Volume 10 Number 23 1982 Nucleic Acids Research

# An immunoglobulin promoter region is unaltered by DNA rearrangement and somatic mutation during B-cell development

Catherine Clarke, James Berenson, Joan Goverman, Paul David Boyer, Stephen Crews*, Gerald Siu* and Kathryn Calame

Department of Biological Chemistry and the Molecular Biology Institute, University of California, Los Angeles, CA 90035 and *Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Received 27 August 1982; Revised and Accepted 19 October 1982

ABSTRACT

The V1 gene encodes the heavy chain variable region of antibodies that bind to phosphorylcholine in the Balb/c mouse. V1 genes have been cloned from mouse sperm DNA, an IgM-producing tumor HPCM2 and an IgA-producing tumor M167. The transcription start site of the V1 gene has been mapped 63± 1 base pairs from the coding sequence for both α and μ transcripts. Comparison of flanking DNA sequence 574 base pairs 5' to the V1 transcription start site in sperm, HPCM2 and M167 DNA reveals that sperm and HPCM2 sequences are completely identical in this region and the M167 sequence differs from them by a single base change. Although the coding region of the V1 gene has undergone a high (4%) rate of somatic mutation in M167 we demonstrate that the somatic mutation mechanism stops near the transcription start site. These results demonstrate that initiation of V1 gene transcription remains unchanged with respect to location and 5' sequences throughout B-cell development.

## INTRODUCTION

Immunoglobulin variable (V) region genes provide a unique opportunity to study differential gene expression in a complex multigene family. Furthermore, these genes may provide an opportunity to understand novel regulatory mechanisms which have not been described for other eukaryotic genes since immunoglobulin genes are known to display several unique properties. During the development of antibody-producing B cells, V genes are subjected to two different types of modification at the DNA level: 1) gene rearrangement and 2) somatic hypermutation (for reviews see 1-5). Recent studies have shown that both rearrangement and hypermutation are involved in generating the immense diversity displayed by the immune system, but the role either may play in gene regulation remains obscure. Our current studies are focused on transcriptional regulation of immunoglobulin genes. As a first step toward understanding V gene transcriptional regulation, therefore, we have assessed the effects of gene rearrangement and hypermutation on the promoter region of a heavy chain V gene.





0305-1048/82/1023-7731 $2.00/0

Two types of immunoglobulin gene rearrangement occur during B-cell development-- V(D)J joining and $C_H$ switching. Both rearrangements involve deletion of DNA 3' to the expressed V gene but do not cause detectable changes in the DNA 5' to the expressed V gene. However, V(D)J joining appears to be a prerequisite for transcription from V gene promoters. A recent study by Mather and Perry (6) established that the transcriptional activity of unrearranged $V_k$ genes in myeloma cells is at least 16,000 fold lower than transcription of joined $V_k$ genes. The mechanism for this effect is not understood. Furthermore, it is not known if $C_H$ switching affects DNA sequences which regulate $V_H$ gene transcription.

Modification of V gene DNA by somatic mutation has been demonstrated for both light chains (7-10) and heavy chains (11). Recent studies suggest somatic mutation of V genes occurs by an undefined hypermutational mechanism late in B-cell development and is correlated with class switching (11). Variant genes contain altered bases scattered throughout the variable region coding sequences (12) but it is not known if these somatic alterations extend into the 5' region involved in initiation of V gene transcription.

Another intriguing aspect of immunoglobulin (Ig) gene regulation is the extremely high level of antibody protein and mRNA which are present in terminally differentiated plasma cells, in the range of 3-4 x $10^4$ copies of mRNA per cell (13). It has been shown that this high level of mRNA results from a combination of high transcription rate, rapid processing and high messenger stability (13). However, at earlier stages of B-cell development, $I_g$ protein and mRNA are present in much lower amounts, probably reflecting a lowered transcription rate of the genes as well as regulation at other levels (14-16).

We have chosen to focus our study on the transcriptional regulation of the V1 heavy chain gene which encodes a major portion of the immune response to phosphorylcholine in the Balb/c mouse (11). This is a particularly good $V_H$ gene for transcription studies because extensive information is available on regulation of the phosphorylcholine response (17), on protein sequences of heavy chains encoded by the V1 gene (18,8) and on DNA sequences of the germline V1 gene, three closely homologous germline genes (11) and somatic variants of the V1 gene (12). In this paper we have investigated the affects of VDJ joining, $C_H$ switching and somatic mutation on the promoter region of the V1 gene. We have determined the exact site of transcription initiation for the V1 gene and have analyzed

the DNA sequence in a region more than 500 base pairs (bp) 5' to this site. We demonstrate that following both types of DNA rearrangement--VDJ joining and $C_H$ switching--neither the location of the transcription initiation site nor the DNA sequence of the 5' region is altered. Finally, we show that a region of 574 bp 5' to the Vl gene transcription initiation site is virtually unaffected by somatic mutation in a tumor line which shows a high degree of somatic mutation in the Vl coding sequence.

## METHODS

Isolation of RNA. Nuclear and cytoplasmic RNA samples for blotting studies and for S1 nuclease studies were prepared from 2-4 x $10^8$ HPCM2 cells grown in suspension. Cells were washed with cold phosphate buffered saline and lysed with 0.1% NP40. Nuclei were isolated and detergent washed as described by Federoff et al., (19) and Penman (20). Cytoplasmic and nuclear fractions were phenol extracted and ethanol precipitated. Nuclear RNA was treated for 15 minutes at 37°C with 10 ug/ml RNase-free DNase, extracted and ethanol precipitated. All buffers up to this stage in the preparation contained 10mM vanadyl ribonucleoside complex to inhibit endogenous RNases (21). Both cytoplasmic and nuclear RNA samples were subjected to chromatography on oligo dT cellulose (Collaborative Research).

For dideoxy sequencing, HPCM2 and M167 heavy chain RNA was prepared from solid tumor tissue. Tumors were pulverized in liquid nitrogen and total RNA was isolated by the guanidium thiocyanate method (22). Poly (A+) RNA was selected by 2-3 passes over an oligo dT cellulose column prior to separation of heavy chain RNA on isokinetic sucrose gradients.

Isolation of DNA Fragments. Restriction fragments used for S1 nuclease studies, as primer in dideoxy sequencing studies and for sequencing were isolated from polyacrylamide gels by the method of Maxam and Gilbert (23) or from agarose gels by electro elution onto DE81 paper (24).

