25 bp in the coding regions, and the calculated homology is given as % difference in Fig. 7. As shown in Fig. 7 *top*, the flanking regions of pCh105 and pCh111 are homologous throughout the entire 3.5 kb for which sequencing data were obtained. It can be seen that the overall homology between the flanking sequences of these two genes is only slightly lower than that between their coding regions (see also Table I). This may indicate that selective pressure operates on the entire region to a similar extent. The more limited compa-

```
                                                +1500
pCh105   GTCCTCTACT CTATGTGTGT TTGGGACCTC ATATCAGCTG GTGTATGCTG CCTGTTTTAT GGTCCAGTGT TTGAGAGATC
pCh111   -C--A-  -- G--------  -----G---- --- ----G- --------C- T----- --- ---A-G---  --A-------
pCh104   ---------- G-------C- GC-------- ---------A ---------- T---A--GG- T-----C--- --------G-

                                                                +1600
pCh105   TCAGGGGTCC AGGAGTGGTT GAGACTGCTG GTCCTCCTAC AGGATTGCCC TTCTCCTCAG CTTCTATCAT CCATCACTAA
pCh111   A- ------- --      -- ---------- ---------- -----A---- --T------- ---------G ----C----G
pCh104   -AG------- -A--T-AA-C ----TG---- ---------- -------ATT ---------- -----T---- -TT-- ---T

                                                                        +1700
pCh105   TTCAACAACA GGAGTCAGCT GCTTCTGTCC ATTGGTTGGG TGCAAATACC TGCCT[           N.D.
pCh111   ---------- ---T------ ---------T -------T-- --------T- -A---GTGAA TCTTTCAGAT GCTTGTTGGA
pCh104   ---    -G- --  -T---- -----A---- C------[            N.D.

pCh105                                        N.D.
pCh111   TCTTTCAGAG GTCAGTCATG ATAGGTCCCT TTTTGTGAAT GCTCCATAGC CTCAGTAACA GTGGTAGGTC ATGGCATCTC
pCh104                                        N.D.

                +1800
pCh111   CACTTGAGCT TGATCCCACT TTGGGCCTGT TTCTGGACCT TCTTTTCCTC AGGCTCCTCT CCATTTGCAT CCTTGTAATA
pCh104              N.D.                 ]----T- ---------- ------T--- ------C--- T-C------T

                                        +1900
pCh111   ATTTCAAACA GGAATAATTA TGGGTCAGAT TTGTGACTGT GAGATGGCAA CCCCATCTGT CACTTGATTT CCTGTCTTCC
pCh104   CA----G--- -T--C--A-- -T------TG ---------- ---------- -------AC- --------GC ----------

                                                        +2000
pCh111   TGCTGGAGGT GGGCTCTATA AGGTCCCTCT CCCTACTGTC ACGCATTTCA TTTAAGGTCC CTCTCCTTGA GACCTGAGAG
pCh104   ---------- A--------- --T---T--- ---------- ---------- ---------- ---C-T---- -T-T-----AA

pCh111   TCTTTAACA
pCh104   ---C-C---
```

Fig. 6. Nucleotide sequence comparison of pCh105, pCh111 and pCh104 in a region between ∿ 1470 and ∿ 2040 bp 3' to the first codon of V_H (see Fig. 2). Symbols as in Fig. 3.

rison of pCh105 and pCh111 to pCh104 shows the same tendency (Fig. 7 *middle* and Table I). The comparison of pCh105 and pCh108B however shows a completely different pattern (Fig. 7 *bottom* and Table I). The coding region of pCh108B is as closely related to pCh105 as is that of pCh111. In the 3' flanking region both sequences start to diverge extensively after ∿ 70 bp following the last codon of V_H. As mentioned before, this effect is more striking when the 5' flanking sequences of both genes are compared. The sequences can be aligned for the approximately 400 bp between Z, and the code of the signal peptide. Starting directly upstream from the Z region however, the sequences are completely unrelated over a stretch of at least 2 kilobase pairs. A computer analysis of these sequences did not detect any significant homology in this region.

