HUMAN J CHAIN GENE

Structure and Expression in B Lymphoid Cells

By EDWARD E. MAX and STANLEY J. KORSMEYER

From the Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, and the Metabolism Branch, National Cancer Institute National Institutes of Health, Bethesda, Maryland 20205

J chain is a 15 kilodalton protein that links together IgM or IgA monomers to form pentameric IgM or dimeric IgA (reviewed in reference 1). The regulation of J chain expression in B cells has been studied by several laboratories in an effort to understand the developmental program of gene activation in the B lymphocyte. Some results suggest that the program for J chain expression may be different between mouse and man. Studies of mouse cell lines representing different stages of B cell differentiation have suggested that J chain expression is a late step of mature B cells; J chain appeared to be absent at the pre-B stage and in cells representative of unstimulated B cells (expressing surface IgM only), but present in large amounts in plasmacytomas and in mitogen-stimulated splenic lymphocytes (2–5). In contrast to this pattern for mouse lymphocytes, several reports have indicated that human lymphocytes may begin J chain expression at an earlier stage (6), perhaps even before synthesis of Ig (7).

To further characterize J chain expression in human lymphocytes and to extend the analysis to the DNA level, we have obtained genomic clones for the human J chain gene. The cloned gene has been examined by nucleotide sequence analysis to establish its identity as a functional J chain gene. The resulting sequence has been compared with the previously published human J chain sequence, determined at the amino acid level, and with a murine J chain complementary DNA (cDNA)[1] clone. Because of the presence of nine candidate polyadenylation signals (AATAAA or AATTAAA) downstream of the C-terminal coding block in the human clone, the 3' end of the gene could not be established from the sequence alone; S1 nuclease protection experiments using RNA from the B lymphoid line BALM-4 have allowed identification of the functional polyadenylation site. Hybridization probes from the clone have been used to explore the human genome (by Southern blot experiments) for related sequences. Finally, in initial experiments investigating the developmental program of this gene, a cloned J chain probe has been used to analyze J chain transcripts in a variety of human lymphoid cells representative of pre-B and B cell stages of differentiation.

Materials and Methods

Cloning Methods. The clones HuJ1 and HuJ2 were constructed by inserting agarose gel–purified (Bull's Eye apparatus, Hoefer Scientific Instruments, San Francisco, CA)

---

[1] *Abbreviations used in this paper:* cDNA, complementary DNA; mRNA, messenger RNA.

restriction endonuclease fragments into phage arms, and packaging (8) using a commercial extract (Packagene, Promega Biotec, Madison, WI). For HuJ1, Bam HI fragments of human placental DNA were ligated into Bam HI arms of Charon 28; for HuJ2, Eco RI fragments of DNA from the cell line REH (9) were ligated into Eco RI arms of λgtWES. Clones were screened by the in situ hybridization method of Grunstein and Hogness (10) using a 330 basepair Bam HI-Hinf I fragment from the mouse J chain cDNA clone (3, 11) as a probe.

Sequence Analysis. Subclones of phage DNA fragments hybridizing to the mouse cDNA probe were constructed in pBR322 or pBR327, and were sequenced by the chemical degradation method of Maxam and Gilbert (12), or, after subcloning into M13 vectors, by the dideoxy terminators method of Sanger et al. (13).

S1 Nuclease Analysis. The 3' end of the messenger RNA (mRNA) encoding the human J chain was determined by the S1 nuclease method of Berk and Sharp (14) as presented in Maniatis et al. (15) with modifications. In a typical experiment, 20 μg of total cellular RNA from the J chain–expressing B cell line BALM-4 was coprecipitated with a double-stranded DNA probe labelled at one 3' end. The nucleic acids were resuspended in 10 μl hybridization buffer (40 mM PIPES, pH 6.4, 1 mM EDTA, 0.4 M NaCl, 80% formamide) and incubated at 72°C for 15 min, 43°C for 30 min, 27°C for 30 min and 20°C for 2 h. Aliquots (5 μl) were diluted into 50 μl of S1 buffer (0.28 M NaCl, 0.5 M sodium acetate, pH 4.6, 4.5 mM ZnCl₂, 20 μg/ml heat-denatured salmon sperm DNA) with or without S1 nuclease (20 U/ml) (Bethesda Research Laboratories, Gaithersburg, MD), and incubated at 20°C for 1 h. The reaction was stopped by chilling on ice. 10 μl of 4 M ammonium acetate, with 100 mM EDTA was added, and the nucleic acids were extracted with 1:1 phenol/CHCl₃, precipitated with isopropranol, resuspended in formamide loading buffer, and electrophoresed on a denaturing 6% acrylamide-urea gel.

Southern Blots. Human placental DNA was digested with restriction endonucleases (Bethesda Research Laboratories, or New England Biolabs, Beverly, MA), electrophoresed (10 μg/lane) on 0.7% agarose gels, and blotted onto nitrocellulose (Millipore, Bedford, MA) by the method of Southern (16). The blots were hybridized to nick-translated probes in a formamide–dextran sulfate buffer (17).

J Chain Expression in Lymphoid Cells. 18 well-characterized cell lines or leukemias were examined for the presence of J chain sequences in total cellular RNA. NALL-1, HPB-Null, NALM-6, NALM-16, REH, and NALM-1JR are B cell precursor lines established from acute lymphoblastic leukemia or chronic myelogenous leukemia. Their Ig expression, Ig gene configuration, and B cell–associated surface antigens are listed in Table I and have been reported previously, in part (9, 18). HL-60 is a promyelocytic leukemia (19) and RPMI 8402 a T cell acute lymphoblastic leukemia; both retain germline Ig gene configuration (9). CLL-DC is a μ, λ surface Ig–positive chronic lymphocytic leukemia that does not secrete Ig (20). BHM-23 (21), RPMI 8392 (21), B85 (21) and SUDHL-6 (22) are IgM-secreting cell lines. SB26-11 (21) and BALM-4 (23) secrete IgG; Arr2.5/2 and Arr2.2/11 (spontaneous B cell lines from a lymph node; Z. Landow, unpublished observations) and GM1056 (Human Genetic Mutant Cell Repository, Camden, NJ) secrete IgA. Examination of Ig gene configuration in these cells was accomplished by Southern blots using a 2.2 kilobase (kb) Sau 3A JH probe, a 1.3 kb Eco RI Cμ probe, a 2.3 kb Hind III-Sma I Cγ probe, a 2.5 kb Eco RI Cκ probe, and a 0.8 kbase Bgl II-Eco RI Cλl probe, as previously described (9). Total cytoplasmic RNA (5–10 μg/lane), isolated from these cells by the guanidine isothiocyanate method of Chirgwin et al. (24) was electrophoresed in 0.7% agarose-formaldehyde gels (15), blotted onto nitrocellulose or Gene Screen Plus (New England Nuclear, Boston, MA) and hybridized to nick-translated probes.