RNA Blots. RNA samples (1-15 ug) were suspended in a solution containing 50% (v/v) recrystallized deionized formamide, 2.2M formaldehyde and 1X MOPS buffer, heated at 57°C for 3 minutes and loaded immediately on 1% agarose gels containing 2.2 M formaldehyde in 1X MOPS buffer. (1X MOPS buffer is 20mM Na-MOPS, pH 7.0, 5 mM NaOAc and 1 mM EDTA.) Ribosomal RNA was run in outside lanes which were excised, stained with ethidium bromide and photographed to provide size standards. The remainder of the gel was blotted without any pretreatment, using 20 X SSC and nitrocellulose as described by Thomas (25). Filters were hybridized and washed as described by Thomas

(25), using 10-20 ng/ml of nick translated probe of specific activity 1-5 x $10^8$ cpm/ug. Filters were exposed to Kodak XAR film using Dupont Cronex Lightening X intensifying screens at -70°C for 3 hrs to 1 week.

S1 Nuclease. S1 nuclease mapping of HPCM2 mRNA was performed according to the method of Berk and Sharp (26). Poly (A+) cytoplasmic RNA (2 ug), 100 ng of nick translated DNA, specific activity 2-4 x $10^6$ cpm/ug, corresponding roughly to one nick per 5 kb (A. Berk, personal communication) and 5 ug of yeast tRNA were ethanol precipitated together and dissolved in 10 ul of 80% (v/v) recrystallized deionized formamide, 0.4 M NaCl, 0.04 M Pipes, pH 6.4, 1 mM EDTA. The DNA was denatured by heating at 70°C for 10 minutes, then hybridization was allowed to proceed for 3 hrs at 49°C. Hybridizations were stopped by the addition of 200 ul ice cold S1 buffer (0.25 M NaCl, 0.03 M NaOAc pH 4.5, 1 mM $ZnSO_4$, 5 ug/ml denatured calf thymus DNA and 200-600 units/ml of S1 nuclease (New England Nuclear)). The S1 reaction was incubated at 5°C for 60 minutes, then ethanol precipitated and electrophoresed on an 8% polyacylamide 7 M urea sequencing gel using end-labeled fragments of HpaII cut pBR322 DNA as standards.

Dideoxy Sequencing of mRNA. These reactions were carried out according to the method of Sanger and Coulson (27) as modified by Levy et al. (28). A 77 bp AluI-DdeI fragment from the V1 cDNA clone was isolated from a 10% polyacrylamide gel. One hundred ng of primer was denatured by boiling for 3 minutes and then hybridized with 200ng of heavy chain RNA in 1.5 X H buffer (1X H buffer is 50 mM tris, pH 8.3, 6 mM $Mg(OAc)_2$, 60 mM NaCl, 10 mM DTT) at 68°C for 45 minutes. cDNA synthesis catalyzed by avian myeloblastosis reverse transcriptase was carried out using 100 uM dATP, dCTP and dTTp, 50 μCi $^{32}P$ labeled dCTP (2000-3000 Ci/mole), 50 uM ddATP, ddGTP and ddTTP and 0.4uM ddCTP. In one set of experiments, template RNA was pretreated with methyl mercury hydroxide at a final concentration of 2.5 mM for 5 minutes, quenched with 13 mM β-mercaptoethanol (29) and then used in the sequencing reaction.

DNA Sequencing. DNA sequence determinations were carried out using the chemical cleavage method of Maxam and Gilbert (23). A subclone of ChM2-423 containing the 7.5 kb Eco R1 fragment was used as a source of HPCM2 DNA and a subclone containing the 0.6 kb BamHI fragment from Ch167α 10 (12) (kindly provided by S. Kim) was used as a source of M167 DNA. HPCM2 and M167 sequences were determined on both strands for most of the sequence.

RESULTS

The Active μ Gene From HPCM2 Contains a Rearranged Vl Gene

HPCM2 is an IgM producing hybridoma line which was shown by Gearhart et al. (8) to secrete phosphorylcholine binding antibody having the T15 idiotype. The amino acid sequence of the $V_H$ segment of μ chains from HPCM2 corresponds exactly to the germline Vl gene segment sequence (11) thus suggesting that Vl is rearranged and expressed in HPCM2 cells. We have confirmed this by isolation and analysis of a Vl-Cμ gene clone from a library constructed in Charon 4A from HPCM2 genomic DNA partially digested with EcoRl. Heteroduplex and restriction enzyme analyses of this clone, ChM2-423, clearly establish that the Vl gene is rearranged and expressed in HPCM2 (Fig. 1). In contrast to α gene clones, this clone represents only one DNA rearrangement of the Vl gene, VDJ joining. Accordingly, we decided to perform initial transcript mapping studies on ChM2-423. The 7.5 kb and 10.5 kb Eco Rl restriction fragments containing the VlDJl coding region and the Cμ coding region respectively were subcloned into pBR322 for further study.

Secreted Mu, Membrane Mu and Cμ-Only Transcripts Are Present in HPCM2 Nuclear RNA. Poly (A)+ nuclear RNA was isolated from HPCM2 cells grown in culture and size-separated on denaturing formaldehyde-agarose gels. The

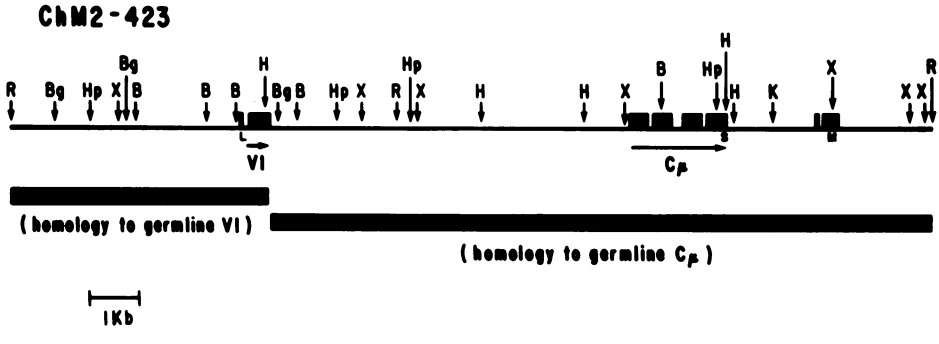


Fig. 1. Map of the μ heavy chain gene cloned from HPCM2 DNA. Restriction sites on the mouse genomic DNA insert in Clone ChM2-423 are indicated by letters: R, EcoRl; Bg, Bgl II; Hp, Hpa II; X, Xba I; B, BamHI; K, KpnI. Bgl II sites are not shown for the 10.5 Kb EcoRl fragment containing Cμ sequences. Coding sequences are indicated by raised boxes. Secreted and membrane carboxyterminal coding regions of the Cμ gene are indicated by S and M respectively; the Vl leader region is indicated by L.

