mRNA transcripts initiating within the human immunoglobulin mu heavy chain enhancer region contain a non-translatable exon and are extremely heterogeneous at the 5′ end

Geoffrey A.M.Neale and Geoffrey R.Kitchingman*
Department of Virology and Molecular Biology, St Jude Children's Research Hospital, Memphis, TN 38101, USA

Received January 11, 1991; Revised and Accepted March 25, 1991 EMBL accession no. X58529

ABSTRACT

Transcription events are thought to precede gene rearrangement in the immunoglobulin (Ig) loci and may be the mechanism by which the various gene regions are made accessible for recombination. If this is the case, identification and characterization of transcripts from the Ig loci should permit a better understanding of the gene rearrangement process. We have isolated a 2.3 kb cDNA clone from the human pre-B cell line Nalm-1 that contains enhancer-specific sequences from the Ig heavy (H) chain gene locus. The 2.3 kb transcript initiated within the enhancer region and showed extreme 5′ heterogeneity, with more than 50 initiation sites mapping near the Ig-specific octamer ATTTGCGT. Sequencing of the cDNA clone demonstrated that 644 nucleotides from the Ig enhancer region were incorporated as a leader exon spliced to the mu constant ($C\mu$) region. This leader exon contained many translation termination codons and may function to inhibit the translation of sterile $C\mu$ polypeptides. Using an enhancer-derived probe, we detected two low-abundancy mRNA transcripts with sizes of 2.3 and 12 kb. Northern blot analysis suggested that the 12 kb transcript was the unspliced precursor mRNA of a VDJ rearrangement. The potential role of these enhancer-containing transcripts in the opening of the IgH chain gene for rearrangement and for class switching is discussed.

INTRODUCTION

The differentiation of a pre-B cell to a mature B cell occurs in a sequential stepwise fashion. Essential to this maturation process is the ability of B cells to undergo somatic gene rearrangement. Gene rearrangement is required for the expression of a functional immunoglobulin (Ig) molecule (1), and is also involved later in the B-cell response during isotype switching (2, 3). Thus, control of gene recombination is crucial to the development of the mature B cell, and to its response to antigen. Substantial information is available regarding the heptamer-nonamer recognition sequences involved in gene rearrangement in both B- and T-lymphoid cells (4). However, the cellular signals that regulate the timing and specificity of these recombination events are not yet known.

One proposed model for Ig gene rearrangement (5, 6) involves transcription of the relevant loci to make these regions accessible to the recombination machinery. Experiments using artificial recombination substrates (7, 8) have suggested that a single recombination system is used for rearrangement of all of the Ig and T cell receptor (TCR) genes. Therefore, the specificity of recombination appears to be conferred by another mechanism and may involve transcriptional activation of the relevant locus (5, 6). Recent observations that kappa light chain gene rearrangement correlates with LPS-stimulation of kappa light chain gene transcription (9) and that growth factor-induced isotype switching of the murine alpha chain gene is preceded by transcription of the alpha switch region (10) strongly support this hypothesis of transcription-linked gene rearrangement in B cells.

Transcription of Ig genes is regulated by many trans-acting and cis-acting elements (11, 12). The octanucleotide ATTTGCAT (or its inverse complement) is one of these cis-acting elements and is found in most Ig gene promoters and in the heavy chain gene enhancer (13, 14). This motif is the principal determinant of B-cell specific transcription and is sufficient by itself to confer lineage specific transcription from Ig heavy and light chain gene promoters (15). After heavy chain gene rearrangement, the heavy chain gene enhancer [located in the intron between the heavy chain joining (J_H) region and $C\mu$] drives transcription from the promoter upstream of the rearranged variable (V_H) region (16, 17).

In addition to the productive transcripts originating from the rearranged heavy chain gene, sterile $C\mu$ transcripts can arise from joinings between the heavy chain diversity (D_H) region and J_H, or from the J_H-$C\mu$ intron. In pre-B and early B cells, these sterile transcripts are nearly as abundant as the mRNA encoding

* To whom correspondence should be addressed

the rearranged mu heavy chain (18–20). Recent reports show that transcription can also initiate upstream of J_1 in non-rearranged kappa light chains (21, 22). Thus, transcription from several different promoters may be involved in regulating the accessibility of the Ig loci to the recombination machinery.

We are examining the pattern of gene expression in the human pre-B cell line Nalm-1. From a preliminary screening of cDNA libraries made from these cells, we have identified a low-abundance transcript that initiates very heterogeneously within the Ig enhancer region. This transcript initiates near the enhancer octanucleotide ATTTGCGT and contains an enhancer-encoded exon that is spliced to Cμ. Here, we discuss the potential functions of this transcript and its enhancer-encoded exon in the transcription and rearrangement of Ig genes.