Results

Cloning the Human J Chain Gene. The substantial homology between the amino acid sequence reported for the human J chain protein and that derived from nucleotide sequence of a mouse J chain cDNA clone suggested that a probe from the mouse clone might cross hybridize with the human gene. Initial

TABLE I

Characteristics of Cells

Cell	Type	J chain mRNA	Ig produced			Ig gene configuration			B cell–associated surface antigens			
			Cyto	Surface	Secreted	H chain	κ	λ	HLA-DR	B4	CALLA	B1
NALL-1	Pre-B ALL	−	−	−	−	2 JH rearr	2 del	germ	+	+	+	+
HPB-null	Pre-B ALL	+	μ	−	−	1 JH rearr, 1 del	germ	germ	+	+	−	−
NALM-6	Pre-B ALL	−	μ	−	−	2 JH rearr	2 del	1 rearr	+	+	+	−
NALM-16	Pre-B ALL	−	−	−	−	1 JH rearr	germ	germ	+	+	−	−
REH	Pre-B ALL	−	λ	−	λ	2 JH rearr	2 del	2 rearr	+	+	+	−
NALM-1 JR	Pre-B CML	+	μ	−	−	1 JH rearr	germ	germ	+	+	+	+
HL-60	Promyelocytic leukemia	−	−	−	−	germ	germ	germ	+			
8402	T cell ALL	−	−	−	−	germ	germ	germ				
SB26-11	EBV B cell	−	ND	γ, κ	γ, κ	2 JH rearr, H chain switch	1 rearr	1 rearr				
Arr 2.5/2	Spontaneous B cell line	+	ND	α, λ	α, λ	1 JH rearr, H chain switch	ND	1 rearr				
Arr 2.2/11	Spontaneous B cell line	+	ND	α, κ	α, κ	1 JH rearr, H chain switch	ND	germ				
BALM-4	B cell acute leukemia	+	ND	γ, κ	γ, κ	2 JH rearr, H chain switch	ND	ND				
SU-DHL-6	B cell DHL	+	ND	μ, κ	μ, κ	1 Cμ rearr	1 rearr, 1 del	1 rearr				
B85	EBV B cell	+	ND	μ, κ	μ, κ	1 Cμ rearr	2 rearr	germ				
GM1056	EBV B cell	+	ND	α, λ	α, λ	1 JH rearr, H chain switch	1 rearr, 1 del	1 rearr				
8392	EBV B cell	+	ND	μ, λ	μ, λ	1 Cμ rearr	2 del	2 rearr				
CLL-DC	B cell CLL	−	ND	μ, λ	−	1 Cμ rearr	2 del	2 rearr				
BHM-23	Burkitt B cell	+	ND	μ, κ	μ, κ	1 Cμ rearr	1 rearr, 1 del	germ				

ALL, acute lymphoblastic leukemia; CML, chronic myelogenous leukemia; EBV, Epstein-Barr virus-transformed; DHL, diffuse histiocytic lymphoma; Cyto, cytoplasmic Ig; H chain, heavy chain gene; rearr, rearrangement; del, deletion; germ, germline; ND, not done. Test for cytoplasmic Ig and the presence of B cell–associated surface antigens were performed on the pre-B cells, as previously described and partially reported (9, 18, 19). The DA-2 antibody was used to detect a nonpolymorphic HLA-DR determinant; B4 (gp 40/80) heterodimer recognized by anti-B4 is a B cell–restricted antigen; CALLA (gp100, common acute lymphoblastic leukemia antigen) was recognized by the J5 monoclonal antibody; B1 (pp35) is a B cell–restricted antigen recognized by anti-B1.

Southern blots of human placental genomic DNA hybridized with a 330 basepair Bam H1 to Hinf I probe (11) from the mouse cDNA (a generous gift of Dr. Marian Koshland, University of California, Berkeley) demonstrated either one or two bands (depending on restriction enzyme used) that hybridized intensely to the probe. Despite this seemingly favorable result, no genomic clones were obtained on screening of several gene libraries of human DNA. Each screen encompassed 8–10 × 10⁵ plaques; libraries were constructed from partial Mbo I digests of DNA from human cell lines or tissues, cloned into the Bam HI site of Charon 28. Each library had already been successfully screened for other genes. As an alternative approach, Bam H1 fragments of human DNA were size-fractionated by preparative agarose gel electrophoresis, and a 7.5 kb positively hybridizing fraction was ligated into Bam H1-cut arms of Charon 28, and packaged. One clone isolated from this enriched material, HuJ1, was found by nucleotide sequence analysis to contain the two 3' exons of a gene encoding the human J chain. In an attempt to obtain the remaining 5' region of the gene, probes from this clone were used for several additional library screens, but once again no isolates were identified. Therefore, we ligated into phage arms and packaged a preparative agarose gel fraction of 8 kb Eco R1 fragments that was expected (on the basis of genomic Southern blots) to contain a DNA segment overlapping the Bam H1 clone but extending further 5' (see Fig. 1, top). The packaged phage were screened with a probe from the 5' end of HuJ1. One resulting clone, HuJ2, was found to contain an exon encoding the remainder of the sequence of the mature J chain protein.

The Human J Chain Gene. The three exons from the cloned DNA segments (see Fig. 1) encode an amino acid sequence very similar to the one previously determined from chemical analysis of the protein by Mole et al. (25) and homologous with the murine sequence determined from a cDNA clone (11) (Fig. 2). The N-terminal amino acid is glutamine, consistent with the blocked pyrrolidone carboxylic acid found at the protein level (25); this residue is absent from the murine protein. Presumably a signal peptide is encoded in an exon lying upstream; this would correspond to the situation in mouse (M. Koshland, personal communication), and would be consistent with the presence of an appropriately located ". . .AG" (ending at position 228 of Fig. 1) that is similar to the consensus sequence for mRNA splice acceptor sites (26).