The homology to germline Vl and Cμ DNA, as indicated by bars, was ascertained by heteroduplex and restriction enzyme analyses.

RNA was transferred to nitrocellulose filters and probed with Cμ (30) and Vl (31) cDNA probes. As shown in Fig. 2[a], the Cμ cDNA probe hybridized to three large RNA species: a very faint band of 12.0 kb and stronger bands at 10.1 and 9.2 kb. A strong broad region of hybridization at 2.4 kb is

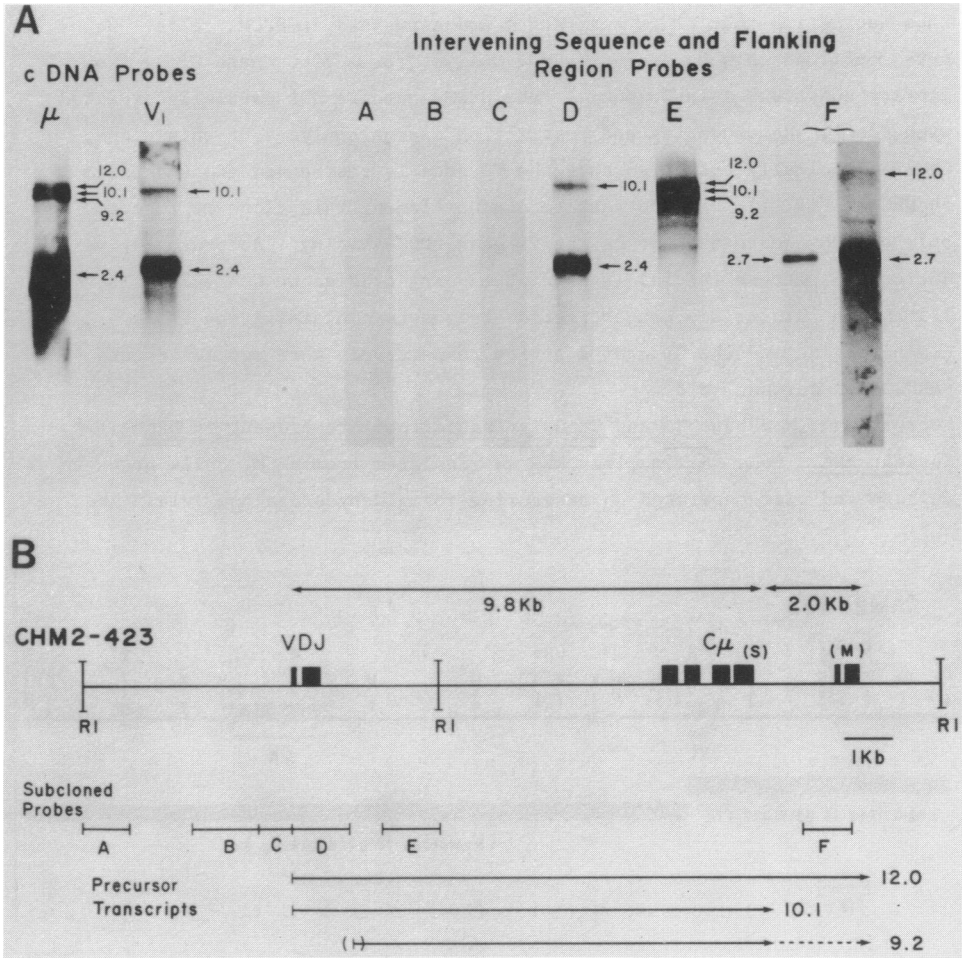


Fig. 2. RNA blot analysis of nuclear transcripts from HPCM2.
2a Poly($A^+$) nuclear RNA from HPCM2 cells was separated on 1% agarose-formaldehyde gels, transferred to nitrocellulose and hybridized with the indicated probes, as described in methods.
2b The genomic location of subcloned probes used for RNA blot hybridizations is indicated. Lines above the map indicate the lengths of genomic DNA between the coding regions indicated. The location of precursor transcripts on the genomic DNA is indicated in the lower portion of the figure.

also seen which corresponds in size to mature μmRNA. The V1 cDNA probe also hybridizes to the 2.4 Kb band and to the 10.1 kb band, but does not hybridize to the 9.2 kb or 12.0 Kb band. Failure to hybridize to the 12.0 kb band may have been caused, however, by the low amount of RNA in this band and the fact that the V1 probe is approximately one third the length of the Cμ probe. Results with the membrane probe discussed below, suggest that the 12.0 Kb transcript does in fact contain V1 sequences.

In order to map V1-Cμ transcripts on the genomic clone, we subcloned several DNA restriction fragments from the 5' flanking region, the J1-Cμ intervening sequence and membrane exon sequences of ChM2-423 as illustrated in Fig. 2b. 5' Probes "A", "B" and "C" do not hybridize with any of the Cμ or V1 containing transcripts in HPCM2 nuclear RNA. This result strongly suggests that the primary transcript of this gene initiates downstream from the 3' end of the "C" probe. The "C" probe contains a BamH1 restriction fragment located approximately 60 bp 5' to the V1 gene leader coding sequences. These results therefore suggest that transcription of the gene initiates less than 100 bp 5' of the V1 coding sequences. These results do not rule out, however, the possibility of a larger, very short-lived primary transcript of the gene which is not detected by our blotting analysis of steady state nuclear transcripts and which is rapidly processed at the 5' end. However, further experiments described in the following section show that the 5' end of the mRNA maps close to the leader coding sequence and are consistent with the interpretation that these large transcripts represent precursors which have not been processed at their 5' ends.

Lanes hybridized with probes "D" and "E" show that the 5' end of the 9.2 kb transcript maps somewhere between the 3' side of probe "D" and the 3' side of probe "E" since this transcript hybridizes with the "E" but not with the "D" probe. This is also consistent with the fact that the 9.2 kb transcript contains Cμ but not V1 cDNA sequences. Since the 9.2 kb transcript does not contain $V_1$ sequence, we assume that it does not encode a μ chain and represents a nonproductive transcript. Genomic southern blots of HPCM2 DNA show that these cells contain only a single Cμ allele, the productively rearranged one represented by ChM2-423 (M. Mercola and K. Calame, unpublished results). Therefore, the 9.2 Cμ-only transcript, as well as the productive 12.0 and 10.1 Kb transcripts, is initiated on this allele.