Deletions occur within repetitive DNA

Our comparison of homologous sequences of five V_H genes confirms the conclusions of Efstratiadis et al. (37) about the occurrence of small dele-

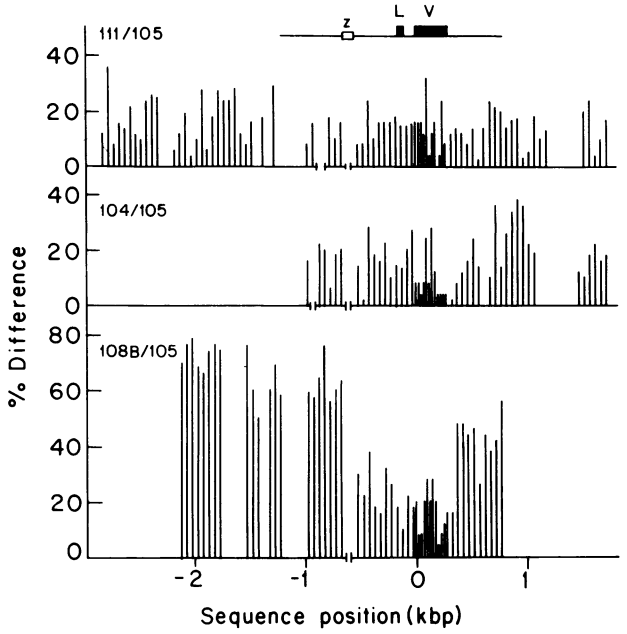

Fig. 7. Comparison of the nucleotide sequences of V_H coding and flanking regions. Each bar in the flanking sequences represents the percent nucleotide differences between two of the clones as calculated for consecutive stretches of 50 bp, while the percent difference in the coding regions was calculated for stretches of 25 bp. In the ordinate 0 represents codon 1 of V_H. Regions for which no bars are given, were not compared. Gaps in the horizontal axis indicate deletions in one of the sequences. No homology comparison is given for the region of simple DNA sequence (see Fig. 4). Clones compared are: *top*, pCh105 vs pCh111; *middle*, pCh105 vs pCh104; *bottom*, pCh 105 vs pCh108B.

tions/insertions in regions of repetitive DNA. An example of this occurrence can be found in Fig. 3 between positions -930 and -780. The pCh104 sequence in this region shows the presence of repetitive TATA, CACA and ATGT elements. Deletions of various sizes occur in this region in the clones pCh105 and pCh111. Two other examples are indicated in Fig. 5: the 13 bp long underlined sequence called I is repeated three times in pCh111 in a region where a 12 bp deletion is found in pCh105. The 13 bp long sequence called II in Fig. 5 is present twice in pCh105, and overlaps a 5 bp deletion in pCh111, where this sequence occurs only once. This phenomenon may be due to base mispairing and slippage during replication, as proposed by Efstratiadis et al. (37), and could also be an explanation for the size heterogeneity of the $(TCA)_n$ stretch in the different Z sequences.

Table 1. % Substitution in coding and flanking regions of V_H genes. The table compares the homology between various segments of the five genes analysed. For each pair of columns the first gives the location of the segment (see Fig. 2) and the second column shows the size of the segment compared in bp. Z is the simple sequence region. ATG is the first codon in V_H.

Genes compared	5'→Z %	size bp	Z→ATG %	size bp	V_H %	size bp	3' %	size bp
pCh105/pCh104	18.0	310	16.7	378	12.5	433	19.3	1030
pCh105/pCh111	15.5	1670	13.9	368	13.6	433	14.2	1023
pCh105/pCh108B	64.2	1040	24.7	368	16.6	433	39.6	470
pCh111/pCh104	21.6	180	13.6	368	16.9	433	24.1	878
pCh108B/pCh108A	13.8	780	18.4	320	9.9	433	15.1	372

DISCUSSION

Hypervariable regions are present in the germ-line

In a recent paper we presented the coding sequences of four embryonic V_H genes, which belong to mouse V_H subgroup II (13). In this paper we describe the flanking sequences of these genes, and add the flanking and coding sequence of a fifth cross-hybridizing gene, pCh105. Comparison of the coding sequence of pCh105 to the other four confirms our earlier conclusions, that closely related germline V_H genes mainly differ in their hypervariable regions but that differences are found over the entire length of the genes (Fig. 3 and ref. 13). The two genes, which are located on the single EcoRI fragment of Ch108 show close homology. pCh105 does not show such a close relation to any of the other four genes, except in its leader region, which differs in only one position from the pCh108A leader. This high homology (98%) between the two leader sequences is remarkable, since it does not extend beyond the leader code.