The J chain sequence contains eight cysteine residues that are thought to play an important role in the linkage of Ig monomers. The sequence deduced from the human gene reveals that all of the cysteine residues are at positions homologous to the cysteines in the murine sequence (Fig. 2). Although the protein sequence reported by Mole et al. (25) also deduced eight cysteines, the cysteine at residue 101 of our sequence was not identified by them, and residue 89 (serine in both our sequence and the murine sequence) was identified as cysteine (see Fig. 2). In addition to the cysteine residues, a glycosylation site (asparagine-isoleucine-serine, residues 49–51) is conserved between the human and murine sequences. The lengths of the two amino acid sequences are identical, 137 residues. The absence of the N-terminal glutamine in the murine sequence is compensated by the insertion of an extra aspartate after residue 95 of the human sequence, in a region that is poorly conserved between the two sequences.
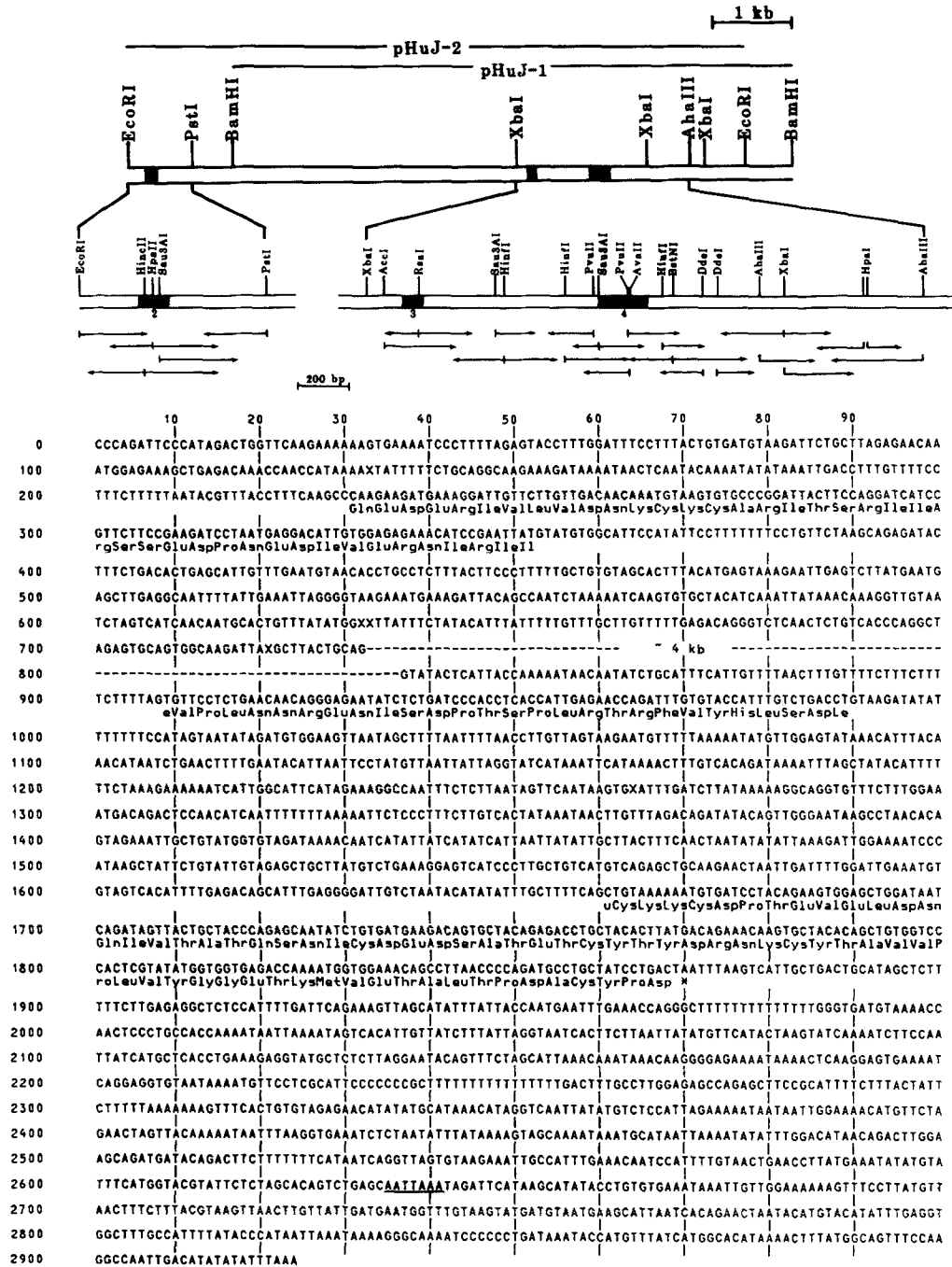
```
_____EXON 2_____          _____EXON 3_____

         10        20        30        40                      50        60
          |         |         |         |                       |         |
    QEDERIVLVDNKCKCARITSRIIRSSEDPNEDIVERNIRII          VPLNNRENISDPTSPLRTRFVYHLSDL      HUMAN GENE
    K--------------------///--------EIV-///--          ---D-----------------------      HUMAN PROTEIN
    D--ATI-A----M-T-V-----P-T-------------V            ----------------RN--------      MOUSE cDNA


                    _____EXON 4_____

      70        80        90  95 96  100       110       120       130   137
       |         |         |   |  |   |         |         |         |     |
    CKKCDPTEVELDNQIVTATQSNICDED SATETCYTYDRNKCYTAVVPLVYGGETKMVETALTPDACYPD      HUMAN GENE
    --X--------------------CB----N X-X-X//--------------------Z-----B----B      HUMAN PROTEIN
    ------V----ED-V----------V--DGVP----V--------TM---R-H------QA-----S----      MOUSE cDNA
```
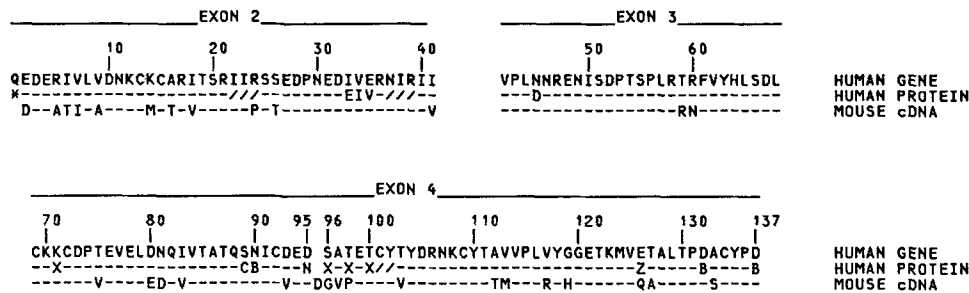
FIGURE 2. J chain amino acid comparisons. The sequence of the mature human J chain protein deduced from the gene is presented, divided into the three coding blocks of exons two, three, and four. For comparison, the chemically determined amino acid sequence of the J chain protein is presented below. Dashes represent residues identical to those of the gene-defined sequence; (/) represents residues omitted from the sequence. The third line gives the amino acid sequence deduced from a mouse cDNA clone (11), incorporating the correction identified by Elliot and Steiner (45).

Although the possibility of gene polymorphism as an explanation for the discrepancies between our sequence and that of Mole et al. (25) has not been rigorously excluded, it seems likely that the latter study concluded that the human J chain is 129 amino acids long because two arginine tripeptides (residues 22–24 and 37–39), as well as residues 101 and 102, were missed in the analysis. If each insertion/deletion between the mouse and human amino sequences is counted as a single substitution, then there are 31 amino acid changes in the 137-residue sequence, or about 77% amino acid concordance between the two J chains.

We believe that our cloned DNA segment corresponds to a functional J chain gene. The nucleotide sequence of Fig. 1 contains all expected mRNA splice signals, no termination codon until the expected end of the amino acid sequence (TAA at position 1,869) and an appropriate polyadenylation signal (AATTAAA at position 2,635, as discussed below), so it is a good candidate for a functional gene. Since the nucleotide sequence encodes an amino acid sequence very similar to that previously reported for the human J chain, probes from this clone should certainly hybridize to the functional J chain gene. Genomic Southern blots of human DNA probed with DNA fragments from our clone identify only two hybridizing bands. One corresponds to the cloned gene, while the other demonstrates homology apparently confined to exon four (as discussed below), and thus could not encode a complete J chain. For these reasons, we conclude that the exons we have cloned derive from a single functional gene for the human J chain.

Polyadenylation Signal. The 3' end of eukaryotic mRNA is generally signalled by the hexanucleotide AATAAA, or the variant AATTAAA, found ~20 nucleotides 5' of the poly(A) tail in most mature mRNA. The human J chain gene reveals nine candidate polyadenylation signals (AATAAA or AATTAAA) in the sequenced region downstream of exon four (Fig. 1), leaving the position of the 3' end of the gene ambiguous. The homology between the mouse and human genes in the 3' untranslated region is much less than in the coding sequences, but there is sufficient similarity to identify in the human sequence the AATTAAA homologous to the polyadenylation signal in the mouse (Fig. 3). This
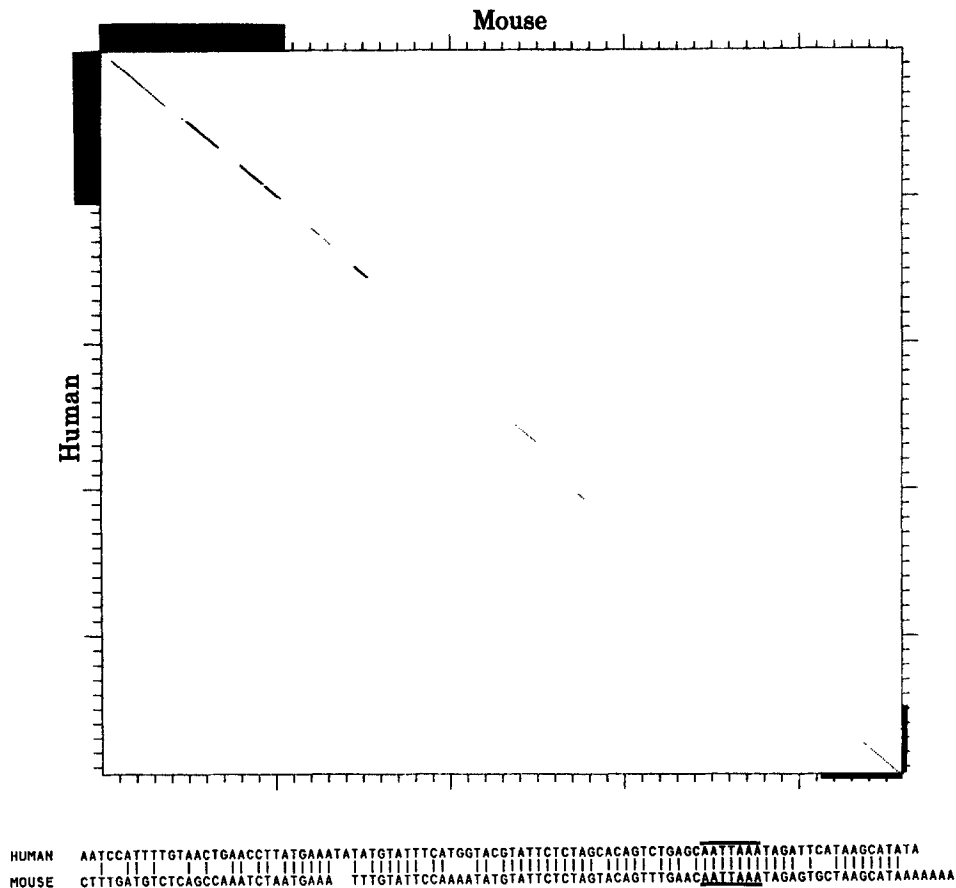
Mouse


```
HUMAN   AATCCATTTTGTAACTGAACCTTATGAAATATATGTATTTCATGGTACGTATTCTCTAGCACAGTCTGAGCAATTAAATAGATTCATAAGCATATA
         | | || |  | |  || || ||||||| |  ||||||| || |  || |||||||||||| ||||| ||| |||||||||||||| |  ||||||||
MOUSE   CTTTGATGTCTCAGCCAAATCTAATGAAA  TTTGTATTCCAAAATATGTATTCTCTAGTACAGTTTGAACAATTAAATAGAGTGCTAAGCATAAAAAAA
```