The "F" probe is specific for membrane sequences and contians a 1.0

kb Kpn I-XbaI fragment which includes the 3' part of the intervening sequence between secreted (s) and membrane (m) and part of the μ exons. It hybridizes to the 12.0 kb transcript and to a 2.7 kb component of the broad mRNA band. The presence of a mature 2.7 kb mRNA which was shown previously to be the size of membrane μmRNA (31,32), is consistent with the assumption that the 12.0 kb transcript does contain V1 sequence because without $V_H$ sequences it is very unlikely that a mRNA of 2.7 kb would be generated. Failure of the membrane specific probe to hybridize to the 2.4 kb secreted mRNA demonstrates that the 3' end of this RNA is 5' to the Kpn site and we assume that it contains the same 3' untranslated sequence as the secreted μmRNA from myeloma M104E (30). Faintly hybridizing species below the 12.0 kb band suggest that a small portion of the Cμ-only transcript (9.2kb) may also contain membrane sequences. (We have not mapped the processing intermediate of 5.2 kb which hybridizes with the membrane probe.) The pattern of hybridization with probe "F" demonstrates: 1) that a fraction of HPCM2 μmRNA is the 2.7 kb membrane species (32,33) and 2) the 12.0 kb transcript represent transcription across the s- m intervening sequence into the m exons, as illustrated in the lower portion of Fig. 3. Our results do not distinguish between two

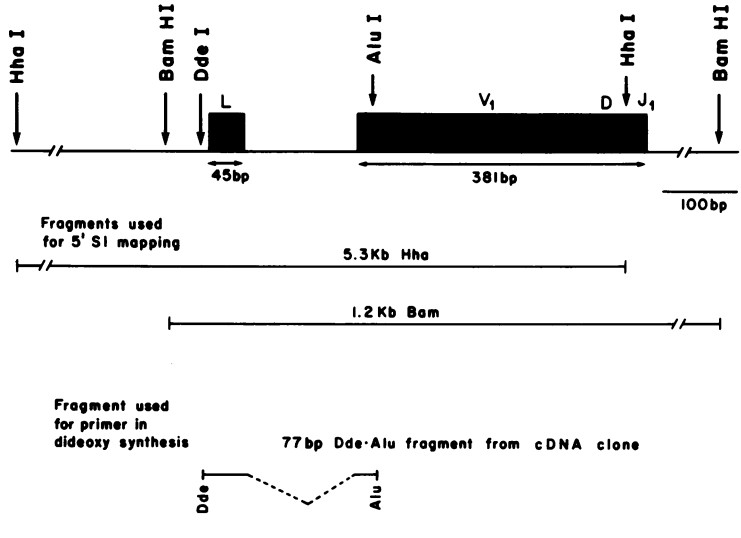


Fig. 3. Map of the V1 gene showing the DNA fragments used for S1 nuclease and dideoxynucleotide sequencing studies. Dotted lines indicate the absence of L/V1 intervening sequences in the fragment isolated from the cDNA clone.

possible synthetic routes for the 10.1 kb secreted transcript. As first suggested by Rogers et al. (33), the 10.1 kb transcript may be derived from the 12 kb transcript by processing at the 3' end, in which case the 10.1 kb transcript does not represent a primary transcript of the gene. Alternatively, the 10.1 kb transcript may represent a primary transcript which terminates 3' to the s sequences and 5' to the m (and probe "F") sequences. In either case, the transcripts map on the genomic clone as shown in the lower portion of Fig. 2b.

The 5' End of HPCM2 μmRNA Maps 63 bp From the V1 Leader Coding Sequence. Capping of nascent eucaryotic transcripts occurs very early in the transcription process (34), and it has been established that the cap site and transcription initiation site are coincident (35-36). Therefore, we have mapped the 5' end of μmRNA from HPCM2 using two different approaches: S1 nuclease protection studies (26) and primer extension with dideoxynucleoside triphosphate sequencing (27). Fig. 3 shows an enlarged map of the region of genomic DNA involved and indicates the restriction fragments which were used for these experiments.

Fig. 4 shows the results of S1 protection studies using two different uniformly labeled restriction fragments from subclones of ChM2-423. Hybridization of poly (A+) cytoplasmic RNA from HPCM2 with the 5.3 kb HhaI fragment yields two S1 protected fragments of 360±10 bp and 108±2bp (lanes C and D). The 360±10 bp protected fragment clearly corresponds to the 355 bp V1 coding region which extends from the HhaI site in the J1 region to the beginning of the intervening sequence between V1 and the leader coding region. The 108±2 bp fragment, therefore, corresponds to 45 bp of leader coding sequence plus 5' untranslated sequences of about 63±2 bp. There is a BamH1 site located 57 bp from the end of the leader coding sequence, as shown by RNA and DNA sequence analyses in the following sections. We made use of this restriction site to map the 5' S1 protected fragment more accurately. As shown in Fig. 5 lanes A and B the Bam 1.2 kb fragment generates a 380±10 bp DNA fragment and a 102±2 bp fragment. The former band corresponds to the VDJ coding region (381 bp). In this experiment the leader-5' untranslated protected fragment is 102±2 bp, 6±2 bp smaller.

The sequence of the 5' end of μmRNA from HPCM2 tumors was determined by dideoxy sequencing using purified μmRNA as the template and a restriction fragment from the V1 cDNA clone as primer for cDNA synthesis catalyzed by reverse transcriptase (27). The primer was a 77bp DdeI-AluI fragment which included the leader coding sequences and 14 base pairs 5'

to the leader, as illustrated in the lower portion of Fig. 3. Results using this primer are shown in Fig. 5a. Synthesis stopped completely after 49 bp in each lane, including a control lane (0) which contained no dideoxy chain terminators. The halt in synthesis is not a result of non-specific termination effects because we have shown by using primer fragments located closer to the 3' end of the V1 sequence, that the primer extension reaction under our conditions can proceed for at least 150 bp (data not shown). The sequence can be read clearly from within 9 bp of the DdeI site to within 2 or 3 bp of the end of the mRNA and corresponds exactly with the DNA sequence in the region 5' to the leader coding sequence (Fig. 6). These results thus place the 5' end of the μmRNA 63±1