Homologous flanking regions may be targets for recombination

Comparison of the flanking sequences of pCh105 and pCh111 (Figs. 3,5,6,7 *top*) shows that homology is present in the entire stretch which we compared ∿ 3.5 kb). There seems to be no correlation between the extent of divergence and the distance from the coding regions. It is of interest that the overall homology between the coding regions is only slightly higher than between the flanking regions (Fig. 7 *top* and Table I). This suggests that these genes evolved by duplication of a common ancestor, and that the duplication unit is

much larger than the coding region itself. The extent of divergence (∿ 18%) indicates that these genes separated a considerable time ago. Hence, the similar extent of homology in flanking and coding regions suggests that some forces operate on the flanking sequences to keep their homology.

Seidman et al. (1) have proposed that flanking regions may play a crucial role in determining the frequency of unequal crossing-over events between genes. Long stretches of homologous DNA, flanking two homologous genes, would considerably increase the chance for recombination. This hypothesis has been extended (18,23-25) to apply also to gene conversion, and implies that homology in flanking regions may be a biologically important tool for certain classes of genes to enable them to undergo frequent recombination events. It is possible that the strong sequence homology around V_H genes of the same subgroup is used to maintain the possibility for interaction between V_H genes which could result in gene correction.

Simple sequences: a role in gene interaction?

Regions of simple DNA sequence in the vicinity of structural genes have been found in other gene families as well. For example, in the human β-like globin genes, this region is located in the second intervening sequence (21). Convincing evidence has been presented that this region serves an important function in the early steps of homologous recombination between two related but non-identical genes (21). In the case of histone genes, the simple sequence is linked to one of the five structural genes of the histone gene repeat (38,39). Kedes has argued that simple DNA sequences, present in analogous positions in the vicinity of homologous genes, will reanneal much faster than any of the homologous but unique DNA surrounding it and may act like a zipper (26). We would like to follow this suggestion and call the simple DNA sequence which is linked to our V_H genes, a "zipper" sequence. It is of interest that a stretch of simple DNA sequence consisting of 31 CA doublets has been found at a distance of about 300 bases at the 3' side of an immunoglobulin V_K gene (40) and that the intron of four T15 V_H genes contains a repetitive GT sequence (12).

The simple DNA sequence ∿ 380 bp upstream from the initiator codon of V_H has the following features: it is composed of repeating trinucleotides TCC and TCA which gives it an asymmetrical structure, it shows deletion or self-duplication (compare pCh105 to pCh104 in Fig. 3), and in one of the genes it shows a mosaic arrangement of the two elements (see pCh111 in Fig. 3). We believe that the mixed arrangement of the basic elements of the zipper in pCh111 versus their linear arrangement in pCh104 and pCh105, and the size

heterogeneity indicate that this region in pCh111 in fact has played a role in unequal recombination events. A kinetically fast reannealing between two zipper regions, like those of pCh105 and pCh104, may be the first step in such a process (21,39). The mixed arrangement of Z in pCh111 may be explained by mismatch repair if such a mechanism could operate on closely spaced targets on both strands of the heteroduplex simultaneously, as proposed by Seidman et al. (1).

Two sets of V_H genes within the V_HII subgroup

The two Ch108 genes show a complete absence of homology to the three other genes in the region 5' to the Z sequence. They do however retain close homology to one another. Hence the event, responsible for the different sequence of the 5' flanking regions of pCh108A and B probably took place before they were duplicated from an ancestral gene. This places these genes on a different branch of the evolutionary tree of this subgroup as compared to the three other genes, and suggests a subdivision of the five V_H genes into two distinct sets within a single subgroup. The lack of homology between the two sets of genes in the region 5' to the zipper implicates that the target for recombination between members of the different sets is much smaller than that for recombination between members of the same set. This suggests (41) that the Ch108 genes will less easily recombine with the Ch105-like genes (pC105, 104 and 111). In this context it may be significant that the zippers of the Ch108 genes are mutated to an extent that their fast annealing to the Ch105-like zippers is at least doubtful. However, since the Ch108 genes are homologous to each other in their 5' regions, and have very similar zippers, they may recombine with each other and with other Ch108-like genes which may be present in the germline.

One of the conclusions from this work is that strong homology (75% or higher) in the flanking regions of immunoglobulin V_H genes is being kept within a set of closely related genes throughout a long distance (several kb). This suggests that some forces other than those operating on the expressed proteins must operate to keep this homology. The preservation of homology in flanking sequences may lead to selective advantage if gene interaction is important for maintaining this gene family in evolution. We speculate that this interaction starts with the alignment of stretches of simple DNA sequence followed by homologous recombination between two genes, which may lead to gene conversion. On the basis of flanking region homology, we distinguish two sets of genes, the Ch105-like and the Ch108-like genes, within the V_HII subgroup. Genes within a set may recombine with one another, but it is

likely that interaction between the sets is infrequent. Members of both sets are preceded by simple sequences of different structure which may indicate their different recombination ability.