MATERIALS AND METHODS

Cells and cell culture

Nalm-1 is a human pre-B cell line that expresses CD10, CD20, CD24 and cytoplasmic Ig (23). KB cells are a human epithelial cell line containing unrearranged Ig genes. Cells were maintained in RPMI 1640 medium supplemented with 10% fetal calf serum.

cDNA library construction and sequencing

Total RNA was isolated by the guanidinium isothiocyanate-CsCl method (24), and poly $(A)^+$ RNA was purified by oligo-dT cellulose chromatography (25). cDNA was prepared from 3 μg of poly $(A)^+$ RNA by the method of Gubler and Hoffman (26) using AMV reverse transcriptase with an oligo-dT primer for first strand synthesis. After second strand synthesis using *E. coli* DNA polymerase I with RNase H, the double-stranded cDNA was blunt-ended with T4 DNA polymerase. Directional cloning of double-stranded cDNA was achieved using a polylinker addition method (27). Internal EcoRI and HindIII restriction sites were protected from subsequent digestion by methylation with EcoRI and AluI methylases. Methylated cDNA was ligated to an oligonucleotide (pGCTTGAATTCAAGC) that is self-complementary and contains an EcoRI site. When this oligonucleotide is ligated to cDNA, the two terminal A residues from the poly A tail at the 3′ end of the cDNA create a HindIII site that is internal to the EcoRI site. Subsequent digestion with EcoRI and HindIII generates cDNA with an EcoRI site at the 5′ end, and a HindIII site at the 3′ end of the cDNA. EcoRI/HindIII digested cDNA was inserted into Charon pBS(−) (28). This vector was constructed from lambda Charon 15 and contains the pBluescript(−) plasmid (Stratagene). The library was screened with the Cμ IgH chain gene probe (see below) and positive clones were plaque-purified, and then subcloned into pBluescript as described by Swaroop and Weissman (28). Sequencing of the 5′ and 3′ ends of cDNA clones using T7 and T3 sequencing primers (Stratagene), respectively, was performed by the dideoxy method using alkaline denatured plasmid according to the Sequenase 2.0 protocol (United States Biochemical). Full-length sequencing of clones was done stepwise in both directions using synthetic oligonucleotide primers generated from sequences obtained with the previous primer.

Probes

The genomic locations of the probes used in this study are shown in Figure 3. High specific activity ^{32}P-labeled probes were made either by random priming (29) of purified DNA fragments using Klenow DNA polymerase or by making complementary sense riboprobes using T3 or T7 RNA polymerase (Stratagene).

The Cμ, J_H, and V_H probes were made by the random-priming method (29). The Cμ probe is a 1.5 kb EcoRI genomic fragment containing Cμ1 to Cμ3. The J_H probe is a 1.6 kb EcoRI/Sau3A genomic fragment containing regions J_3 to J_6. The V_H probe is a 0.3 kb SacI fragment from a Nalm-1 cDNA clone containing a VDJ rearrangement.

The enhancer, switch, and mu membrane (μmem) probes were generated from cloned DNA by making complementary-sense ^{32}P-UTP-labeled RNA (Stratagene). The enhancer probe is a 0.3 kb BglII/TaqI genomic fragment containing the IgH chain gene enhancer octamer. The switch probe is a 2.5 kb SacI genomic fragment containing the Cμ switch region. The μmem probe was generated from a cDNA clone containing the Cμ membrane exons.

RNA analyses

RNA mapping experiments using complementary-strand RNA probes were performed with slight modifications of the method described by Melton et al (30). Ten μg of poly $(A)^+$ RNA or 50 μg of total RNA were hybridized overnight at 45°C with a ^{32}P-UTP-labeled complementary-sense RNA probe in 30 μl of 80% formamide-40 mM PIPES (pH 6.4)-400 mM NaCl-1mM EDTA. After digestion for 1 h at 30°C in 300 μl of 10 mM Tris-HCl (pH 7.5)-300 mM NaCl-5 mM EDTA-40 μg/ml RNase A, protected fragments were ethanol precipitated and analyzed on 6% polyacrylamide/7M urea sequencing gels. Size markers were an end-labeled HinfI digest of pBR322 and sequencing reactions of the cloned 0.3 kb BglII/TaqI enhancer fragment.

For Northern blotting, 3 μg of poly $(A)^+$ RNA were denatured in formamide and formaldehyde, electrophoresed through 1% formaldehyde-agarose gels (25), and blotted overnight to Nytran membranes (Schleicher and Schuell). Prehybridization, hybridization, and washing of membranes were done as recommended by the supplier.

RESULTS

cDNA library construction and screening

Toward our overall goal of examining the pattern of gene expression in Nalm-1 cells, we made two cDNA libraries from poly $(A)^+$ RNA isolated from these cells. Each library consisted of about 12×10^6 primary recombinant lambda clones, more than 96% of which contained cDNA inserts. As a prelude to more extensive screening, 5×10^4 primary recombinants from each library were screened with the Cμ probe. From this screening, six positive clones, with inserts ranging from 0.9 kb to 2.5 kb, were obtained and the inserts subcloned for sequence analysis. Because the cloning procedure we used ensures directional cloning of cDNA into the polylinker of the pBluescript plasmid, 5′ and 3′ sequence data were obtained directly using T7 and T3 sequencing primers, respectively. Matching this preliminary sequence data with published sequences (31–33) of the IgH chain gene allowed tentative identification of the origins of our clones.

