Structure and Expression of the Human Immunoglobulin \( \lambda \) Genes

By Thomas J. Vasicek and Philip Leder

From the Department of Genetics, Harvard Medical School; and the Howard Hughes Medical Research Institute, Boston, Massachusetts 02115

Summary

We determined the DNA sequence of two large regions of chromosome 22: 33.7 kb containing the \( \lambda \) complex; and 5.2 kb 5' of the functionally rearranged \( \lambda \) gene from the human myeloma, U266. Analysis of these sequences reveals the complete structure of the human \( \lambda \) complex and a previously undescribed seventh \( \lambda \) region that may encode the \( \text{Ke}^+\text{Oz}^- \lambda \) protein. The seven constant regions are organized in a tandem array, and each is preceded by a single \( J_\lambda \) region. \( \lambda_1, \lambda_2, \lambda_3, \) and \( \lambda_7 \) are apparently active genes, while \( \lambda_4, \lambda_5, \) and \( \lambda_6 \) are pseudogenes. There are no other \( J_\lambda \) or \( \lambda \) regions within a 60-kb region surrounding the \( \lambda \) complex; however, there are at least four other \( \lambda \)-like genes and \( \lambda \) pseudogenes in the human genome. The \( \lambda \) genes appear to have evolved via a series of gene duplication events resulting from unequal crossing over or gene conversion between the highly conserved \( \lambda \) regions on mispaired chromosomes. The lack of Alu sequences in this large segment of DNA suggests that the \( \lambda \) complex resulted from a recent amplification of a smaller Alu-free segment of DNA. Illegitimate recombination between repeated sequences containing \( \lambda_2 \) and \( \lambda_3 \) may be responsible for variable amplification of the \( \lambda \) genes. We also found a 1,377-bp open reading frame (ORF) located on the opposite strand in the region containing \( \lambda_7 \). While this ORF is flanked by potential RNA splicing signals, we have no evidence that it is part of a functional gene. We also discovered a \( V_\lambda \) pseudogene, called \( \psi V_\lambda 1, \) 3 kb upstream of the U266\( \lambda \) gene. Using primer extension analysis to map the transcription start in the human \( \lambda \) gene, we have identified its initiation point 41 bp upstream of the initiation codon. Analysis of the \( \lambda \) promoter reveals that it contains a TATAA box at position -29 relative to the transcription initiation site and an octamer sequence at -67. Computer analysis of 40 kb of DNA sequences surrounding the human \( \lambda \) locus has revealed no sequences resembling the \( \kappa \) or IgH transcriptional enhancers, nor have in vitro analyses for function revealed enhancer activity. A comparison of these results with those obtained in separate studies with transgenic mice point to a complex, developmentally linked mechanism of transcriptional activation.

The Ig genes are among the most intensely studied and clinically relevant loci in man. Ig proteins are encoded at three independent loci: IgH, IgK, and Ig\( \lambda \). Each locus consists of a complex of \( V, J, \) and \( C \) gene segments, along with diversity segments in the IgH locus, which require rearrangement during B cell development to produce active genes (for review, see reference 1). The detailed structures of the human \( \kappa \) and IgH genes have been determined, and control sequences such as transcriptional enhancers and conserved promoter elements required for Ig gene activation have been delineated. Because of the importance of the Ig genes, we have set out to characterize the third of these loci, the human \( \lambda \) locus.

\( \lambda \) L chains are present in \( \sim 40\% \) of human IgGs (2). Four \( \lambda \) isotypes, termed Mcg*, Ke^+Oz^-, Ke^-Oz^-, and Ke^-Oz^-, have been described on the basis of their reactivity with the Oz (3), Kern (4), and Mcg (5) antisera, which were raised against \( \lambda \) Bence Jones proteins isolated from patients with multiple myeloma (6). The Oz, Kern, and Mcg antisera bind to specific epitopes on the \( C_\lambda \) regions that result from specific amino acid substitutions. In addition to the four well-known \( \lambda \) isotypes, peptide sequence analysis of human \( \lambda \) proteins reveals 10 additional, distinct \( C_\lambda \) sequences (7). Whether these unique \( \lambda \) proteins represent as yet uncharacterized \( C_\lambda \) genes or discrete polymorphic alleles of known genes has yet to be determined. This contrasts with the human \( \kappa \) gene, which is represented as a single \( C_\kappa \) region sequence with two polymorphic alleles.

