Protein-binding site at the immunoglobulin $\mu_{\text{membrane}}$
polyadenylation signal: Possible role in transcription termination

(immunoglobulin gene control/termination factor/DNA-binding protein/murine B-cell differentiation)

RONALD LAW*†, MICHIKO D. KUWABARA*‡, MICHAEL BRISKIN*, NICOLAS FASEL†, GARY HERMANSOON*,
DAVID S. SIGMAN*‡, and RANDOLPH WALL*§

* Molecular Biology Institute and † Department of Biological Chemistry, University of California at Los Angeles, Los Angeles, CA 90024; ‡ Department of Microbiology and Immunology, UCLA School of Medicine, Los Angeles, CA 90024; and § Institute of Biochimie, ch. des Boveresses, 1066 Epalinges, Lausanne, Switzerland

Communicated by M. Frederick Hawthorne, August 24, 1987

ABSTRACT mRNAs specifying immunoglobulin $\mu$ and $\delta$ heavy chains are encoded by a single large, complex transcription unit ($\mu+\delta$ gene). The transcriptional activity of $\delta$ gene segments in terminally differentiated, IgM-secreting B lymphocytes is 10–20 times lower than in earlier B-lineage cells expressing $\delta$ mRNA. We find that transcription of the $\mu+\delta$ gene in IgM-secreting murine myeloma cells terminates within a region of 500–1000 nucleotides immediately following the $\mu_{\text{membrane}}$ ($\mu_m$) polyadenylation signal. Transcription decreases only minimally through this region in murine cell lines representative of earlier stages in B-cell development. A DNA fragment containing the $\mu_m$ polyadenylation signal gives protein-DNA complexes with different mobilities in gel retardation assays with nuclear extracts from myeloma cells than with nuclear extracts from earlier B-lineage cells. However, using a recently developed “footprinting” procedure in which protein-DNA complexes resolved in gel retardation assays are subjected to nucleolytic cleavage while still in the polycrylamide gel, we find that the DNA sequences protected by factors from the two cell types are indistinguishable. The factor-binding site on the DNA is located 5' of the $\mu_m$ polyadenylation signal AATAAA and includes the 15-nucleotide-long A+T-rich palindrome CTGTAAACAAATGTC. This type of palindromic binding site exhibits orientation-dependent activity consistent with the reported properties of polymerase II termination signals. This binding site is followed by two sets of directly repeated DNA sequences with different helical conformation as revealed by their reactivity with the chemical nucleases 1,10-phenanthroline-copper. The close proximity of these features to the signals for $\mu_m$ mRNA processing may reflect a linkage of the processes of developmentally regulated $\mu_m$ polyadenylation and transcription termination.

Transcription termination sites delineate the 3' extremities of eukaryotic transcription units for mRNA (reviewed in refs. 1 and 2). Transcription termination has been studied in a variety of eukaryotic genes (reviewed in ref. 3). These include the genes encoding globin (4), ovalbumin (5), $\alpha$-amylase (6), dihydrofolate reductase (7), gastrin (8), and immunoglobulin $\kappa$ light chain (9). In all these studies, several features of termination were observed. Transcription termination follows poly(A) sites. The sites of transcription termination vary considerably in distance beyond the polyadenylation signals of different genes. Termination occurs at multiple sites as reflected by a gradual decline in transcription.

Developmentally regulated transcription termination plays an important role in controlling gene expression from the complex $\mu+\delta$ transcription unit encoding the heavy chains of IgM ($\mu$) and IgD ($\delta$) (10–14). Although the features of transcription termination in the $\mu+\delta$ gene generally resemble termination in other eukaryotic genes, termination between the $C_{\mu}$ and $C_{\delta}$ constant gene segments is only exerted in terminally differentiated plasma cells or in their myeloma-cell equivalents (10–14). The precise location and the mechanism for termination within the $\mu+\delta$ transcription unit has not been defined. Platt (2) has classified terminators as either constitutive or regulatable. This classification implies that regulatable terminators necessarily require trans-acting factors. Accordingly, the termination of $\delta$ transcription is presumably controlled by trans-acting factors, reflective of different stages in B-lymphocyte development, that interact with relevant termination signals. Other factors, including higher-order structures reflected in altered chromatin conformation or in unique DNA secondary structure, may also influence the process of transcription termination (reviewed in refs. 1 and 2). The observed changes in chromatin structure at the 3' ends of the lysozyme (15), globin (16), ovalbumin (17), and histone (18) genes support this view.