FIGURE 3. Dot-matrix comparison of exon four of human and mouse J chain nucleotide sequences. Sequence homology between the two nucleotide sequences was displayed by a dot-matrix program similar to that described by Maizel and Lenk (46). In such a program, two sequences to be compared are linearly arrayed in computer memory as if one were strung on the X axis and the other on the Y; for every point (X, Y) in the matrix, the computer compares the corresponding sequences on the two axes, and places a dot at (X, Y) if the sequences meet a specified criterion of similarity. For the present matrix, the criterion for a dot was eight nucleotide matches in a window of 10 basepairs. The sequences compared represent the entire exon four, i.e. the fourth coding block (represented by black rectangles on the axes at upper left) and 3' untranslated region. The markers on the axes represent 20 basepair intervals. Only short homology diagonals are present in the 3' untranslated region. The insert below shows the sequence represented by the bars on the axes at lower right; this region corresponds to the 3'-most homology diagonal, including the sequences flanking the AATTAAA polyadenylation signal in both species.

sequence occurs at position 2,635 (Fig. 1), 766 basepairs from the end of the coding sequence, and is the seventh of the nine polyadenylation signal candidates. In the mouse cDNA clone, the DNA segment between exon four and the poly(A) tail contains only the one functional polyadenylation signal. To define the 3' end of the human J chain gene and determine whether the AATTAAA homologous to the murine signal was the functional one, S1 nuclease mapping was used (see Fig. 4).

Either of two 3′ end-labelled DNA probes was hybridized with total RNA from BALM-4, a human B cell line documented by Northern blotting to contain abundant J chain RNA. The RNA-DNA hybrid was treated with S1 nuclease, and the samples were run on a denaturing gel to measure the size of the DNA protected from S1 digestion. The formation of RNA-DNA hybrids is favored by high G-C content. Probe A, an Xba I-Aha III fragment of 524 basepairs, was extremely A-T rich (G-C content only 29%), and so the S1 protection was difficult to detect. Nevertheless, on overexposed gels a band at 260 basepairs was reproducibly found after RNA hybridization and S1 treatment, but was absent in control experiments (left panel, Fig. 4). To verify this S1 protection, the experiments were repeated with probe B. This 1,106 basepair fragment had a slightly higher overall G-C content (32%), but included a 72 basepair portion of coding sequence with a G-C content of 51%. In addition to its potential for stronger hybridization, this probe was capable of detecting transcripts ending at the four upstream polyadenylation candidates. As shown in the middle panel of Fig. 4, probe B detected only a single large transcript. When this band was sized by running the gel further (right panel), the end of the transcript coincided with that detected by the shorter A probe; this position is approximately at nucleotide 2,660 (Fig. 1) about 20 basepairs 3′ of the AATTAAA homologous to the polyadenylation signal found in mouse.

A Second Homologous Sequence. Initial genomic Southern blots of human DNA digested with various restriction enzymes demonstrated, for most enzymes, two bands hybridizing with a probe from the mouse J chain cDNA clone. Characterization of the human J chain gene allowed identification, for each enzyme, of the band contributed to the Southern blot by this gene. A map could then be constructed of the "extra" bands. To determine the extent of homology between the J chain gene and this extra homologous sequence, probes derived from the cloned gene were hybridized to Eco RI digests of human DNA. In such digests, the functional J chain gene corresponding to our clones appears as an 8 kb band, and the second homologous sequence as a 4.5 kb band. As shown in Fig. 5, the 4.5 kb band hybridizes to probes containing exon four, the 3′ untranslated region, and 3′ flanking region (probes B through F) but not to probes 5′ of exon four (probe A). Thus, the 4.5 kb band does not appear to contain a sequence homologous to exon three. An alternative explanation is that the extra sequence might represent an intact gene that is cut by Eco RI such that its exon three resides on an 8 kb Eco RI fragment fortuitously comigrating with the 8 kb fragment containing the cloned gene. This possibility is diminished by the fact that only a single band hybridizing to probe A is also apparent in an Eco RI/Sst I double-digest (Fig. 5). It is unlikely that two sequences would yield bands comigrating under both digestion conditions. Full characterization of the second hybridizing sequence will require cloning and sequence analysis of this DNA segment.

Several probes derived from the cloned gene in the region flanking exon two contain reiterated sequence, as judged by the intense smear they produced on hybridization to genomic Southern blots. One reiterated sequence identified by nucleotide sequence analysis is an *Alu* repeat (27), which begins around position 640 of Fig. 1, and extends downstream into a region of the clone that has not

Probe A

Probe B

Probe B
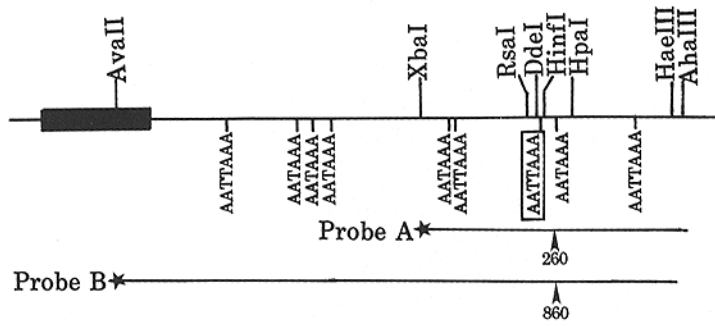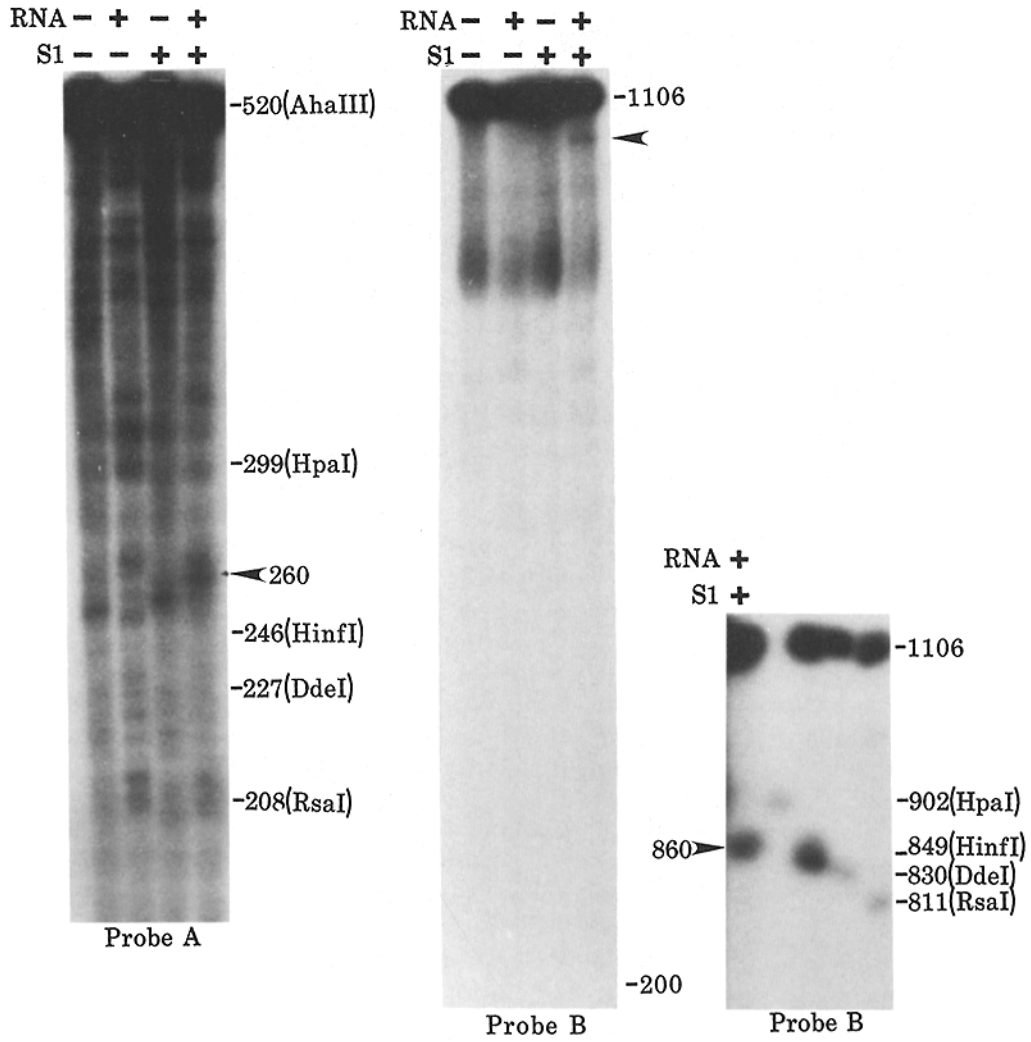
been sequenced. A computer search of the GenBank (Bolt, Beranek, and New-man, Cambridge, MA) library revealed no other sequence homologies of apparent significance.