In their initial characterization of the human \( \lambda \) locus, Hieter et al. (8) cloned and partially characterized the \( C_\lambda \) gene complex and predicted that the human \( \lambda \) locus contained a long tandem array of at least six \( C_\lambda \) regions spanning \( \sim 30 \) kb of DNA. The DNA sequence of the first three \( C_\lambda \) regions (numbered from 5' to 3') revealed that they potentially encoded the Mcg*, Ke^+Oz^-, and Ke^-Oz^+ polypeptides,
Materials and Methods

All coding region and pseudogene sequences were confirmed by Promega Biotech (Madison, WI) (T7 and Sp6 promoter primers). England Biolabs (Beverly, MA) (pBR322 and pUC primers) or Applied Biosystems Inc., Foster City, CA) or purchased from New England Biolabs. Primers were either synthesized on a DNA synthesizer (380A; Applied Biosystems Inc., Foster City, CA) using Sequenase (United States Biochemical Corp., Cleveland, OH). DNA was performed on double-stranded plasmids, cosmids, and λ phage using Sequenase (United States Biochemical Corp., Cleveland, OH). Primers were either synthesized on a DNA synthesizer (380A; Applied Biosystems Inc., Foster City, CA) or purchased from New England Biolabs (Beverly, MA) (pBR322 and pUC primers) or Promega Biotech (Madison, WI) (T7 and Sp6 promoter primers). All coding region and pseudogene sequences were confirmed by sequencing on both strands with dITP in the place of dGTP (19), as were most of the intronic and intergenic regions, which were sequenced on only one strand. Sequences were analyzed on 4.0 or 5.0%, denaturing, electrolyte gradient (20) gels, 100 cm long by 0.4 mm thick, run in an Epphorteck (Haake/Buchler, Saddle Brook, NJ) gel apparatus. This arrangement allowed resolution of 600–850 bases from each primer.

**Primer Extension Analysis.** An antisense strand deoxyoligonucleotide primer, 5'CTGTGCCCTG AGTGAGGAGG GTGAG-GATGA-3', complementary to the 3'-most bases of the first exon (the V leader coding exon; see Fig. 9) was end labeled with T4 polynucleotide kinase and used to prime cDNA synthesis with AMV reverse transcriptase (Life Sciences, Tampa, FL) on total cellular RNA (15) from U266 myeloma cells. The resulting cDNAs were subjected to electrophoresis on an 8% sequencing gel with adjacent lanes containing chain termination sequencing reactions.

**Computer-aided DNA Sequence Analysis.** Computer analyses were performed using the University of Wisconsin Genetics Computer Group (UWGCG) software (21), except for the evolutionary tree, which was generated by the progressive alignment program described by Feng and Doolittle (22).

**Results**

**DNA Sequence of the C Complex.** To define the structural features of the C complex and establish the mechanism for its evolution, as well as search for potential enhancer sequences, we determined the nucleotide sequence of the entire C complex. The DNA sequencing strategy for the C complex is shown in Fig. 1. We subcloned each of the λ genes separately and performed plasmid-directed, chain termination sequencing reactions with primers that specifically hybridize to either the J or the C regions to determine the sequence of large regions surrounding each J and C gene. In addition, we used commercially available plasmid primers to sequence inward from the ends of each cloned fragment and obtained the complete sequence of each plasmid insert by extending the known regions with additional synthetic primers. We sequenced across the subclone boundaries and small EcoRI restriction fragments by priming directly on fragments cloned in cosmids and λ phage. In all, 42.2 kb of DNA were sequenced, largely on both strands.

**Structure of the Human C Complex.** Analysis of the complete sequence of human C complex reveals that it consists of a tandem array of seven C genes, each preceded by a single J region (Fig. 1). We determined that no additional λ genes exist within 10 kb 5' of J1 or within 17 kb 3' of C7 by probing Southern blots of cloned DNA with labeled J and C fragments (data not shown). Southern blot analysis with J and C probes and partial DNA sequence analysis of the C complex led previous investigators to propose that the λ locus contained a long array of six J-C gene pairs arranged in tandem (6, 9). X may have been missed in previous studies because the distance between λ' 7 and λ 7 is much smaller than that between the other λ genes. The distance between each C region polyadenylation site and the next J region is as follows: C1 to J2, 3,800 bp; C2 to J3, 3,600 bp; C3 to J4, 3,780 bp; C4 to J5, 3,080 bp; C5 to J6, 2,110 bp; C6 to J7, 1,420 bp. In addition, the J-C intron length decreases from 1,540 bp in λ1 to