The clones identified using the Cμ probe contained a number of distinct 5′ and 3′ ends. As might be expected, we identified both μ membrane and μ secreted transcripts from the 3′ end sequence data. One cDNA had a VDJ rearrangement at the 5′ end, consisting of 300 bp of a variable region from the V_H4 family (34), followed by a D_H region and the J_4 region. The 3′

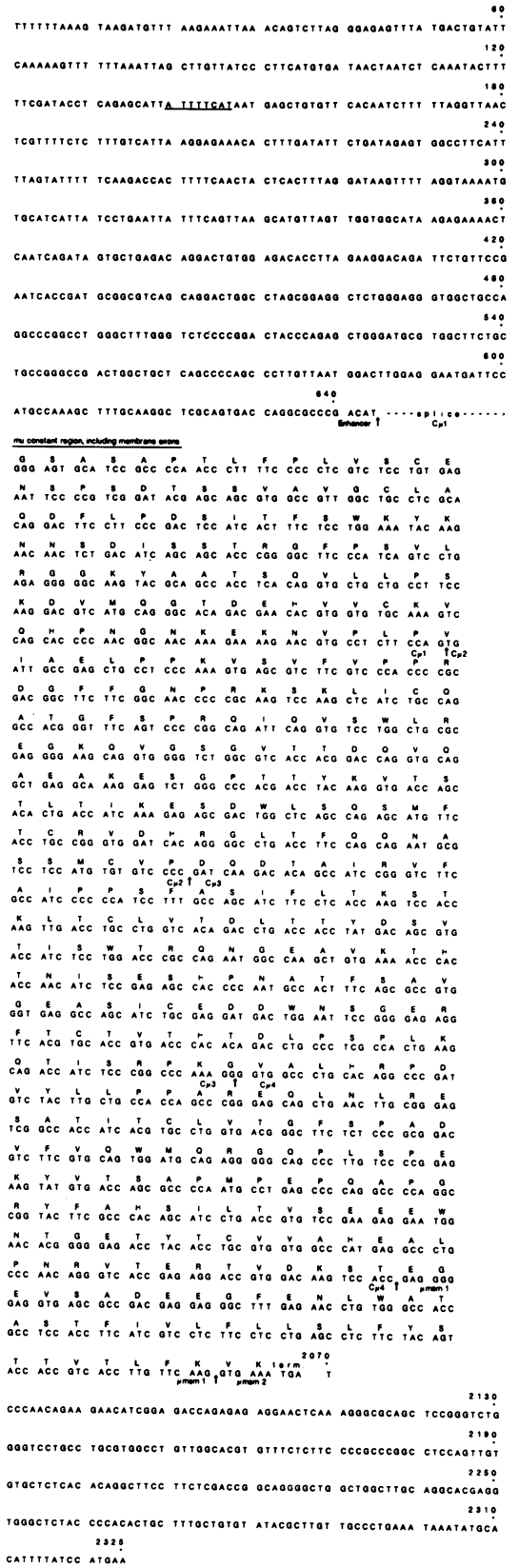

Figure 1. Nucleotide sequence of a cDNA clone initiating within the human IgH chain enhancer region. Nucleotides 1–644 are from the human IgH chain enhancer, while nucleotides 645–2325 are from Cμ. An imperfect octamer from the enhancer region is underlined. Exon boundaries are indicated by arrows.

end of this clone terminated with Cμ secreted sequences. We assume that this cDNA clone represents the functional heavy chain gene rearrangement expressed in these cells.

We also found sequences at the 5′ ends of three other cDNA clones that suggested an unusual pattern of transcription in these cells. The 5′ end of one cDNA clone started with the J_5-J_6 intervening sequence, followed by the entire J_6 region and then the first exon of the constant region. The 3′ end of this cDNA clone terminated with Cμ membrane sequences. Thus, it appears that this cDNA clone arose from a transcript originating upstream of J_6, and that the intervening sequence between J_6 and the constant region was excised precisely even though the J_6 region had not undergone gene rearrangement. This cDNA clone could have arisen by aberrant splicing of mRNA precursor transcripts from either Ig allele.

The two most striking cDNA clones had μmem sequences at their 3′ ends, were about 2.3 kb in size, and had Ig enhancer sequences at their 5′ ends. During normal processing of IgH chain gene transcripts, the intervening sequence between J_H and the constant region (containing the enhancer and switch regions) is removed by splicing. The enhancer sequences in these clones could have arisen by (i) cloning of a non-spliced precursor transcript, (ii) utilization of alternative donor/acceptor splice sites in the heavy chain locus, or (iii) initiation of transcription within the enhancer region. To distinguish among these possibilities, we characterized these cDNA clones by further sequencing and Northern blot analysis.