Here we report that the active region of transcription termination within the $\mu+\delta$ gene closely follows the $\mu_m$ polyadenylation site ($\mu_m$ represents the region encoding the carboxyl terminus of $\mu$ heavy chain of membrane-associated IgM). We further show that nuclear extracts from B-lineage cells active or inactive in transcription termination yield different patterns of protein-DNA complexes in electrophoretic mobility-shift assays with a DNA segment containing the $\mu_m$ polyadenylation signals (19, 20). These complexes were "footprinted" directly in the gel by a recently developed procedure employing the chemical nucleases 1,10-phenanthroline-copper ion (oPhen-Cu) (21). The protein-binding domain was indistinguishable from terminating and nonterminating cells and included a 15-nucleotide-long A+T-rich palindrome immediately preceding the $\mu_m$ polyadenylation signal AATAAA. This factor-binding site and the $\mu_m$ polyadenylation signal are closely followed by two sets of short directly repeated sequences that display abrupt changes in helix geometry as revealed by their reactivity to oPhen-Cu, a nucleolytic reagent sensitive to geometric parameters of the minor groove of DNA (reviewed in ref. 22). The presence of a protein-binding site and the unusual features of DNA secondary structure in this region are considered relative to a possible mechanism for developmentally regulated transcription termination in the immunoglobulin $\mu+\delta$ gene.

Abbreviations: $\mu_m$ and $\mu_m$, regions encoding the carboxyl termini of membrane-associated and secreted $\mu$ heavy chains, respectively; $C$, constant-region gene segment; oPhen, 1,10-phenanthroline; MPE, methidiumpropyl-EDTA.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
MATERIALS AND METHODS

Cells. The murine pre-B-cell line JS61-10 is an Abelson virus-transformed fetal liver cell line obtained from Owen Witte (23). M2 is an IgM-secreting murine myeloma cell line (24). The hybridoma cell line MXW231 and the 70Z/3 pre-B-cell line have been described (12).

Nuclear Transcription Assays. In vitro transcription assays (25) were carried out as described (10, 25). Unlabeled poly(A) and poly(C) were included along with sheared calf thymus DNA in the hybridization reaction mixtures. The complete region from the Cε3 domain through Cε4 was subcloned and used as probes for hybridization with labeled RNA from transcription assays.

Electrophoretic Mobility-Shift (Gel Retardation) Analysis. Nuclear extracts were prepared according to Dignam et al. (25). Binding reactions were performed using the general conditions described by Sen and Baltimore (26). The με141 DNA (10⁴ cpm), specifically labeled at either end, was incubated in 10 μl of 10 mM Tris-HCl, pH 7.5/10 mM NaCl/0.1 mM EDTA containing 5.1 μg of poly(dI-dC). Binding was initiated by the addition of 7.5 μg of protein from nuclear extract and samples were incubated for 30 min at room temperature. Samples then were loaded on a 6% (39:1 crosslinking) acrylamide gel and electrophoresed in 1 X TBE (89 mM Tris borate/2 mM EDTA, pH 8.3) for 2.5 hr at 200 V.

oPhen-Cu “Footprinting” in the Acrylamide Matrix. The following reagents for the oPhen-Cu footprinting reaction were purchased from the indicated suppliers: 1,10-phenanthroline (oPhen) and 2,9-dimethyl-1,10-phenanthroline (Me2oPhen) (G. F. Smith); CuSO4·5H2O (Mallinkrodt); 3-mercaptopropionic acid (Aldrich). Analysis of protein-binding sites protected from oPhen-Cu cleavage (footprinting) in the acrylamide matrix was carried out according to the method of Kuwabara and Sigman (21). End-labeled με141 DNA (10⁹ cpm) was incubated in 20 μl of poly(dI-dC) and 15 μg of nuclear extract protein. The wet acrylamide retardation gel was exposed to x-ray film for 30 min at room temperature until the retarded band was visible. The bands of interest were then excised and immersed in 100 μl of 50 mM Tris-HCl (pH 8.0). Then 10 μl of a solution containing 2.0 mM oPhen and 0.45 mM CuSO4 was added, followed by 10 μl of 58 mM 3-mercaptopropionic acid. Digestion was allowed to proceed for 10 min at room temperature and was then quenched by addition of 10 μl of 28 mM Me2oPhen (100% ethanol) and 270 μl of 0.5 M ammonium acetate/1 mM EDTA. The DNA was then eluted overnight at 37°C. The eluted DNA was ethanol-precipitated, resuspended in 80% (vol/vol) formamide/10 mM NaOH/1.0 mM EDTA/0.1% bromophenol blue/0.1% xylene cyanol, and analyzed by electrophoresis in a 10% acrylamide DNA-sequencing gel. Autoradiograms of digested-DNA gels were scanned with an LKB 2222-010 Ultroscan XL densitometer.