J Chain Gene Transcription in Human Cells of B Lymphocyte Lineage. To initiate our investigation of J chain regulation, we have used Northern blot analysis to examine J chain RNA in several human cell lines and tumors representing different stages of B lymphocyte development. Total cellular RNA from selected cells was electrophoresed on a formaldehyde-agarose gel, blotted onto Gene Screen Plus or nitrocellulose membranes, and probed with the ~600 basepair Pvu II/Xba I exon four probe from the HuJ1 clone.

Fig. 6 and Table I reveal that J chain may be transcribed as early as the pre-B cell stage in humans. Both HPB-Null and NALM-1JR are well-characterized human B cell precursor lines that contain cytoplasmic μ chain, but do not synthesize light chain or secrete Ig; yet they are strongly positive for J chain mRNA. However, J chain mRNA was not apparent in four of the six pre-B–like lines examined. Of these, three demonstrated no cytoplasmic μ (NALL-1, NALM-16, REH), and only one (NALM-6) was positive for cytoplasmic μ. Thus, while the J chain gene clearly can be transcribed at the B cell precursor stage, J chain mRNA is not always present in all such human cells, appearing in our sample only in those cells with cytoplasmic μ. Of the ten lines we examined that were representative of mature B cells, eight, incuding all of the IgM or IgA secretors, contained J chain RNA easily detectable in samples of total cellular RNA. The correlation of J expression with heavy chain isotype (i.e., J expression in cells making μ or α) is not strict; BALM-4, a γ, κ cell, contains abundant J message, whereas none was apparent in our blot of RNA from the μ, λ cell CLL-DC. The lack of detectable J chain message in the latter cell may reflect a developmental stage slightly less mature than that of the other B cell lines in that, typical of CLL cells, it displays surface Ig but does not secrete Ig. Such nonsecreting B cells would not synthesize pentameric IgM, and would have no apparent need for J chain protein.

RNA from two cell lines not in the B lymphocyte lineage, the promyelocytic line HL-60 and the T cell line 8402, showed no hybridization to our J probe, as expected.

The J chain mRNA was observed as a single band that migrates at the same position in all cells with a detectable transcript. This result is consistent with the single functional polyadenylation site detected by S1 nuclease analysis. The size