---

Abbreviation used in this paper: ORF, open reading frame.
1,145 bp in \( \lambda_7 \); \( \lambda_2 \), \( \lambda_3 \), \( \lambda_5 \), and \( \lambda_6 \) are all \( \sim 1,310 \text{ bp} \). Furthermore, the \( J_6 \), region in \( \Lambda \) was not previously recognized, probably because of its unusual location. \( \Lambda_4 \) contains a large deletion beginning 220 bp 3' from \( J_4 \) and ending at bp 65 of the \( C_4 \) region. This deletion removes most of the intron, placing \( J_4 \) anomalously close to \( C_4 \). While a computer search of the entire \( C_4 \) complex sequence reveals no additional \( J_\lambda \) or \( C_\lambda \) regions, the similarity between the \( \lambda \) genes extends beyond the coding sequences (see below).

**Sequence Conservation Flanking the \( J_\lambda \) and \( C_\lambda \) Genes.** We performed a dot matrix comparison of the \( C_\lambda \) complex sequence to itself (Fig. 2) to determine the extent of identity between the seven \( \lambda \) genes. Direct repeats within the sequence are revealed by diagonals offset from the identity line, but parallel to it. The longest diagonals are partially due to two direct repeats \( \sim 3 \text{ kb} \) long in which are embedded \( \lambda_2 \) and \( \lambda_3 \). These repeated sequences are shown as the shaded boxes in Fig. 1 and share \( >95\% \) identity over their entire lengths. The repeats extend through the \( J_\lambda-C_\lambda \) intron and 1.4 kb 5' of the \( J_\lambda \) regions. An additional kilobase 5' of the \( J_\lambda \) regions is \( >85\% \) identical, but 3' of the \( C_\lambda \) regions, the sequences diverge rapidly and are \( <60\% \) identical.

Comparison of the remaining \( J_\lambda \) and \( C_\lambda \) regions on the dot matrix results in very short diagonals except in a few cases. The plot shows that sequences 5' of \( J_2, J_3, J_4, \) and \( J_5, 3' \) of \( C_5 \) and \( C_6 \), and in the introns of \( \lambda_6 \) and \( \lambda_7 \) share persistent identity extending well beyond the coding regions.

Notably, computer searches of the sequenced regions of

![Figure 1. Structure and organization of the human CA complex. The \( J_\lambda \) and \( C_\lambda \) exons, labeled \( \lambda_1-\lambda_7 \), are shown as black boxes; the ragged left edge on \( C_\lambda \) indicates that a portion of the \( C \) region is deleted (see text). The dotted line indicates the physical continuity of the locus. Restriction sites shown are BglI, Bg; EcoRI, E; and HindIII, H. Horizontal arrows beneath the restriction map depict some of the individual sequencing gel runs. Representative plasmids, \( \lambda \) phages, and one cosmid are shown beneath the ruler, which is labeled 0-34 kb. The numbers in parentheses indicate the total length of the clone inserts. The shaded boxes at \( \lambda_2 \) and \( \lambda_3 \) mark direct repeats that share \( >95\% \) identity. The box under \( \lambda_7 \) shows the location, and the white arrow indicates the orientation, of a long ORF in the opposite orientation from the \( \lambda \) genes. These sequence data have been submitted to the EMBL GenBank Data Libraries under the accession number X51755.](image1.png)
the λ locus reveal no Alu repeat sequences. On average, Alu sequences occur every 5–6 kb in the human genome (23), and we might expect six such sequences within the λ locus. Alu sequences have been implicated in genomic rearrangements at other loci (24), however, it appears that they have not been involved with the duplications observed in the λ locus. Indeed, lack of Alu sequences in a region this large suggests that it arose from recent amplification of a sequence that itself lacked an Alu sequence.

Distinguishing between Functional and Pseudogenes. Having precisely located all of the remaining J3 and C3 genes in the λ complex, we set out to determine which of them were functional. All active Ig J regions are preceded by conserved nonamer (G2T3GT) and heptamer (CACTGTTG) sequences separated by either 12 or 23 unconserved base pairs. The reverse complement of these sequences (CACTGTTG and ACAAAC) separated by 23 or 12 bp is found downstream of each corresponding V region (or D in the IgH locus) (1). These recombinational signals are necessary and sufficient for accurate V-J rearrangements in normal pre-B cells (25, 26). Mutation of the recombinational signals is one potential mechanism for inactivating an Ig gene. Fig. 3 A shows that each human J3 has an intact copy of the consensus nonamer and heptamer sequences separated by 12 bases. The recombinational signals in λ2 and λ3 exactly match the consensus, and the substitutions present in the others are known to have no effect on activity (26). Thus, V-J rearrangements with all of these J3 regions are theoretically possible.