Nuclease Digests. The general conditions for DNA cleavage in solution were as described by Spassky and Sigman (27). με141 (4 × 10⁴ cpm), labeled at one end, in 20 μl of 50 mM Tris-HCl, pH 7.4/75 mM NaCl/5 mM MgCl2 was digested with oPhen-Cu (170 μM oPhen/40 μM CuSO4/4.8 mM 3-mercaptopropionic acid, 2 min at room temperature, 2.1 mM Me2oPhen quench), DNase I (50 μg/ml, 2 min at room temperature, 50 mM EDTA quench), or methyllumpropyl-EDTA-iron(II) (MPE-Fe) (10 μM MPE/10 μM Fe(NH4)2(SO4)2/4 mM dithiothreitol, 15 min at room temperature, 50 mM EDTA quench, ref. 28) and then ethanol-precipitated. Purified DNA was electrophoresed in a 10% acrylamide sequencing gel calibrated with G+A lanes as described by Maxam and Gilbert (29).

RESULTS

Most Transcription Termination Occurs Within 543 Nucleotides Downstream of the με Poly(A) Site. The subcloned DNA probes used in mapping the location of transcription termination between Cε4 and Cε3 are shown in Fig. 1. (All of these probes except number 4 lack repetitive sequences.) In vitro transcription assays were carried out on nuclei isolated from JS61-10 pre-B cells, representative of an early stage in B-cell development, and from IgM-secreting M2 myeloma cells, representative of the terminally differentiated stage. Labeled RNA from in vitro nuclear transcription reactions was hybridized to an experimentally established excess of probe DNA in slot blots (Fig. 1). The intensities of labeled RNA hybridized with the different probe DNAs were determined by densitometry and normalized per nucleotide of probe DNA. The histograms in Fig. 1 show the normalized hybridization relative to the με-specific probe (probe 3). The transcription of the με and μm regions is essentially identical in both the pre-B cells and the myeloma cells. This result is consistent with previous findings that transcription does not terminate between the με and μm exons in most B-lineage cells, including myeloma cells (10–14, 31). In agreement with previous results (10–14, 31), we observed that Cε3 transcription was reduced >95% in the M2 myeloma cells. Repeated measurements of transcription in isolated nuclei indicated that ongoing transcription in M2 myeloma cells declined 60–70% over the 543-nucleotide-long DNA fragment immediately following the με poly(A)-addition site (Fig. 1). Transcription normalized over the length of DNA probe 6 [beginning with the HindIII site 543 nucleotides past the μm poly(A) site] was reduced 80–90% from the μm values. These results establish that the majority (60–70%) of the transcription termination events in M2 myeloma cells occurred within DNA.
543 nucleotides past the $\mu_m$ poly(A) site and that most residual transcripts terminated over the next 514 nucleotides. Thus transcription termination in these IgM-secreting M2 myeloma cells was completed within 1 kb following the $\mu_m$ polyadenylation site. These results extend earlier studies, which did not directly measure transcription in these segments immediately following the $\mu_m$ polyadenylation site (10–14). In contrast, transcription over this 1-kb region in the JS61-10 pre-B cells decreased only 15–20% below $\mu_m$ values. The $C_4$ sequences are transcribed in pre-B cells at a level 20% that of $\mu_\nu$ or $\mu_m$ sequences. This decrease in $C_4$ transcription in the pre-B cells primarily occurred over the intergenic region containing the large, unique inverted sequence elements (i.e., probe 7; see ref. 25) located 1.2 kb past the $\mu_m$ polyadenylation site. Most B-cell lines exhibit a pattern of transcription like that described in pre-B cells (10–14, 31).

Identification of a Protein-Binding Domain Near the $\mu_m$ Poly(A) Site. The exonuclease III protection assay of Wu (32) was initially employed to detect nuclear protein-binding sites in the region of active transcription termination. For this assay, we used a 660-base-pair (bp) Xba I–HindIII fragment which included the 543-bp region downstream of the $\mu_m$ poly(A) site and about 100 bp of the 3’ untranslated region which included the 543-bp $C_4$ 2-exon (30). Nuclear extracts from both the pre-B-cell line JS61-10 and the IgM-secreting hybridoma MXW231 generated strong exonuclease III stops, indicative of protein binding, which mapped near the $\mu_m$ poly(A) site (data not shown).