FIGURE 4. The 3' end of the message as defined by S1 nuclease mapping. Total cellular RNA from the B cell line BALM-4 was hybridized to 3' end-labelled probe A (left panel) or B (middle panel) as described in Methods. Samples were treated with S1 nuclease and run (with controls) on denaturing urea-polyacrylamide gels. As size markers, samples of the same probes were cut by restriction enzymes Dde I, Hinf I, Rsa I, and Hpa I (partial digestions were used in the case of the longer probe B), and run on the same gels. The last lane of the middle panel was duplicated in the first lane of the right panel; this gel was run much longer in order to compare the protected band with the size marker bands shown in the remaining lanes of this panel. The map below illustrates the relationship of the two probes to the exon four coding block of the gene (black rectangle), the nine candiate polyadenylation signals, and the restriction sites used in the analysis. The boxed AATTAAA sequence is homologous to the polyadenylation signal in the mouse cDNA clone (Fig. 3). The gels indicate a single protected fragment terminating at a position 260 nucleotides from the labelled Xba I end of probe A and 860 nucleotides from the labelled Ava II end of probe B.
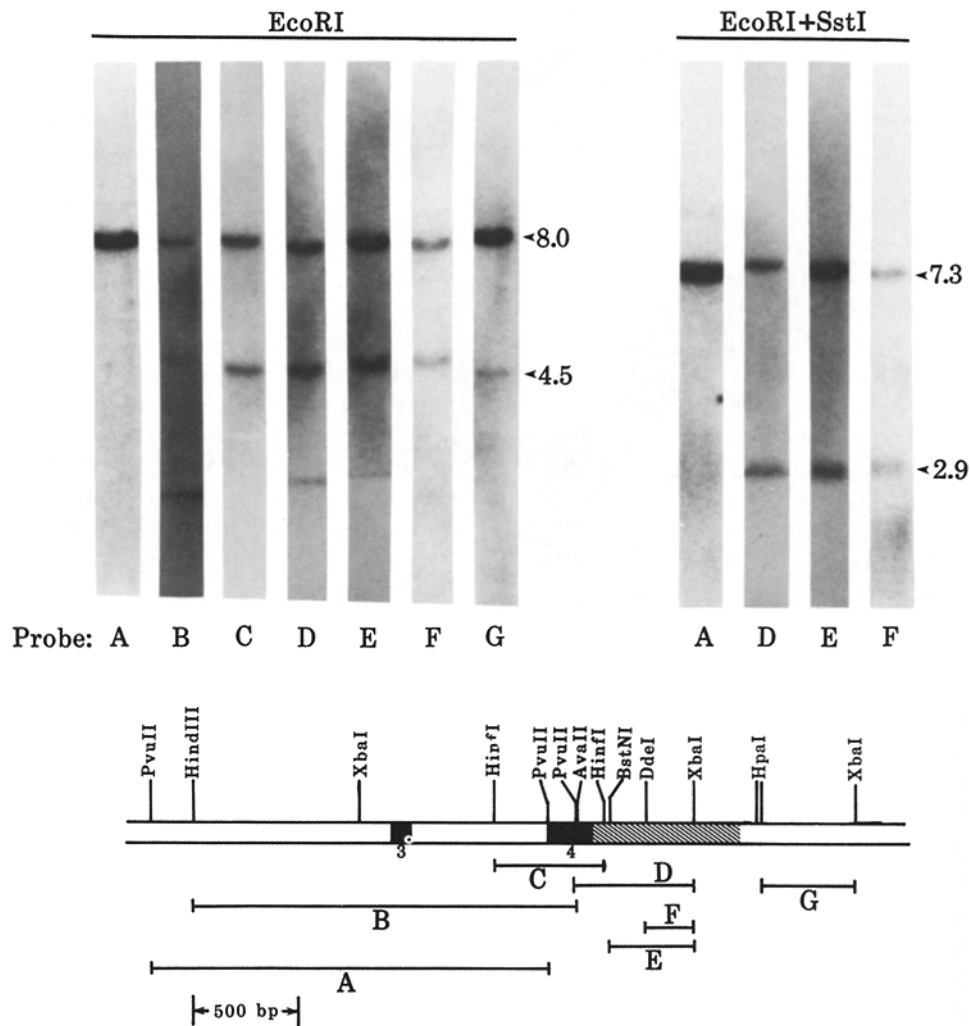
FIGURE 5.   An additional sequence in human DNA that hybridizes to exon four probes from the J chain gene. Human placental DNA was digested by the indicated enzymes, electrophoresed in agarose gels, transferred to nitrocellulose, and hybridized with probes derived from the cloned human J chain gene. Lanes from several experiments have been assembled for this figure; this explains slight differences in the migration of the bands. The probes, indicated below each lane, represent DNA segments from around exons three and four, as shown in the map at bottom. In the map, black rectangles represent the coding blocks; the hatched rectangle represents 3′ untranslated region. The cloned gene corresponds to the 8 kb Eco RI band and the 7.3 kb Eco RI/Sst I band seen in the blots. The 4.5 kb Eco RI band represents another sequence in the human genome that contains enough similarity with exon four of the J chain gene to hybridize with these probes. The sharp lines near the bottom of several lanes are edge artifacts resulting from cutting the blots.
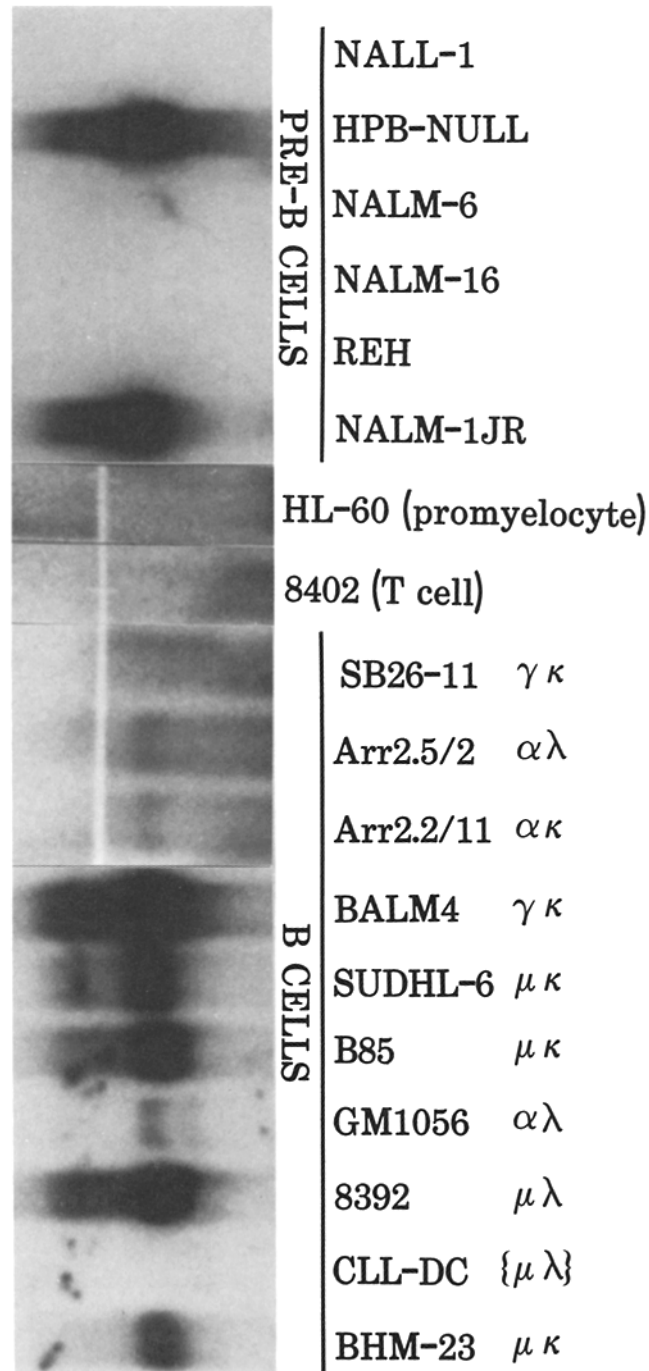
NALL-1
HPB-NULL
NALM-6
NALM-16
REH
NALM-1JR

PRE-B CELLS

HL-60 (promyelocyte)

8402 (T cell)

SB26-11 $\gamma \kappa$

Arr2.5/2 $\alpha \lambda$

Arr2.2/11 $\alpha \kappa$

BALM4 $\gamma \kappa$

SUDHL-6 $\mu \kappa$

B85 $\mu \kappa$

GM1056 $\alpha \lambda$

8392 $\mu \lambda$

CLL-DC $\{\mu \lambda\}$

BHM-23 $\mu \kappa$

B CELLS

FIGURE 6. J chain sequences in RNA from human cells. Total cellular RNA was used for Northern blots hybridized to the Pvu II/Xba I J chain probe. The cell lines are identified above each lane; the Ig isotypes synthesized by the B cell lines are indicated. The spots appearing in the NALM-6 lane and the six right-most lanes are believed to be hybridization artifacts; the faint band slightly above the position of the J chain mRNA in the GM1056 and CLL-DC lanes apparently represents residual binding to 18 S RNA, which migrates exactly at this position. In these experiments, an attempt was made to load consistent amounts of RNA on the gels and to follow standardized protocols for blotting, hybridization, etc. However, no attempt at quantitation has been made, nor have we determined the detection limits of our blotting technique in terms of numbers of RNA molecules per cell.

of the hybridizing RNA (estimated from the migration of transfer RNA, 18 S RNA, and 28 S RNA on the same gels) was ~1,500 nucleotides.

Discussion

The cloning of the human gene encoding the J chain protein has been accomplished, with some difficulty, owing to the apparent absence of the gene in several gene libraries. Failure of a gene to be represented in a given library may, theoretically, be explained by peculiar distribution of sites for the restriction endonuclease originally used to generate the DNA fragments from which the library was constructed, Mbo I in our case. The fact that all the libraries we screened were amplified is consistent with the alternative possibility that the gene is associated with sequences that handicap phage multiplication during the amplification step. Such hypothetical sequences did not, however, prevent the growth of the clones we isolated from size-fractionated restriction fragments.

The identification of the DNA segment contained in the overlapping clones HuJ1 and HuJ2 as deriving from a functional gene for the human J chain protein is based on the facts that (a) the cloned segment encodes an amino acid sequence nearly identical to that previously reported for this protein, and (b) Southern blots of human genomic DNA reveal no other candidate for a related gene that could encode the complete protein. The "extra" band of Fig. 5 apparently represents a sequence closely similar to exon four of the functional gene. It seems unlikely that this extra band represents a short, fortuitous sequence similarity without real homology, because the hybridization is so strong, and because it is detected by nonoverlapping probes that span more than 0.5 kb. The band is not likely to represent a "processed pseudogene" (28) because it does not hybridize to probe A, which contains all of exon three. One intriguing possibility is that the presence of exon four in both the J chain gene and in the extra band may represent an example of the evolutionary exon shuffling postulated by Gilbert (29) as an explanation for the splitting of eukaryotic genes by noncoding intron sequences. In the framework of this model, exon four may be functioning as an exon in another gene, whose remaining structure may be completely unrelated to the J chain gene. Alternatively, the extra band may represent a duplicated copy of the J chain gene that has undergone a substantial deletion. An evaluation of these possibilites awaits analysis of the extra hybridizing sequence.