Another potential mechanism for inactivating a gene is mutation of the RNA processing signals for polyadenylation or splicing. Each of the seven λ genes has at least one copy of the consensus polyadenylation site (AATAAA) in its 3' untranslated region (Fig. 3 B). Two polyadenylation sites are present in λ2, λ3, λ5, and λ6, but the 5' site is mutated in λ1 and λ4, and is completely removed from λ7 by a 20-bp deletion. The other signals critical to RNA processing are those required for splicing. Nearly all eukaryotic introns begin with the AGGT/AGT splice donor and end with (T/C)NCAGG splice acceptor sequences (27). However, only λ1, λ2, λ3, and λ7 contain these conserved mRNA splice signals (Fig. 3 C). Note that the 5' splice donors in λ4, λ5, and λ6 all lack the GT dinucleotide, and the deletion in the λ4 gene completely eliminates the 3' splice acceptor sequence (the underlined GT and AG dinucleotides are invariant and define the beginning and the end of the intron, respectively). Thus, it is unlikely that transcripts from λ4, λ5, and λ6 could be appropriately spliced.

The Coding Regions: λ7 Encodes the Ke+Oz− C Region. Only λ1, λ2, λ3, and λ7 can direct synthesis of a functional mRNA containing a complete Jx-Cx ORF. Hieter et al. (8) showed that λ1, λ2, and λ3 could encode the Mcg, Ke− Oz−, and Ke− Oz+ proteins, but could not determine which of the remaining λ genes encoded the Ke− Oz− protein. The nucleotide sequence data presented in Figs. 3 and 4 indicate that the newly discovered λ7 has all the hallmarks of a functional λ gene segment. It has the consensus signals for V-J rearrangement, mRNA splicing, and polyadenylation, as well as an open reading frame that could encode a protein much like the published Ke− Oz− λ chain (28). It has been shown that the Gly residue at position 152 is required for reactivity with the Kern antiserum and, indeed, λ7 contains a Gly at this position (Fig. 5). However, λ7 differs from published amino acid sequences of λ proteins at three other positions: 157 (Val), 195 (Arg), and 212 (Ala). To determine whether this discrepancy is due to a cloning artifact or polymorphism, we cloned and sequenced this same region from two additional individuals and found them to be identical to our original sequence (data not shown). Therefore, the possibilities remain that the published Ke− Oz− amino acid sequence is encoded on an as yet uncharacterized gene, that additional, rare polymorphic forms of the characterized λ genes exist in the human population, or that the published Ke− Oz− amino acid sequence is incorrect at these positions.

λ4, λ5, and λ6 Are Pseudogenes. Multiple defects in λ4, λ5, and λ6 render them incapable of coding for functional λ proteins. Most dramatically, λ4 contains a deletion of ~1,150 bp with respect to λ3 and λ5. Fig. 1 shows that the J3 and C3 regions are much closer together in λ4 than in the other genes. The bases missing from λ4, make up most

Figure 3. Comparison of flanking sequences surrounding the J3 and C3 regions. The complete sequences for λ2 are shown. Identities in the other genes are represented by dashes, and deletions are represented by asterisks. The amino acids are numbered as in Kabat et al. (40). (A) Heptamer-nonamer recombinational signals 5' of the J3 regions. (B) 3' flanking sequences showing the translation terminator (ter) and the AATAAA polyadenylation signals. (C) Splice donor and acceptor sequences in the J3-C3 intron. Add 60 to the numbers between the arrows to calculate the total intron lengths.
of the intron and the 3' mRNA splice site (see the line 4 in Fig. 3 C), plus 64 bp of Cβ coding sequence in the λ3 gene (see Fig. 4, ψλ4, 3' of the Jβ). In addition, a three-base deletion at positions 226-228 in A4 eliminates one codon, and a single-base deletion at position 270 causes a framesshift. λ5 cannot code for a functional protein due to a deletion of 11 bp in Cα5 (positions 151-161; Fig. 4), resulting in a shift to a missense frame that would not be terminated until the polyadenylation site. λ6 has three framesshift mutations in its coding sequences, and Jβ6 has a four-base deletion that would result in premature termination four codons later. Moreover, Cβ6 also contains a four-base duplication at position 247 and a single-base insertion at position 349. Therefore, while λ4, λ5, and λ6 have the appropriate signals for Vβ-Jβ recombination, rearrangements with these genes cannot produce functional proteins.