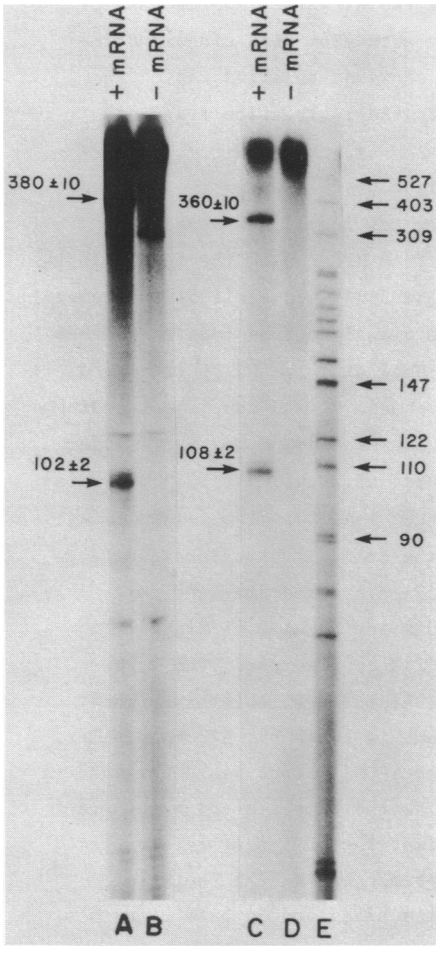


Fig. 4. S1 nuclease analysis of μmRNA from HPCM2. Poly ($A^+$) cytoplasmic RNA from HPCM2 cells was hybridized with uniformly labeled probes, Bam 1.2 kb (A&B) or Hha 5.3 kb (C&D) under R-loop conditions. After treatment with S1 nuclease, the protected fragments were separated on sequencing gels. Lane E is standard pBR322 digested with Hpa II.

bp from the leader coding sequence and 4±1 bp from the BamHl restriction site as shown in Fig. 6. This is completely consistent with the Sl protection results and thus rules out the possibility that synthesis stopped at some secondary structure block in the mRNA rather than at the 5' end. Furthermore, when mRNA was treated with methylmercury hydroxide (29) just prior to cDNA synthesis to remove secondary structure, results identical to those shown in Fig. 5a were obtained (data not shown). Placement of the 5' end of mRNA 63±1 bp 5' to the coding sequences is also consistent with the RNA blotting experiments (Fig. 2); accordingly, we believe that this is the site where Vl gene transcription initiates in HPCM2.

Transcription of the Vl Gene Initiates at the Same Site in IgA Producer M167 as in IgM Producer HPCM2.

The IgA producing myeloma M167 has been shown to express the Vl gene

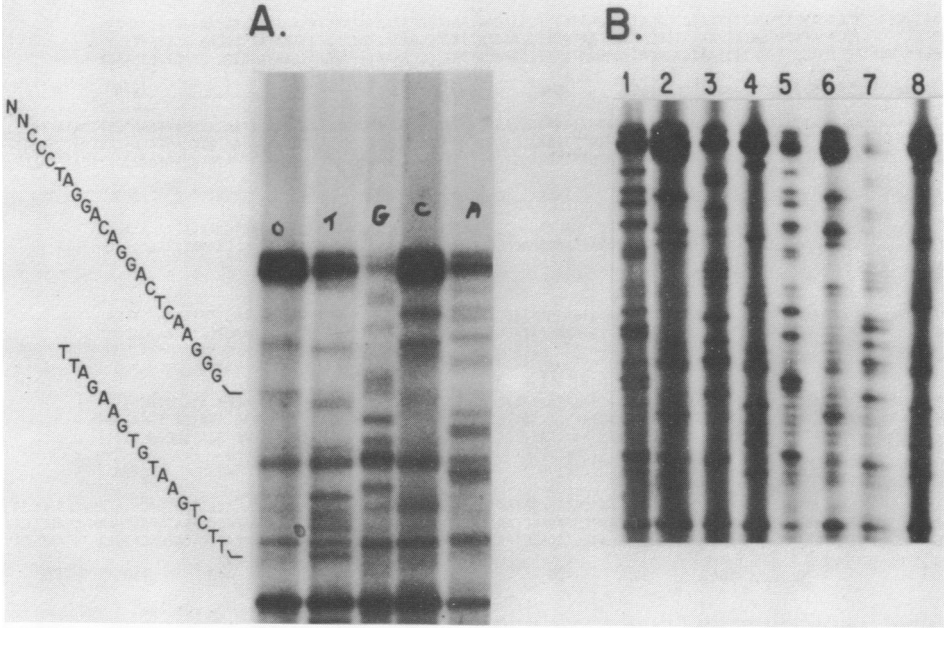


Fig. 5. Primer extension and dideoxynucleoside triphosphate sequencing of Vl gene transcripts.
5a Purified μmRNA from HPCM2 was hybridized to the 77 base pair AluI-DdI fragment shown in Fig. 4 and used to prime DNA synthesis catalyzed by avian myeloblastosis reverse transcriptase in the presence of chain terminators. Lanes indicate: A, ddATP; C, ddCTP; G, ddGTP; T,ddTTP and O, no chain terminator.
5b Products of similar dideoxy sequencing reactions using purified μmRNA from HPCM2 (lanes 1-4) and αmRNA from M167 (lanes 5-8).

(11). In order to determine if transcription of the Vl gene initiates at the same site in M167 as in HPCM2, we repeated the dideoxy sequencing experiment described above using purified αmRNA from M167 tumors and the same DdeI-AluI primer fragment. The results shown in Fig. 5b clearly show that cDNA synthesis on the M167 μmRNA template stops at exactly the same point where it stops using the HPCM2 μmRNA template. The sequence again corresponds exactly to the M167 DNA sequence 5' to the Vl leader (12) and differs in this region at one site from HPCM2 and germline sequences.

DNA Sequence Analyses of 574 bp 5' to the Cap Site of the Vl Gene Show Minimal Differences Between Germline, HPCM2 and M167 DNA.

The sequence of 574 bp 5' to the transcription initiation site of the Vl gene was determined for HPCM2 and M167 DNA; a short stretch of HPCM2 DNA 3' to the cap site was also sequenced. This analysis, shown in Fig. 6,