Sequence analysis of the enhancer cDNA clones

To determine the composition of the 2.3 kb transcripts, we determined the complete sequence of the longest cDNA clone (Figure 1). The 5′ end of this clone contained 644 bp from the enhancer region that were followed by the first exon of the constant region. Comparison with published sequences (31–33) showed that each intron of the constant region was spliced correctly to generate a μmem transcript that was preceded by the enhancer sequences. The genomic locations of the segments making up the cDNA clone are depicted in Figure 2.

Further analysis of the sequences derived from the enhancer region showed that there were eleven potential translation initiation codons, suggesting that a novel protein product may be encoded by this unusual mRNA. However, for the first ten initiation codons, there were in-frame stop codons that would terminate translation of any potential polypeptides. The last AUG codon, some 600 bp downstream from the 5′ end of the cDNA clone, was out-of-frame with the constant region polypeptide and has the potential to code for a 74 amino acid polypeptide before terminating within exon 1 of the constant region.

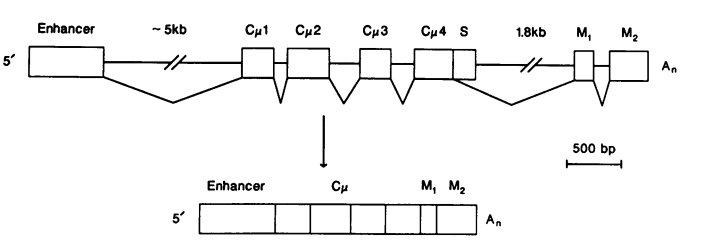

Figure 2. Schematic representation of a human cDNA clone containing human IgH chain enhancer sequences. Abbreviations used: Cμ1-Cμ4, exons 1–4 of the mu constant region; S, mu constant region secreted sequences; M_1 and M_2, mu membrane exon sequences; A_n, poly (A).

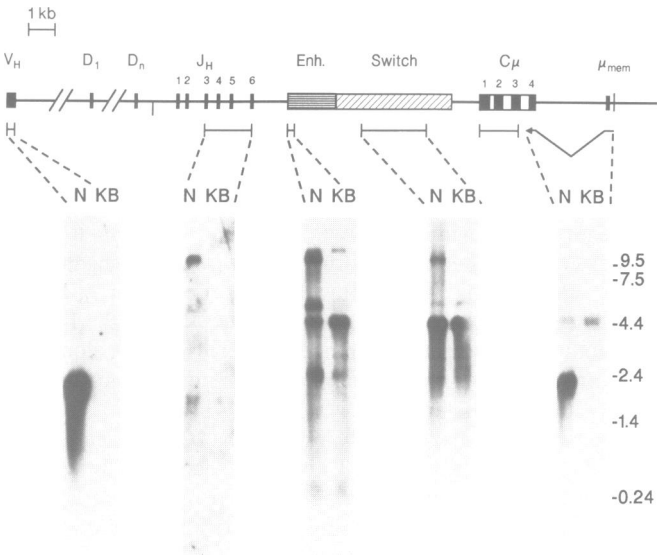

Figure 3. Northern blot analysis of poly $(A)^+$ RNA from Nalm-1 cells using probes from the human IgH chain gene locus. 3 μg per lane of poly $(A)^+$ isolated from Nalm-1 or KB cells were electrophoresed through formaldehyde-agarose gels and blotted. The genomic locations of the probes tested with the blots are shown above the autoradiograms and are defined in Materials and Methods. Abbreviations: V, variable; D, diversity; J, joining; E, enhancer; Cμ, mu constant; μmem, mu membrane. Size markers are shown in kilobases. Exposure time of blots: V_H4 probe, 6 h; J_H probe, 24 h; Enh probe, 6 h; Switch probe, 15 min; μmem probe, 15 min.

Northern blot analysis of poly (A)+ RNA using an enhancer probe

To confirm the existence of enhancer-containing transcripts in Nalm-1 cells, and to determine their size, we tested poly $(A)^+$ RNA isolated from Nalm-1 cells with an enhancer-specific probe (Figure 3). A transcript of about 2.3 kb was seen. An additional transcript of approximately 12 kb was also detected in poly $(A)^+$ RNA extracted from Nalm-1 cells but not in control poly $(A)^+$ RNA isolated from KB cells (Figure 3). The intensity of hybridization of the enhancer probe to these transcripts was much lower than that observed using either a constant region probe (μmem) or a variable region probe (V_H4), suggesting that the enhancer-containing mRNAs were relatively low in abundance (see legend to Figure 3 for exposure times of blots). The size of the 2.3 kb band on the Northern blot was in good agreement with the size of our cDNA clones, suggesting that our clones were close to full-length copies of this mRNA transcript.