An unexpected feature of the J chain gene was the existence of multiple potential polyadenylation signals 3′ of the final coding block. The sequence AATAAA was first recognized as a polyadenylation signal by its appearance near the 3′ end of many RNA sequences (30). Subsequently, it has been shown, in several systems, that mutations in this hexanucleotide can prevent normal cleavage and polyadenylation of RNA transcripts (31). AATAAA, or the variant AATTAAA found in several genes, is apparently necessary but not sufficient as a functional signal. That it is not sufficient is demonstrated in the human J chain gene by the presence of nine such sequences, only one of which appears to function as a true polyadenylation signal in BALM-4, as shown in Fig. 4. Multiple polyadenylation signals have been reported in several other genes in which, however, more than one of these signals is generally functional [e.g. mouse β_2-microglobulin (32), chicken pro-α2(1) collagen (33), mouse α-amylase (34), and

mouse dihydrofolate reductase (35)]. In the Ig heavy chain genes the function of alternate polyadenylation sites is specifically regulated during maturation of the cell (36); this raises the formal possibility that, in the human J chain system, different polyadenylation signals might be functional under some circumstances. The fact that the functional signal observed in BALM-4 is the only one of the nine candidates showing flanking homology with the single signal found in the murine J chain cDNA studied by Cann et al. (11) suggests that this flanking homologous segment may contain sequences that contribute to the signaling for cleavage and polyadenylation. This homologous segment is 5' of the AATTAAA found in both genes (Fig. 3); whether significant homology is also present further 3', in a sequence that (at the RNA level) is replaced by poly(A), is unknown at present because a genomic sequence of this region of the mouse gene is not yet available. Sequence comparisons of the 3' regions would be particularly interesting because experiments in several systems (e.g. bovine growth hormone [37], adenovirus E2A transcript [38], hepatitis B surface antigen [39], and SV40 late messages [M. Sadofsky and J. Alwine, personal communication]) suggest that sequences 20–60 basepairs 3' of the AATAAA are critical in determining the position of poly(A) addition. No evidence has yet been obtained for function of a sequence upstream of the AATAAA signaling the region where we find sequence conservation between mouse and human J chain genes, but a role for this region cannot at present be ruled out.

Developmental Program of J Chain Expression. The observation that several mouse lymphoma lines with characteristics of early B cells make no detectable J chain or J chain mRNA (2–5) has been taken to indicate that in mouse, J chain synthesis is a late-developing function in the maturation of the B lymphocyte. Some studies (40) have supported this notion for human B cells. Thus Mestecky et al. (40) did not detect J chain in either pre-B or B cells, but found that synthesis of J chain and Ig appeared simultaneously after pokeweed mitogen stimulation of B lymphocytes, as if expression of J chain and Ig were regulated in a linked manner. These results fit in an appealing way with the fact that the only known functions of the J chain involve interaction with Ig: the formation of Ig multimers (IgM pentamers and IgA dimers), and the facilitation of binding of these molecules to secretory component (1, 41). Thus, a simplistic expectation of J chain expression might predict synthesis of J chain only in lymphocytes requiring these functions (i.e., those secreting IgM or IgA). Consistent with this viewpoint, we found that a nonsecreting but IgM-bearing chronic lymphocytic leukemia cell lacked detectable J chain RNA, while all the IgM- or IgA-secreting B cell lines that we examined contained it.

However, accumulating evidence suggests that, in humans at least, J chain expression may not always be tightly linked to Ig expression. Mason and Stein (42), for example, identified intracellular J chain (by immunohistochemical staining) in three Ig⁻ human lymphomas, as well as in occasional Ig⁻ cells in reactive lymph node biopsies. McCune et al. (6) used in vitro translation of RNA and immunoprecipitation to document J chain expression in a pre-B-like cell line and two Ig⁻ round cell lines. Hajdu et al. (7) examined J chain by radioimmunoassay and immuno–electron microscopy, and suggested that J chain expression actually precedes μ synthesis in B lymphocyte maturation.

Our Nothern blot analyses show that, although four of the cells of pre-B phenotype we examined do not express detectable J RNA, two pre-B RNA samples contain abundant J chain sequences (Fig. 6). These two J$^+$ lines have been unequivocally characterized as pre-B-like: they both express CALLA, an antigen lost on most mature B cells; they both express cytoplasmic but not surface μ chains; and they both have their light chain genes in the germline configuration. These traits rule out the interpretation, which can be entertained for some other cases of J$^+$ Ig$^-$ cells, that they represent not pre-B cells, but mature B cells that have lost their capacity to synthesize antibody. The model of Hajdu et al. (7), suggesting that J chain expression precedes μ expression, was based on the finding of immunoreactive J chain in μ^- leukemic null cells, as well as in two μ^+ pre-B cell lines. Although we have not examined fresh leukemic null cells, the six pre-B lines that we used for Nothern blot analysis of J chain RNA do not support such a model. Given the absence of a J$^+$ μ^- line in our set, and the presence of a J$^-$ μ^+ cell (NALM-6), our data are more consistent with a model in which μ expression precedes J chain expression; but in consideration of the data of other investigators (7), it is probably most reasonable to conclude that no clear sequence of activation of these two genes has yet been established for human lymphoid cells, if indeed such an invariant sequence exists.

The observation of J chain expression in human pre-B lines is at variance with published results in the mouse. As indicated above, all results on murine lymphoid lines representing the pre-B or early unstimulated B cells are consistent with J chain being a late characteristic of mature, stimulated B cells (5). Conceivably, the malignant transformed cells that are generally studied (human or mouse) might not accurately reflect the phenotype of normal differentiating B lymphocytes. The process of transformation and maturation arrest could activate or repress specific gene expression, and such processes could operate differently between the two species. Alternatively, it is also possible that more extensive studies of mouse pre-B cells will reveal some J chain expression at stages earlier than has so far been observed. A most interesting possibility is that a real species difference may exist in the developmental program of gene expression of murine and human B lymphocytes, but more evidence will be necessary to substantiate this.

Of the mature B cells examined in the present study, J chain message was found in almost all. The absence of detectable J chain RNA in the nonsecretory, IgM-bearing CLL-DC is consistent with the phenotype of other early B cells such as WEHI 231 (43) and unstimulated murine B lymphocytes (5) reported from Koshland's laboratory, and the comparable human CLL and lymphoma material reported by Laurent et al. (44). These cells suggest that the transcription of the J chain gene may be turned off at certain periods within the mature but nonsecretory stage of B cell differentiation. The presence of J chain RNA in an IgG-secreting cell argues that the gene may continue to be transcribed after heavy chain class switching to Cγ regions, even though the J chain protein is not utilized in secreted IgG. This confirms similar observations at the protein level reported by several laboratories (40, and references therein).

The availability of a human J chain clone should allow the regulation of this gene during B cell differentiation to be more finely dissected.

Summary

As part of an ongoing investigation of the regulation of gene expression in B cell development, we have obtained a genomic DNA clone encoding the human J chain protein. The nucleotide sequence of exons encoding the mature protein defines a 137 amino acid primary sequence similar to that previously determined at the protein level. Probes from the gene have been used to analyze J chain expression in human cell lines corresponding to pre-B and B lymphocytes. J chain RNA was detected in two of six human pre-B cell lines and in 8 of 10 B cell lines expressing various Ig isotypes. The expression of the J chain gene is, thus, not tightly linked to IgM or IgA secretion. Our data do not, however, support the recent suggestion (7) that synthesis of J chain precedes that of μ chain in B lymphocyte differentiation. Because of the presence of nine candidate polyadenylation signals (AATAAA or AATTAAA) downstream of the C-terminal coding block of the J chain gene, the 3' end of the gene could not be determined from sequence data alone. To define the 3' end, J chain RNA from a human B lymphocyte line was used to protect an end-labelled DNA fragment from S1 nuclease digestion. The sequence 40 basepairs 5' of the functional polyadenylation site identified by these S1 experiments is homologous the same region of a previously reported mouse J chain complementary DNA clone.