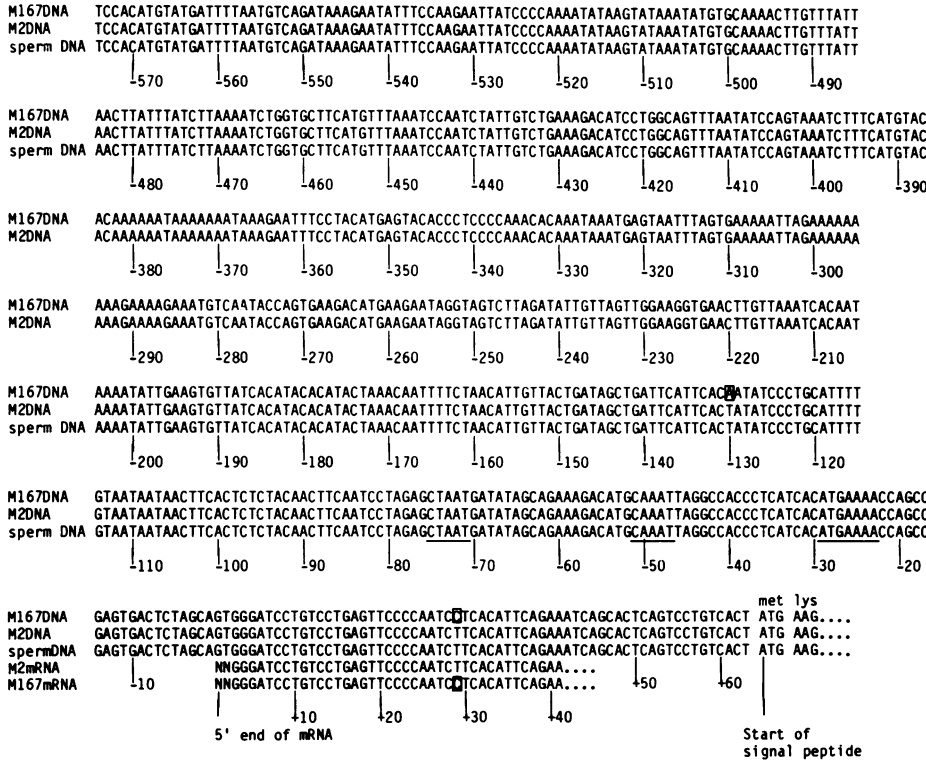


Fig. 6. DNA sequence of the N terminal portion of the Vl gene and the region 5' to it in mouse sperm DNA, HPCM2 DNA and M167 DNA.

reveals that the V1 gene has a variant of the "AATAA" box, "ATGAAAA", between -23 and -29, a "CAAT" sequence, CAAAT between -46 and -51, and a variant of the "CAAT" sequence, CTAAT, between -71 and -75 (numbering from the transcription initiation site at +1). The region 5' to these sequences appears to be very AT rich, for example from-282 to -330 is 84% AT, -466 to -521 is 86% AT and -359 to -385 is 93% AT. We also note that the region 5' to the "CAAT" sequence at -71 and extending about 450 bp in the 5' direction contains many runs of direct and inverted repeats which vary in length from ten to twenty base pairs.

Comparison of the DNA sequences from mouse sperm, HPCM2 and M167 show no differences for 129 bp 5' to the transcription initiation site. Moreover, there is only one base pair change in the entire 574 bp which were sequenced 5' to the transcription initiation site. This change occurs at -130 where a T residue in HPCM2 is replaced by an A residue in M167. The germline and HPCM2 sequences are completely identical throughout the region sequenced.

## DISCUSSION

The transcription start site of the V1 heavy chain gene has been mapped in this study 63±1 bp 5' to the leader coding sequence. Primer extension and dideoxy sequencing results were in excellent agreement with S1 protection studies in mapping the 5' end of HPCM2 μmRNA to within 1 base pair of the same location. In all known cases, the 5' end of mRNA corresponds to the transcription start site (34-38). We have confirmed that this is true for HPCM2 u transcripts by probing blots of HPCM2 nuclear RNA with subcloned fragments of 5' flanking DNA. Sequences located between 60 bp to 5 kb 5' to the V1 gene did not hybridize to u transcripts (Fig. 2).

Recently the cap site of the $V_k$ gene in MPC11 was mapped by Kelley et al. (37). Their results showed the 5' untranslated region of the kappa mRNA consisted of only three base pairs. Similar results were obtained by Hamlyn and Milstein (39) by dideoxy sequencing of the k mRNA from MOPC21. Our results demonstrate that both μ and α mRNAs containing the V1 gene contain 63 bp of 5' untranslated sequence. Kataoka et al. (40) have used primer extension and S1 protection studies to locate the 5' end of the γ1 mRNA from MC101 about 30 bp from the coding sequence. These data suggest that every functional immunoglobulin V gene segment has a transcription start site located relatively near the coding sequences. DNA sequence analyses 5' to several other V genes reveal putative "AATAA" sequences in a

region within 100 bp of V coding sequence, although transcription start sites in these cases can only be inferred because they have not been mapped (41-44). DNA sequence studies on $V_H$ genes closely related to V1 are also consistent with this hypothesis (S. Crews, G. Siu and L. Hood unpublished results).

The DNA sequence analysis in Fig. 6 shows that the region 5' to the V1 gene contains sequences characteristic of the promoters for other Pol II transcripts (38). Studies on many eukaryotic genes have shown that the "AATAA" homology region is a component of the RNA polymerase II promoter and has a role in directing specific initiation of transcription about 25-30 bp to its 3' side (38,45-49). We note that the V1 "AATAA" sequence contains an unusual G residue. Although their role in initiation of transcription is not well understood, "CAAT" sequences have been noted 5' to several genes and have been shown to be important for modulating transcription of the rabbit beta globin gene in vivo (50). There is a sequence CAAAT at -51 and another CTAAT at -70. Determination of whether the "CAAT" sequences are significant components of Ig gene promoters is not possible until more promoter regions are mapped and sequenced, although it is interesting to note that the $\gamma$1 gene analyzed by Kataoka et al. (40) has a "CAAT" sequence at -50 and the K gene described by Kelly et al. (37) also has a "CAAT" sequence at -77.

A complete understanding of the components of an RNA polymerase II promoter is not possible, however, at present (38). In addition to "ATAA" and "CAAT" sequences, transcription of some genes is also modulated by sequences two or three hundred base pairs 5' to the transcription initiation site (51,52). Therefore, we extended our DNA sequence analyses more than 500 base pairs 5' to the V1 transcription initiation site.

We have been able to compare systematically the 5' DNA sequence of a single $V_H$ gene in the germline form, after VDJ joining, and after VDJ joining plus $C_H$ switching. Our results establish that neither VDJ joining nor $C_H$ switching alters V1 DNA sequences 5' to the coding region. This result is not surprising since heteroduplex and restriction enzyme analyses had previously shown there was no detectable alteration in this region (53,54). DNA sequence analyses of regions 5' to other V genes have also shown no differences between germline and variously rearranged genes (40,10,42), although no systematic comparison of both rearrangements has been carried out previously. In addition, the dideoxy sequencing experiments on M167 mRNA demonstrate that the site of transcription initiation

for the Vl gene is not altered after class switching from u to α. Thus, the same promoter, virtually unchanged in sequence for at least 574 bp, is operative throughout B-cell development for the transcription of the Vl gene. We expect that this pattern will prove to be true for other Ig genes.