We attempted to define the limits and composition of the 12 kb enhancer-containing transcript by further Northern blot analysis, testing parallel blots with probes from different regions of the heavy chain gene (Figure 3). The transcript hybridized with probes from the V_H4, J_H, enhancer, switch, and constant regions. Longer exposure times were required to visualize hybridization of the V_H4 and constant region probes to the 12 kb transcript (data not shown). Thus, it appears that the 12 kb transcript was the unspliced precursor mRNA transcribed from the VDJ gene rearrangement seen in these cells.

RNase protection assay analysis

To determine precisely where transcription of the 2.3 kb mRNA initiated within the enhancer region, we performed RNase protection assays using the BglII/TaqI 300 bp enhancer fragment. ^{32}P-UTP-labeled complementary-sense transcripts were generated from the cloned BglII/TaqI genomic fragment and hybridized to 10 μg of poly $(A)^+$ RNA. After digestion with RNase A, protected fragments were ethanol precipitated and analyzed on sequencing gels (Figure 4). Surprisingly, a very large number of protected bands were seen on the autoradiogram when the probe was hybridized with poly $(A)^+$ RNA isolated from Nalm-1 cells (Figure 4, Lane 7). Only low molecular bands resulting from incomplete RNase digestion were seen when the probe was hybridized with 10 μg of control poly $(A)^+$ RNA from the KB epithelial cell line (Figure 4, Lane 8). Numerous repetitions of this experiment using samples from one poly $(A)^+$ RNA preparation produced identical results (data not shown). To ensure that the presence of multiple bands was not an artifact arising from a partially degraded poly $(A)^+$ RNA stock, we prepared six independent samples of RNA from Nalm-1 cells

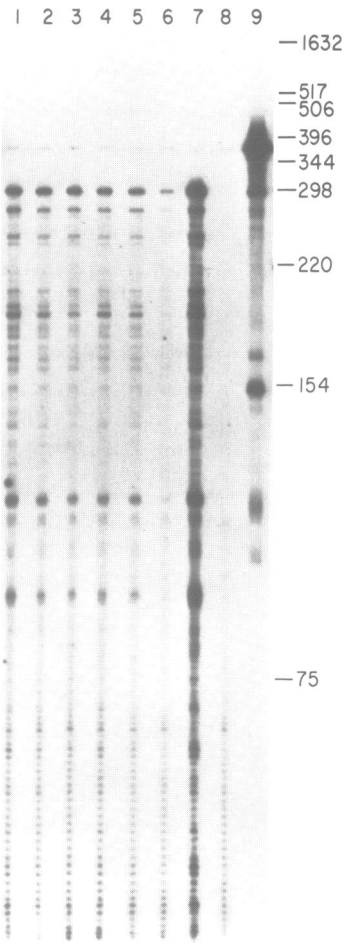

Figure 4. RNase A analysis of transcription initiation sites within the enhancer region of the human IgH chain gene locus. A 300 bp complementary-sense riboprobe generated from a cloned BglII/TaqI enhancer fragment was hybridized with RNA, digested with RNase A, and then analyzed by polyacrylamide/urea gel electrophoresis. RNA samples tested with the enhancer probe were: lanes 1–5, 50 μg of total RNA isolated from six independent samples of Nalm-1 cells; lane 7, 10 μg of poly $(A)^+$ RNA isolated from Nalm-1 cells; lane 8, 10 μg of poly $(A)^+$ RNA isolated from KB cells. Lane 9 shows a sample of non-hybridized, undigested probe. (Note: undigested probe is larger than 300 nucleotides because of vector sequences between the RNA polymerase promoter and the cloned genomic fragment.) Size markers are indicated in nucleotides.

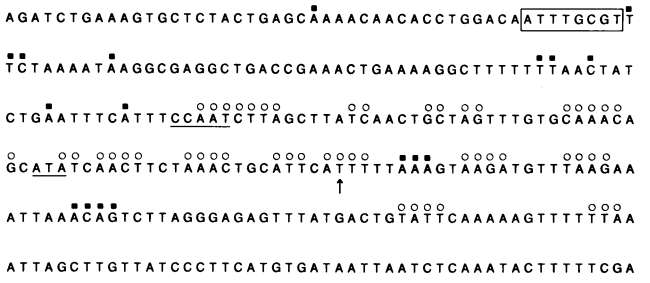

Figure 5. Transcription initiation sites within the human IgH chain enhancer region. The sequence of the BglII/TaqI 300 bp enhancer probe is presented and the locations of the transcription initiation sites found from RNase A analysis are shown. Major sites of transcription initiation are shown by squares, and other sites are indicated by circles above the sequence. The 5′ end of the longest enhancer-containing cDNA clone is shown by an arrow. The human mu heavy chain enhancer octamer sequence is boxed, and a CAAT/TATA element is underlined.

and tested them with the enhancer probe. An identical pattern of protected bands was seen for each RNA preparation (Figure 4, Lanes 1−6). From these experiments, we conclude that the observation of multiple bands in the RNase protection assay reflects extreme heterogeneity at the 5′ end of the transcripts, rather than artifacts arising from the assay system. Attempts to confirm this finding using primer extension were unsuccessful, as sufficient product could not be generated from these low-abundance transcripts.