To further define the protein-binding site located near the $\mu_m$ poly(A) site, a 141-bp HindII fragment (designated $\mu_m$141, see Fig. 2) was used for gel retardation assays and footprinting. First, protein–DNA interactions in the $\mu_m$141 fragment were analyzed by acrylamide gel electrophoretic mobility-shift assays (19, 20). Nuclear extracts were prepared from two pre-B-cell lines, JS61-10 and 70Z/3, and the IgM-secreting hybridoma MXW231. Binding reactions were carried out using end-labeled $\mu_m$141 fragment and nuclear extracts in the presence of poly(dI-dC) to suppress nonspecific protein–DNA interactions (26). The results (Fig. 2) reveal that the major species displaying altered electrophoretic mobility are distinct when JS61-10 and 70Z/3 pre-B cells are compared with the MXW231 hybridoma cells, which exhibit active transcription termination like the M2 myeloma cells (10–14). However, binding of protein to this fragment is not lymphoid-cell-specific, as a nuclear extract from HeLa cells gave a retarded complex that comigrated with the pre-B-cell complexes (data not shown).

Footprinting with the Chemical Nuclease oPhen-Cu Reveals Identical Protein-Binding Sites in Pre-B and IgM-Secreting Cells. We attempted to localize protein-binding sites on the $\mu_m$141 fragment by standard solution footprinting techniques using oPhen-Cu (27), DNase I, and MPE-Fe (28). All attempts to footprint this region under conditions of solution binding proved unsuccessful. This failure to obtain a protected DNA fragment (footprint) could have resulted from an intrinsically weak protein–DNA interaction, yielding unstable complexes in solution, or from a low abundance of the binding protein in the crude nuclear extracts. These limitations can be circumvented by using a method developed by Kuwabara and Sigman (21) in which protein–DNA complexes identified in a gel retardation assay can be treated within the gel matrix with the chemical nuclease oPhen-Cu. The oPhen-Cu cleavage reaction uses readily diffusible reactants, which work effectively in the gel matrix and provide footprints completely consistent with those obtained in solution (21).

Results of such footprinting experiments on the protein–DNA complexes identified in the gel retardation assay with the $\mu_m$141 DNA fragment (Fig. 2) are presented in Fig. 3. The sequences protected from oPhen-Cu cleavage by bound nuclear proteins were confirmed by quantitative densitometry of autoradiograms of control DNA in comparison to protein–DNA complexes. Indistinguishable patterns of protection were obtained with either JS61-10 or MXW231 nuclear extracts. The region in DNA fragment $\mu_m$141 protected with both extracts covers 19 nucleotides, 5’ GTAAAACAAATGTCACATGG 3’, with the $\mu_m$ AAT AAA polyadenylation signal positioned 1 nucleotide from its 3’ border (Figs. 3 and 5). This protein-binding site is located 27 nucleotides 5’ of the $\mu_m$ cleavage/polyadenylation site. The most striking feature of the protein-binding domain is its overlap with the A+T-rich palindrome 5’ CTGTAAACAAATGTC 3’, where the underlined sequence comprises the 5’ end of the site of protection to oPhen-Cu cutting in the $\mu_m$141 fragment.

Unusual DNA Structures in the $\mu_m$141 DNA Detected with Nuclease Probes. The 44-nucleotide-long sequence immediately following the $\mu_m$ cleavage/polyadenylation site is highly homologous (>80% nucleotide sequence identity) between mice and humans (10, 33). Since this homology is higher than for most of the $\mu$ coding sequences, this region may play an important regulatory function. The sequences surrounding the $\mu_m$ cleavage/polyadenylation site are characterized by the presence of multiple directly repeated sequences (Figs. 4 and 5). The reactivity of these sequences to the nucleolytic agents oPhen-Cu F4/F5 (22), DNase I, and MPE-Fe (28) is presented in Fig. 4. Cleavage of the coding strand of end-labeled $\mu_m$141 DNA with oPhen-Cu reveals an unusual pattern of reactivity. The directly repeated sequences 3’ TAAATATA 5’ (R1) are hyperreactive to cleavage by oPhen-Cu in both locations. These two repeated elements are located immediately 3’ of the $\mu_m$ polyadenylation signal and cleavage site, respectively. In contrast, both copies of the second set of directly repeated sequences, 3’ AACACACTT 5’ (R2), are inefficiently cut by oPhen-Cu in both locations.
**DISCUSSION**