It is of particular interest that the DNA sequence of 129 bp 5' to the Vl gene cloned from M167 DNA was shown to be identical to that of HPCM2 and mouse sperm. M167 is a somatic variant of the Vl gene which expresses a heavy chain protein differing from the T15 prototype by eight amino acid residues (18). We chose to analyze the 5' flanking region of M167 because it represents the most altered form of the Vl gene which has been observed. DNA sequence analyses have shown previously that M167 has 46 alterations out of 1158 nucleotides analyzed (4% overall mutation) when compared to the germline Vl sequence (12). The alterations, which included small deletions and insertions as well as single base substitutions, occurred throughout the Vl coding region, in the Jl coding region and in the intervening sequence 3' to Jl extending thru J2 and in the 5' untranslated region. However, 5' to the transcription start site no alterations were found in the first 129 bp of sequence (0% mutation) and only one was found in the entire 574 bp 5' to the transcription start site (0.17% mutation). Although it has not been formally proved, we consider it unlikely that this single change makes a functional difference in the region. These observations establish two important points. First, the hypermutational mechanism which operates on the Vl gene is very specific with respect to where it ends; it stops near the transcription start site. It will be interesting to determine if hypermutation of V genes is related to their transcriptional state. Secondly, since 574 bp 5' to the Vl gene remains virtually unaltered in M167, it is apparent that somatic changes in this region do not cause promoter-up mutations which might have explained how variants could be selected or the increased transcription rate in plasma cells.

The results of this study clearly rule out transcriptional regulation by gene rearrangement or somatic mutation in the region extending 574 bp 5' to the transcription start site of the Vl gene. Thus we must consider and test other models for transcriptional regulation of Ig genes. At least two types of regulation are evident--the activation of transcription which appears to be concommitant with VDJ joining and the increased transcription which occurs in plasma cells. The results of Mather and Perry

(6) showing that high rates of V gene transcription are limited to rearranged V genes, in conjunction with the present work, strongly suggest that 3' DNA sequences distant from promoter regions affect the transcriptional activity of V genes. In the case of the V1 gene, these putative regulatory sequences must be located at least 649 bp 3' to the transcription initiation site (63 bp of 5' untranslated, 45 bp of leader coding, 160 bp of leader/V intervening sequence and 381 bp of V1DJ1 coding). One possibility is that negative regulatory sequences which are removed upon joining might be located 3' to germline V genes. Alternatively, positive regulatory sequences might be located 5' to C gene segments. Several lines of evidence, in fact, support the second possibility. We have presented evidence (Fig. 2) which suggests that there is a promoter site located 3' to J1 which transcribes the Cμ gene segment. Previous RNA blotting studies have demonstrated Cμ-only transcripts in other lymphoid and myeloid cell lines (16,55,56). Perry et al. (57) and Van Ness et al. (58) have studied in detail a similar transcript of the germline Ck gene. In addition, Storb et al. (59,60) have shown that both Cμ and Ck genes in B and T cells are sensitive to DNase I digestion and Rogers and Wall (61) have noted that both expressed and non-expressed Cμ alleles are undermethylated in the B cell lymphoma WEHI279. In sum, these studies show that C genes exist in a transcriptionally active state prior to V(D)J joining and suggest the presence of activating sequences near C genes. Alt et al., (56) have noted recently that deletions in the region between J4 and Cμ result in lower levels of μ transcripts in an Abelson transformed pre-B cell line. Experiments are in progress in our laboratory to search for transcriptional enhancer sequences (62,63) in the region between J4 and Cμ. One expects, however, that other developmentally regulated cellular factors may also play a role in C gene activation, because C genes are not transcriptionally active outside of the B, T or myeloid cell lineages.

Enhancement of Ig gene transcription which occurs later in B cell development could also be mediated by alterations in putative regulatory sequences located between J and C gene segments. If such regulatory sequences can be identified, this hypothesis is easily tested. One possibility, first suggested by Yaoita et al., (64) is that the GC rich region near the Cμ switch site may hinder efficient transcription required for high levels of expression in plasma cells. About 2 kb of DNA from

this region in HPCM2 have been deleted (Fig. 1), which is consistent with this hypothesis.

ACKNOWLEDGMENTS

We thank Arnold Berk, Kevin McEntee, Larry Simpson and Mark Mercola for helpful discussions concerning this work and Kevin McEntee for critically reading the manuscript.

We gratefully acknowledge J. Beard for supplying avian myeloblastosis reverse transcriptase, S. Kim for supplying a subclone of M167 genomic DNA and P. Gearhart for the HPCM2 cell line.

This work was supported by a grant from the National Institute of General Medical Sciences to K.C., J.G. was supported by a Gianini postdoctoral fellowship; C.C. was supported by a Genetics predoctoral training grant; P.D.B. by a Cell and Molecular Biology predoctoral training grant, S.C. & G.S. by NRSA fellowships.

We are grateful to Vivian Windsor for preparation of the manuscript and to John Prehn for computer analyses of DNA sequences.

REFERENCES

1. Adams, J. (1980) Immunol. Today 1: 10-17.
2. Hood, L., Davis, M., Early, P., Calame, K., Kim, S., Crews, S. and Huang, H. (1981) C.S.H.S.Q.B. 45: 887-898.
3. Honjo, T., Nakai, S., Nishida, Y., Kataoka, T., Yamawki-Kataoka, Y., Takahashi, N., Obata, M., Shimizu, A., Yaoita, Y., Nikaido, T. and Ishida, N. (1981) Immunol. Rev. 59:
4. Adams, J., Kemp, D., Bernard, O., Gough, N., Webb, E., Tyler, B., Gerondakis, S. and Cory, S. (1981) Immunol. Rev. 59: 5-32.
5. Baltimore, D. (1981) Cell 26: 295-296.
6. Mather, E. and Perry, R. (1981) Nuc. Acid Res. 9: 6855-6867.
7. Weigert, M. and Riblet, R. (1976) C.S.H.S.Q.B. 41: 837-846.
8. Gearhart, P., Johnson, N., Douglas, R. and Hood, L. (1981) Nature 291: 29-34.
9. Selsing, E. and Storb, U. (1981) Cell 25: 47-58.
10. Pech, M., Hochtl, J., Schnell, H. and Zachau, H. (1981) Nature 291: 668-670.
11. Crews, S., Griffin, J., Huang, H., Calame, K. and Hood, L. (1981) Cell 25: 59-66.
12. Kim, S., Davis, M., Sinn, E., Patten, P. and Hood, L. (1981) Cell 27: 573-582.
13. Schibler, U., Marcu, K. and Perry, R. (1978) Cell 15: 1495-1509.
14. Siden, J., Baltimore, D., Clark, D. and Rosenberg, N. (1979) Cell 16: 389-396.
15. Levitt, D. and Cooper, M. (1980) Cell 19: 617-625.
16. Zuniga, M., D'Eustachio, P. & Ruddle, N. (1982) Proc. Natl. Acad. Sci. 79: 3015-3019.
17. Bottomly, K. (1981) ICN-UCLA Symp. on Mol. and Cell. Biol. XX: 517-532.