From the RNase protection assays, we estimate that there may be more than 50 sites of transcription initiation within the enhancer region of the IgH chain gene (Figure 5). Seventeen of these sites appeared to be utilized more frequently, as indicated by the intensity of the protected bands on the autoradiogram (Figure 4). Within the BglII/TaqI genomic fragment is the IgH chain enhancer octamer ATTTGCGT. This octanucleotide is located about 65 nucleotides upstream from a CAAT/TATA-box motif (Figure 5). Nearly all of the transcripts seen on the autoradiogram initiated near these two potential promoter elements. The 5′ end of our longest cDNA began 5 nucleotides upstream from one of the major initiation sites, and 20 nucleotides downstream from the TATA-like motif.

Significantly, a small amount of full-length enhancer probe was protected in the assay with RNA isolated from Nalm-1 cells but not KB cells (Figure 4), indicating that initiation of transcription also occurred upstream of the enhancer region. This finding was consistent with results from the Northern blot experiments showing that two mRNA species hybridized to the enhancer probe, with the larger (12 kb) transcript originating upstream of the J_H region (Figure 3). Presumably, the larger transcripts were responsible for protecting the entire enhancer region probe.

DISCUSSION

We have isolated and characterized human cDNA clones that contain enhancer-specific sequences from the IgH chain gene locus. Two different size species of enhancer-containing transcripts were identified by Northern blot analysis. These transcripts are far less abundant than transcripts from the functionally rearranged Ig gene. Using RNase protection assays, we observed that the 2.3 kb mRNA initiates with extraordinary 5′ heterogeneity near the enhancer region octanucleotide ATTTGCGT. Sequencing of our cDNA clone indicated that an alternative splice site is utilized to incorporate 600−700 nucleotides from the enhancer region as an exon that is joined to the mu constant region. The questions remain as to how and why are these transcripts generated and what function the enhancer-encoded exon has.

How do these transcripts arise? The Ig octanucleotide ATTTGCAT is well conserved in heavy and light chain gene promoters (13, 14) and, in conjunction with a TATA box, is sufficient for lymphoid-specific promoter activity (15). The enhancer region of the human heavy chain gene also contains two similar, but not identical, octanucleotide sequences (Figures 1 and 5)(35, 36) that may function much like the perfect Ig octanucleotide; that is, as a target site for a lymphoid-specific regulatory protein that permits or stimulates transcription initiation. Thus, within the enhancer region there are several DNA motifs that might allow initiation of transcription. These include the Ig octanucleotide ATTTGCGT in close proximity to a TATA-like box, and a CCAAT/TATA box motif just 65 nucleotides downstream (Figure 5). Transcription within the enhancer region could initiate from either of these promoter elements.

We observed extreme heterogeneity at the 5′ end of transcripts originating from the enhancer region. While we were unable to confirm the RNase protection results with primer extension data, the consistency of the results with multiple, independent RNA preparations indicates that bands due to RNA breakdown were not present. Another possible origin for the heterogeneity observed using the uniformly labeled RNA probe is utilization of alternative 3′ splice sites. However, there are few precendents for this possibility and only one enhancer exon-Cμ splice junction was observed in the cDNA clones. Very few cellular genes with this degree of 5′ heterogeneity have been described (37−39). A transcript with multiple 5′ ends was reported recently for the B29 B-cell-specific gene, where 15−20 different transcription initiation sites were found within a span of 60 nucleotides (37). Significantly, these transcripts originated immediately downstream of Ig enhancer and promoter motifs, including the octanucleotide ATTTGCAT. Additional evidence for the ability of the murine heavy chain enhancer octamer to act directly as a promoter comes from a recent report (40) of heterogeneous initiation of transcription when the heavy chain octamer was placed 5′ to reporter genes. It thus appears that binding of the enhancer protein to its target site, or the close proximity of the octanucleotide to promoter elements, may result in great imprecision of transcription initiation. We suggest that it is the octanucleotide sequence that facilitates initiation of transcription within the enhancer region, and that the extreme heterogeneity of initiation within this region is due to the close proximity of the enhancer octanucleotide to the other promoter elements.