ACKNOWLEDGEMENTS

We thank Mrs. S. Hazum for her excellent assistance. This work was supported in part by a grant from the National Council for Research and Development, Israel and the DKFZ, Heidelberg, FRG.

REFERENCES

1. Seidman, J.G., Leder, A., Nau, M., Norman, B. and Leder, P. (1978) Science 202, 11-17.
2. Rabbitts, T.H., Matthyssens, G. and Hamlyn, P.H. (1980) Nature 284, 238-243.
3. Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980) Cell 19, 981-992.
4. Max, E.E., Seidman, J.G. and Leder, P. (1979) Proc. Natl. Acad. Sci. USA 76, 3450-3454.
5. Sakano, H., Hüppi, K., Heinrich, G. and Tonegawa, S. (1979) Nature 280, 288-294.
6. Weigert, M., Perry, R., Kelley, D., Hunkapiller, T., Schilling, J. and Hood, L. (1980) Nature 283, 497-499.
7. Gough, N.M. and Bernard, O. (1981) Proc. Natl. Acad. Sci. USA 78, 509-513.
8. Bernard, O., Hozumi, N. and Tonegawa, S. (1978) Cell 15, 1133-1144.
9. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. and Tonegawa, S. (1980) Nature 286, 676-683.
10. Crews, S., Griffin, J., Huang, H., Calame, K. and Hood, L. (1981) Cell 25, 59-66.
11. Bothwell, A.L.M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. and Baltimore, D. (1981) Cell 24, 625-637.
12. Kim, S., Davis, M., Sinn, E., Patten, P. and Hood, L. (1981) Cell 27, 573-581.
13. Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D. and Cohen, J.B. (1981) Nature 292, 426-430.
14. Bentley, D.L. and Rabbitts, T.H. (1980) Nature 288, 730-733.
15. Wellauer, P.K., Dawid, I.B., Brown, D.D. and Reeder, R.H. (1976) J. Mol. Biol. 105, 461-486.
16. Edelman, G.M. (1970) In: The Neurosciences, p. 885-900. Schmitt, F.D. ed., Rockefeller Univ. Press, N.Y.
17. Edelman, G.M. and Gally, J.A. (1970) In: The Neurosciences, p. 962-972. Schmitt, F.D. ed., Rockefeller Univ. Press, N.Y.
18. Lauer, J., Shen, C.-K.J. and Maniatis, T. (1980). Cell 20, 119-130.
19. Hood, L., Campbell, J.M. and Elgin, S.C.R. (1975) Ann. Rev. Genet. 9, 305-353.
20. Snyder, L.R.G. (1980) Evolution 34, 1077-1098.
21. Slightom, J.L., Blechl, A.E. and Smithies, O. (1980) Cell 21, 627-638.
22. Baltimore, D. (1981) Cell 24, 592-594.
23. Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W. and Wilson, A.C. (1980) Proc. Natl. Acad. Sci. USA 77, 2158-2162.
24. Proudfoot, N.J. and Maniatis, T. (1980) Cell 21, 537-544.
25. Liebhaber, S.A., Goossens, M. and Kan, Y.W. (1981) Nature 290, 26-29.

26. Kedes, L.H. (1979) Ann. Rev. Biochem. 48, 837-870.
27. Zakut, R., Givol, D. and Mory, Y.Y. (1980) Nucl. Acids Res. 8, 453-466.
28. Zakut, R., Cohen, J.B. and Givol, D. (1980) Nucl. Acids Res. 8, 3591-3601 and 8, 4841.
29. Benton, W.D. and Davis, R.W. (1977) Science 196, 180-182.
30. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
31. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L. Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) Gene 2, 95-113.
32. Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. USA 72, 3961-3965.
33. Smith, H.O. and Birnstiel, M.L. (1976) Nucl. Acids Res. 3, 2387-2398.
34. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
35. Smith, D.R. and Calvo, J.M. (1980) Nucl. Acids Res. 8, 2255-2274.
36. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. 87, 107-110.
37. Efstratiadis, A. et al. (1980) Cell 21, 653-668.
38. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978) Cell 14, 655-671.
39. Sures, I., Lowry, J. and Kedes, L.H. (1978) Cell 15, 1033-1044.
40. Nishioka, Y. and Leder, P. (1980) J. Biol. Chem. 255, 3691-3694.
41. Sodergren, E.J. and Fox, M.S. (1979) J. Mol. Biol. 130, 357-377.