18. Barstad, P., Farnsworth, V., Weigert, M., Cohn, M. and Hood, L. (1974) Proc. Natl. Acad. Sci. 71: 4096-4100.
19. Federoff, N., Wellauer, P. and Wall, R. (1977) Cell 10: 597-610.
20. Penman, S. (1969) in Fundamental Techniques in Virology (K. Habel and N. Salzman, eds.) Academic press, N.Y. pp. 35-48.
21. Berger, S. and Birkenmeier, C. (1979) Biochem. 18: 5143-5149.
22. Enea, V. and Zinder, N. (1975) Sci. 190: 584-586.
23. Maxam, A. and Gilbert, W. (1980) Meth. in Enzymol. 65: 499-560.
24. Dretzen, G., Vellard, M., Sassone-Corsi, P. and Chambon, P. (1981) Anal. Biochem. 112: 295-298.
25. Thomas, P. (1980) Proc. Natl. Acad. Sci. 77: 5201-5205.
26. Berk, A. and Sharp, P. (1977) Cell 12: 721-732.
27. Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. 74: 5463-5467.
28. Levy, S., Sures, I. and Kedes, L. (1979) Nature 279: 737-739. Nishioka, Y. and Leder, P. (1980) J. Biol. Chem. 255: 3691-3694.
29. Puyvar, F. and Schimke, R. (1979) J. Biol. Chem. 224: 7636-7642.
30. Calame, K., Rogers, J., Early, P., Davis, M., Livant, D., Wall, R. and Hood, L. (1980) Nature 284: 452-455.
31. Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980a) Cell 19: 981-992.
32. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) Cell 20: 313-319.
33. Rogers, J., Early, P., Carter, C., Calame, K., Wall, R. and Hood, L. (1980) Cell 20: 303-312.
34. Shatkin, A., Darzynkiewicz, E., Furuichi, Y., Kroath, H., Morgan, M., Tahara, S. and Yamakawa, M. (1982) Proc. of Brit. Biochem. Soc. in press.
35. Hagenbuchle, O. & Schibler, U. (1981) P.N.A.S. 78: 2283-2286.
36. Gidoni, D., Kahana, C., Canaani, D. & Groner, Y. (1981) P.N.A.S. 78: 2174-2178.
37. Kelley, D., Coleclough, C. and Perry, R. (1982) Cell 29: 681-689.
38. Breathnach, R. and Chambon, P. (1981) Ann. Rev. Biochem. 50: 349-383.
39. Hamlyn, P., Gait, M. and Milstein, C. (1981) Nuc. Acid Res. 18: 4485-4492.
40. Kataoka, T., Nikaido, T., Miyata, T., Moriwaki, K. and Honjo, T. (1982) J. Biol. Chem. 257: 277-285.
41. Seidman, J., Max, E. and Leder, P. (1979) Nature 280: 370-375.
42. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. and Tonegawa, S. (1980) Nature 286: 676-683.
43. Bernard, O., Hozumi, N. and Tonegawa, S. (1978) Cell 15: 1133-1144.
44. Nishioka, Y. and Leder, P. (1980) J. Biol. Chem. 255: 3691-3694.
45. Benoist, Ch. and Chambon, P. (1981) Nature 290: 304-310.
46. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Sci. 209: 1406-1414.
47. Faye, G., Leung, D., Tatchell, K., Hall, B. and Smith, M. (1981) Proc. Natl. Acad. Sci. 78: 2258-2268.
48. Osborne, T., Schell, R., Burch-Jaffe, E., Berget, S. and Berk, A. (1981) Proc. Natl. Acad. Sci. 78: 1381-1385.
49. Wasylyk, B., Kedinger, C., Corden, J., Brison, O. and Chambon, P. (1980) Nature 285: 367-373.
50. Dierks, P., van Ooyen, A., Mantei, N. and Weissman, C. (1981) Proc. Natl. Acad. Sci. 78: 1411-1415.
51. Grosschedl, R. and Birnstiel, M. (1980a) Proc. Natl. Acad. Sci. 77: 1432-1436.
52. Grosschedl, R. and Birnstiel, M. (1980b) Proc. Natl. Acad. Sci. 77: 7102-7106.

53. Davis, M., Calame, K., Early, P., Livant, D., Joho, R., Weissman, I. and Hood, L. (1980) Nature 283: 733-738.
54. Davis, M., Early, P., Calame, K. and Hood, L. (1979) ICN-UCLA Symp. on Mol. and Cell. Biol. VIII: 393-406.
55. Kemp, D., Harris, A., Cory, S. & Adams, J. (1980) Proc. Natl. Acad. Sci. 77: 2876-2880.
56. Alt, F., Rosenberg, N., Casanova, R., Thomas, E. and Baltimore, D. (1982) Nature 296: 325-331.
57. Perry, R., Kelley, D., Coleclough, C., Seidman, J., Leder, P., Tonegawa, S., Matthyssens, G. and Weigert, M. (1980) Proc. Natl. Acad. Sci. 77: 1937-1941.
58. Van Ness, B., Weigert, M., Coleclough, C., Mather, E., Kelley, D. and Perry, R. (1981) Cell 27: 593-602.
59. Storb, U., Wilson, R., Selsing, E. and Walfield, A. (1981a) Biochem. 20: 990-996.
60. Storb, U., Arp, B. and Wilson, R. (1981b) Nature 294: 90-92.
61. Rogers, J. and Wall, R. (1981) Proc. Natl. Acad. Sci. 78: 7497-7501.
62. Gruss, P., Dhar, R. and Khoury, G. (1981) Proc. Natl. Acad. Sci. 78: 943-947.
63. Levinson, B., Khoury, G., Van de Woude, G. and Gruss, P. (1982) Nature 295: 568-572.
64. Yaoita, Y., Kumagai, Y., Okumura, K. and Honjo, T. (1982) Nature 297: 697-699.