Why do these transcripts occur? A model proposed for the rearrangement of Ig and TCR genes suggests the existence of a common recombination system, with the specificity of recombination directed by transcriptional activation of the specific loci (5, 6). Recent evidence (9, 10) strongly supports the transcription-linked gene rearrangement model. When the kappa light chain gene is transcriptionally activated by treatment of murine pre-B cells with LPS, kappa chain gene rearrangement is detected in cells that do not normally produce a functional heavy chain (9). Additional support for the proposed model comes from experiments in which isotype switching in mouse B cell lines were examined. Germline transcripts originating 5′ to the switch

regions of the alpha (41), epsilon (42, 43) and gamma 2b (44) loci have been isolated prior to growth factor-induced isotype switching in murine B cells. Like the mu transcripts decribed here, these transcripts initiate heterogeneously about 1.5 to 2.0 kb upstream of the respective switch region sequences, but the number of initiation sites for the epsilon (42, 43) and gamma2b (44) sterile transcripts appeared to be fewer than ten. The alternative exons for the sterile transcripts from the alpha, epsilon and gamma 2b heavy chain genes are much smaller than those found attached to the human (~700 bp; this report) or mouse (~200 bp; 38) mu gene transcripts. Another difference is the presence upstream of the transcription initiation sites in the sterile mu transcripts of the enhancer elements, including the octamer and the CCAAT/TATA box; no such elements are found upstream of initiation sites in the other sterile transcripts. Thus, while similarities exist between the sterile exons added to the germline heavy chain gene transcripts, the differences indicate that expression of the mu sterile transcripts is controlled differently than transcription of similar transcripts from the other heavy chain genes.

What is the purpose of the enhancer-encoded exon in the 2.3 kb transcript? The presence of this exon in sterile mRNA transcripts was demonstrated first in murine lymphoma cells by Lennon and Perry (38). These authors have proposed that the enhancer-encoded 'nontron' functions to prevent expression of sterile Cμ polypeptides that might otherwise act as a false cellular signal of productive gene rearrangement. Sequence analysis of our cDNA clones supports this hypothesis because the enhancer-encoded exon contains 11 translation initiation codons and 35 translation termination codons (Figure 1). This number of initiation and termination codons is roughly what would be expected statistically in a random stretch of 600 nucleotides, and such a sequence of nucleotides would be untranslatable. In addition, none of the AUG codons are in a favorable context (45) to act as translation initiation codons because the sequences flanking them do not contain the appropriate recognition signals that are found almost invariably in eukaryotes (45). Thus, the enhancer-encoded exon most likely functions to inhibit the translation of the Cμ polypeptide.

The derived peptide sequence of Cμ from our cDNA clones is nearly identical to the peptide sequence obtained from the germline sequence (31–33). Our sequence is identical to the published sequence (31) with two exceptions: there is a glycine to serine substitution 26 amino acids upstream of the exon 2/exon 3 splice site, and the fourth exon of our sequence contains an extra valine residue 18 amino acids from the mu membrane exon splice junction. The origin of the discrepancies between these two sequences is unknown but may involve allelic differences.

The differentiation of B cells is an ordered and complex process. The evidence accumulated to date suggests that Ig gene rearrangement is linked to transcription from the relevant loci. We have found two low-abundance mRNA transcripts originating from the heavy chain gene locus, and have characterized the one that initiates from within the enhancer region. If this transcript is translated, non-productive heavy chain polypeptides could accumulate. Our results support the existence of a regulatory mechanism to prevent such an occurrence, by the use of an alternative splice site to create a novel enhancer-encoded 'nonsense' exon that is spliced to the Cμ region. The existence of this leader exon should prevent the translation of sterile Cμ polypeptides that might otherwise interfere with the normal regulation of gene expression in the cell.

ACKNOWLEDGMENTS

We would like to thank Drs. Sherman Weissman and Anand Swaroop for instructing one of us (G.A.M.N.) in the art of directional cloning. Dr. Zijie Sun supplied the mu switch region probe. Christy Wright provided editorial assistance. This work was supported by the American Cancer Society (IM-587) and by the American Lebanese Syrian Associated Charities.