References

1. Koshland, M. E. 1975. Structure and function of the J chain. *Adv. Immunol.* 20:41.
2. Mosmann, T. R., Y. Gravel, A. R. Williamson, and R. Baumal. 1978. Modification and fate of J chain in myeloma cells in the presence and absence of polymeric immunoglobulin secretion. *Eur. J. Immunol.* 8:94.
3. Mather, E. L., F. W. Alt, A. L. M. Bothwell, D. Baltimore, and E. Koshland. 1981. Expression of J chain RNA in cell lines representing different stages of B lymphocyte differentiation. *Cell.* 23:369.
4. Koshland, M. E. 1983. Presidential address: molecular aspects of B cell differentiation. *J. Immunol.* 131:i.
5. Lamson, G., and M. E. Koshland. 1984. Changes in J chain and μ chain RNA expression as a function of B cell differentiation. *J. Exp. Med.* 160:877.
6. McCune, J. M., S. M. Fu, and H. G. Kunkel. 1981. J chain biosynthesis in pre-B cells and other possible precursor B cells. *J. Exp. Med.* 154:138.
7. Hajdu, I., A. Moldoveanu, M. D. Cooper, and J. Mestecky. 1983. Ultrastructural studies of human lymphoid cells. μ and J chain expression as a function of B cell differentiation. *J. Exp. Med.* 158:1993.
8. Hohn, B. 1979. *In vitro* packaging of λ and cosmid DNA. *Methods Enzymol.* 68:199.
9. Korsmeyer, S. J., A. Arnold, A. Bakhshi, J. V. Ravetch, U. Siebenlist, P. A. Hieter, S. O. Sharrow, T. W. LeBien, J. H. Kersey, D. G. Poplack, P. Leder, and T. A. Waldmann. 1938. *J. Clin. Invest.* 71:301.
10. Grunstein, M., and D. Hogness. 1975. Colony hybridization: a method for the

isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA.* 72:3961.

11. Cann, G. M., A. Zaritsky, and M. E. Koshland. 1982. Primary structure of the immunoglobulin J chain from the mouse. *Proc. Natl. Acad. Sci. USA.* 79:6656.

12. Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499.

13. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.

14. Berk, A. J., and P. A. Sharp. 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. *Cell.* 12:721.

15. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: *A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY. 207–209.

16. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503.

17. Wahl, G. M., M. Stern, and G. R. Stark. 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethal-paper and rapid hybridization by using dextran sulfate. *Proc. Natl. Acad. Sci. USA.* 76:3683.

18. Bakhshi, A., J. Minowada, A. Arnold, J. Cossman, J. P. Jensen, J. Whang-Peng, T. Waldmann, and S. J. Korsmeyer. 1983. Lymphoid blast crises of chronic myelogenous leukemia represent stages in the development of B-cell precursors. *N. Engl. J. Med.* 309:826.

19. Collins, S. J., F. W. Rossetti, R. E. Callagher, and R. L. Gallot. 1978. Terminal differentiation of human promyelocytic leukemia cells induced by dimethyl sulfoxide and other polar compounds. *Proc. Natl. Acad. Sci. USA.* 75:2458.

20. Hieter, P. A., S. J. Korsmeyer, T. A. Waldmann, and P. Leder. 1981. Human immunoglobulin kappa light chain genes are deleted or rearranged in lambda producing B-cells. *Nature (Lond.).* 290:368.

21. Minowada, J. 1982. Immunology of leukemic cells. *In* Leukemia. F. Gunz and E. Henderson, editors. Grune & Stratton Inc., New York. 119–139.

22. Winter, J. N., D. Vaviakojis, and A. L. Epstein. 1984. Phenotypic analysis of established diffuse histocytic lymphoma cell lines utilizing monoclonal antibodies and cytochemical techniques. *Blood.* 63:140.

23. Lok, M. S., H. Kashiba, T. Han, S. Abe, J. Minowada, A. A. Sandberg. 1979. Establishment and characterization of human lymphocytic lymphoma cell lines (BALM-3, 4 and 5): intraclonal variation in the B cell differentiation stage. *Int. J. Cancer.* 24:S72.

24. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry.* 18:5294.

25. Mole, J. E., A. S. Bhown, and J. C. Bennett. 1977. Primary structure of human J chain: alignment of peptides from chemical and enzymatic hydrolyses. *Biochemistry.* 16:3507.

26. Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:460.

27. Schmid, C. W., and W. R. Jelinek. 1982. The Alu family of dispersed repetitive sequences. *Science (Wash. DC)* 216:1065.

28. Hollis, G. F., P. A. Hieter, O. W. McBride, D. Swan, and P. Leder. 1982. Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. *Nature (Lond.).* 296:321.

29. Gilbert, W. 1978. Why genes in pieces? *Nature (Lond.).* 271:501.

30. Proudfood, N. J., and G. G. Brownlee. 1976. 3' Non-coding region sequences in eukaryotic messenger RNA. *Nature (Lond.).* 263:211.

31. Fitzgerald, M., and T. Shenk. 1981. The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. *Cell.* 24:251.

32. Parnes, J. R., R. R. Robinson, and J. G. Seidman. 1983. Multiple mRNA sp :cies with distinct 3' termini are transcribed from the β_2-microglobulin gene. *Nat\t re (Lond.).* 302:449.

33. Aho, S., V. Tate, and H. Boedtker. 1983. Multiple 3' ends of the chic] .en pro $\alpha2$(I) collagen gene. *Nucleic Acids Res.* 11:5443.

34. Tosi, M., R. A. Young, Hagenbüchle, and U. Schibler. 1981. Multir.e polyadenylation sites in a mouse a-amylase gene. *Nucleic Acids Res.* 9:2313.

35. Setzer, D. R., M. McGrogan, and R. T. Schimke. 1982. Nucleot:de sequence surrounding multiple polyadenylation sites in the mouse dihydrofola ce reductase gene. *J. Biol. Chem.* 257:5143.

36. Early, P., J. Rogers, M. Davis, K. Calame, M. Bond, R. Wall., and L. Hood. 1980. Two mRNAs can be produced from a single immunoglobulin M gene by alternative RNA processing pathways. *Cell.* 20:313.

37. Woychik, R. P., R. H. Lyons, L. Post, and F. M. Rottman. 1984. Requirement for the 3' flanking region of the bovine growth hormone gene for accurate polyadenylylation. *Proc. Natl. Acad. Sci. USA.* 81:3944.

38. McDevitt, M. A., M. J. Imperiale, H. Ali, and J. R. Nevins. 1984. Requirement of a downstream sequence for generation of a poly(A) addition site. *Cell.* 37:993.

39. Simonsen, C. C., and A. D. Levinson. 1983. Analysis of processing and polyadenylation signals of the hepatitis B virus surface antigen gene by using simian virus 40-hepatitis B virus chimeric plasmids. *Mol. Cell. Biol.* 3:2250.

40. Mestecky, J., J.-L. Prud'homme, S. S. Crago, E. Mihaesco, J. T. Prchal, and A. J. Okos. 1980. Presence of J chain in human lymphoid cells. *Clin. Exp. Immunol.* 39:371.

41. Brandtzaeg, P., and H. Prydz. 1984. Direct evidence for an integrated function of J chain and secretory component in epithelial transport of immunoglobulins. *Nature (Lond.).* 311:71.

42. Mason, D. Y., and H. Stein. 1981. Reactive and neoplastic human lymphoid cells producing J chain in the absence of immunoglobulin: evidence for the existence of 'J chain disease'? *Clin. Exp. Immunol.* 46:305.

43. Raschke, W. C., E. L. Mather, and M. E. Koshland. 1984. Assembly and secretion of pentameric IgM in a fusion between a nonsecreting B cell lymphoma and an IgG-secreting plasmacytoma. *Proc. Natl. Acad. Sci. USA.* 76:3469.

44. Laurent, G., G. Delsol, F. Reyes, M. Abbal, and E. Mihaesco. 1981. Detection of J chain in lymphomas and related disorders. *Clin. Exp. Immunol.* 44:620.

45. Elliott, B. W., Jr., and L. A. Steiner. 1984. Amino- and carboxy-terminal sequence of mouse J chain and analysis of tryptic peptides. *J. Immunol.* 132:2968.

46. Maizel, J. V., Jr., and R. P. Lenk. 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. *Proc. Natl. Acad. Sci. USA.* 78:7665.