We have determined that the majority of developmentally regulated transcription termination in the complex \( \mu + \delta \) gene occurs over a segment of 543 nucleotides following the \( \mu_m \) poly(A) site. Using a method that combines an acrylamide gel electrophoretic mobility-shift assay with oPhen-Cu footprinting (19-21), we identified a protein-binding domain that is recognized by factors present, both in pre-B and IgM-secreting cells. This factor-binding site is located immediately 5' of the \( \mu_m \) AATAAA polyadenylation signal and is situated over an A+T-rich palindrome. Although the protein-DNA complexes generated with either pre-B-cell or hybridoma nuclear extracts display different electrophoretic mobilities in mobility-shift assays (Fig. 2), the DNA sequences protected from oPhen-Cu cleavage in both cases are indistinguishable. Since mobility-shift assays are also sensitive to protein-protein interactions, the interaction of an additional protein(s) with a DNA-binding protein common to both cell types (as footprinted by oPhen-Cu) could account for the slower-migrating complexes observed with pre-B-cell extracts versus myeloma extracts. Alternatively, limited proteolysis of the DNA-binding proteins in myeloma cell extracts may provide a trivial explanation for the altered migration. Following the polyadenylation signal are two pairs of repetitive sequences (Fig. 5) with markedly different reactivities toward oPhen-Cu. Since this reagent senses minor-groove geometry, fluctuations in this structural feature of DNA may influence termination of transcription in conjunction with protein factors bound upstream of the AATAAA polyadenylation signal.

The study of transcription termination in a variety of eukaryotic genes transcribed by polymerase II suggests the involvement of multiple elements that vary in their sequence and spatial proximity to one another (reviewed in refs. 1 and 2). Principally, these studies suggest that the signals for termination precede the regions in which termination (i.e., polymerase disengagement) occurs. A major impediment to the deciphering of termination signals in eukaryotic genes transcribed by polymerase II has been the inability to detect any significant homology at the nucleotide level (1-3). Computer homology searches have identified loosely related,
short, A+T-rich sequences analogous to a yeast consensus termination signal, TTGGTTTG (1), in the 3' flanking regions of eukaryotic genes where termination occurs. Based on their identification of an A+T-rich termination element from the gastrin gene, Sato et al. (8) proposed that oligo(dT) tracts present in A+T-rich elements might promote transcript release from the DNA template due to the marked instability of oligo(dA-U) hybrids (34).

For the ρ+δ gene, the unique structure of the protein-binding site located upstream of the ρ+δ polyadenylation signal suggests a possible role in termination apart from its A+T-richness. The palindrome in the binding site, CTGTAAACAAATGTC, is an inverted repeat that lacks dyad symmetry (sequences with dyad symmetry have the same 5'→3' sequence on both strands). Such a sequence, unlike those possessing a dyad axis of symmetry, would direct a binding protein to assume a unique orientation relative to upstream and downstream flanking sequences. Orientation-specific activity for eukaryotic termination signals has been demonstrated for both the β-globin (35) and gastrin (8) genes. An inverted repeat with a 2-fold axis of symmetry would not impose a similar constraint on a binding factor and is less well suited for the direction-specific process of transcription termination. Possible sequence-specific binding to a single strand (i.e., coding strand) of the template DNA could be essential for events that diminish the processivity of RNA polymerase.

Recent studies indicate that the signals involved in cleavage/polyadenylation are required in transcription termination (35, 36). The polyadenylation signals and the 3' flanking sequences of a murine α-globin gene were found to be necessary for efficient termination of transcription in the adenovirus EIA gene (35). In a thalassemic human α-globin gene, mutation of the AATAAA polyadenylation signal to AATAAG abolished cleavage and polyadenylation at the normal site and also resulted in transcription read-through well beyond the normal termination region (36). That the factor-binding site and the unusual structural features revealed in our studies are in close proximity to the signals for ρm processing implies that processes in ρm cleavage/polyadenylation and developmentally regulated transcription termination may be linked.

This work was supported by research grants from the National Institutes of Health (GM21199 to D.S.S.) and CA12800 to (R.W.) and the National Science Foundation (8311332 to R.W.) and by Public Health Service National Research Awards CA09056-11 (G.H.), AI07126 (M.B.), and GM07104 (R.L.).