REFERENCES

1. Blackwell,T.K. and Alt,F.W. (1989) Annu. Rev. Genet., **23**, 605–636.
2. Dunnick,W. and Stavnezer,J. (1990) Mol. Cell Biol., **10**, 397–400.
3. Dunnick,W., Wilson,M. and Stavnezer,J. (1989) Mol. Cell Biol., **9**, 1850–1856.
4. Tonegawa,S. (1983) Nature, **302**, 575–581.
5. Alt,F.W., Blackwell,T.K. and Yancopoulos,G.D. (1987) Science, **238**, 1079–1082.
6. Blackwell,T.K., Moore,M.W., Yancopoulos,G.D., Suh,H., Lutzker,S., Selsing,E. and Alt,F.W. (1986) Nature, **324**, 585–589.
7. Lieber,M.R., Hesse,J.E., Mizuuchi,K. and Gellert,M. (1987) Genes Dev., **1**, 751–761.
8. Yancopoulos,G.D., Blackwell,T.K., Suh,H., Hood,L. and Alt,F. (1986) Cell, **44**, 251–259.
9. Schlissel,M.S. and Baltimore,D. (1989) Cell, **58**, 1001–1007.
10. Lebman,D.A., Nomura,D.Y., Coffman,R.L. and Lee,F.D. (1990) Proc. Natl. Acad. Sci. USA, **87**, 3962–3966.
11. Araki,K.H., Maeda,H., Wang,J., Kitamura,D. and Watanabe,T. (1988) Cell, **53**, 723–730.
12. Maeda,H., Araki,K., Kitamura,D., Wang,J. and Watanabe,T. (1987) Nucleic Acids Res., **15**, 2851–2869.
13. Falkner,F.G. and Zachau,H.G. (1984) Nature, **310**, 71–74.
14. Parslow,T.G., Blair,D.L., Murphy,W.J. and Granner,D.K. (1984) Proc. Natl. Acad. Sci. USA, **81**, 2650–2654.
15. Wirth,T., Staudt,L. and Baltimore,D. (1987) Nature, **329**, 174–178.
16. Banerji,J., Olsen,L. and Schaffner,W. (1983) Cell, **33**, 729–740.
17. Gillies,S.D., Morrison,S.L., Oi,V.T. and Tonegawa,S. (1983) Cell, **33**, 717–728.
18. Alt,F.W., Rosenberg,N., Enea,V., Siden,E. and Baltimore, D. (1982) Mol. Cell Biol., **2**, 386–400.
19. Nelson,K.J., Haimovich,J. and Perry, R.P. (1983) Mol. Cell Biol., **3**, 1317–1332.
20. Perry,R.P. and Kelley,D.E. (1979) Cell, **18**, 1333–1339.
21. Leclerq,L., Butkeraitis,P. and Reth,M. (1989) Nucleic Acids Res., **17**, 6809–6819.
22. Martin,D.J. and Van Ness,B.G. (1989) Mol. Cell Biol., **9**, 4560–4562.
23. Minowada,J., Tsubota,T., Greaves,M.F. and Walter,T.R. (1977) J. Natl. Cancer Inst., **59**, 83–87.
24. Chirgwin,J.M. Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) Biochemistry, **18**, 5294–5299.
25. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
26. Gubler,U. and Hoffman,B.J. (1983) Gene, **25**, 263–269.
27. Dorssers,L. and Postmes,A.M.E.A. (1987) Nucleic Acids Res., **15**, 3629.
28. Swaroop,A. and Weissman,S.M. (1988) Nucleic Acids Res., **16**, 8739.
29. Feinberg,A.P. and Vogelstein,B. (1983) Anal. Biochem., **132**, 6–13.
30. Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zin,K. and Green,M.R. (1984) Nucleic Acids Res., **12**, 7035–7056.
31. Dorai,H. and Gillies,S.D. (1989) Nucleic Acids Res., **17**, 6412.
32. Rabbitts,T.H., Forster,A. and Milstein,C.P. (1981) Nucleic Acids Res., **9**, 4509–4524.
33. Friedlander,R.M., Nussenzweig,M.C. and Leder,P. (1990) Nuc. Acids Res., **18**, 4278.
34. Berman,J.E., Mellis,S.J., Pollock,R., Smith,C.L., Suh,H., Heinke,B., Kowal,C., Surti,U., Chess,L., Cantor,C.R. and Alt,F.W. (1988) EMBO J., **7**, 727–738.
35. Mills,F.C., Fisher,L.M., Kuroda,R., Ford,A.M. and Gould,H.J. (1983) Nature, **306**, 809–812.
36. Rabbitts,T.H., Forster,A., Baer,R., and Hamlyn,P.H. (1983) Nature, **306**, 806–809.
37. Hermanson,G.G., Brisken,M., Sigman,D. and Wall,R. (1989) Proc. Natl. Acad. Sci. USA, **86**, 7341–7345.

38. Lennon,G.G. and Perry,R.P. (1985) Nature, **318**, 475–478.
39. Reynolds,G.A., Basu,S.K., Osborne,T.F., Chin,D.J., Gil,G., Brown,M.S., Goldstein,J.L. and Luskey,K.L. (1984) Cell, **38**, 275–284.
40. Su,L-K. and Kadesch,T. (1990) Mol. Cell Biol., **10**, 2619–2624.
41. Lebman,D.A., Nomura,D.Y., Coffman,R.L. and Lee,F.D. (1990) Proc. Natl. Acad. Sci. USA, **87**, 3962–3966.
42. Gerondakis,S. (1990) Proc. Natl. Acad. Sci. USA, **87**, 1581–1585.
43. Rothman,P., Chen,Y-Y., Lutzker,S., Li,S.C., Stewart,V., Coffman,R. and Alt,F.W. (1990) Mol. Cell. Biol., **10**, 1672–1679.
44. Lutzker,S. and Alt,F.W. (1988) Mol. Cell. Biol., **8**, 1849–1852.
45. Kozak,M. (1989) J. Cell Biol., **108**, 229–241.