**Evolution of the Ig λ Genes.** While the λ evolutionary tree and the structure of the mouse λ genes (Fig. 6, A and C) suggest that the mouse λ1-λ4 and λ1-λ5 clusters arose by duplication of an ancestral cluster (29), similar description of the origins of the human λ genes is complicated by intergenic exchanges that appear to have taken place among them. This effect is revealed on examination of the sequences shown in Figs. 3 and 4. For example, as shown in Fig. 4, Cα3 and Cα7 are identical from nucleotide positions 348 and 349. The Jβ regions are enclosed in the box.

**Figure 5.** Amino acid differences between the λ genes. The residues are numbered as Kabat et al. (40), and only the differences that determine the isotype of the active genes and those that occur in more than one gene are shown. The sequences are arranged as in Fig. 4 to place the most similar sequences closest together. The predicted reactivity of λ1, λ2, λ3, and λ7 to the Mcg, Ke, and Oz antisera are shown on the left, and the relevant amino acids are boxed.
from Cα3. It is possible that there is differential representation of these exchanges at the level of alleles. For example, a similar exchange has occurred in certain alleles of the human Ig Cα1 and Cα2 genes (30). The results of Udey and Blumberg (31) also suggest that exchanges are present in some alleles of the Cα complex and not in others.

A Long ORF Overlaps Cα7. In addition to the λ C region coding exons, several ORFs were found in the computer analysis of the Cα complex sequence using the UWGCG Frames program. One strikingly long ORF (1,377 bp) is encoded in the region of Cα7 on the strand opposite that encoding the λ genes (Fig. 1). The probability that a 458-codon ORF would occur in a random DNA sequence ([(61/64)^458] = 2.8 × 10^-10), is very small. While there are several ATG codons within this ORF (Fig. 7), none match the initiator consensus (32) and, therefore, are unlikely to be capable of efficient translation initiation. However, there are several consensus splice acceptor sequences near the beginning of this ORF, as well as splice donor sequences near its end, suggesting that it could be spliced to other exons at either or both ends. The 2 kb we have sequenced upstream of this ORF contains no sequences resembling transcriptional promoters; moreover, there are no polyadenylation sites within 10 kb downstream of this region. Comparison of the deduced amino acid sequence of the long ORF to the sequence data banks reveals no similarities to published sequences. To determine whether this region is transcribed, we analyzed RNA purified from fetal liver, adult PBL, three Burkitt Lymphoma cell lines, and a myeloma cell line in an RNase protection assay (data not shown). No protected fragments were detected using 40-μg samples of total cellular RNA from these tissues (the lower limit of detection is 0.1 pg of a 700-base fragment). Thus, we were unable to detect transcripts from this ORF in the tissues analyzed, but it may be transcribed in some other tissue, or in these tissues at a different developmental stage. While we cannot demonstrate that this ORF is part of an active gene, its discovery remains an intriguing observation.

The Search for a Potential Human λ Enhancer. Although cis-acting enhancer elements have been identified for the κ and IgH genes, such regions have not been described in the human λ locus. To initiate a search for the potential human λ enhancer, we isolated the active, rearranged λ gene (Fig. 8) from the U266 cell line (17). Using this clone, we attempted to demonstrate tissue-specific expression when the gene was transfected into a variety of cultured lymphoid cells. However, we found, as others have reported, that cloned λ genes direct very weak transcription in transfected and retrovirus-infected tissue culture cells (33, 34; Vaisceck, et al., manuscript in preparation). The inability to detect substantial transcriptional activity from the human λ gene in tissue culture systems suggests that either sequences necessary for its expression are simply not present on the large clone or that λ expression cannot be achieved using DNA transfection. In studies to be published elsewhere (Vaisceck, et al., manuscript in preparation), we have shown that this cloned, rearranged λ gene is expressed in a tissue-specific manner in transgenic mice, suggesting a more complex regulatory sequence

![Diagram](image-url)
involving activation of the gene in B cell precursors (see Discussion).

Bich-Thuy and Queen (35) have reported that a 4-kb DNA segment 3' of the mouse \( \lambda 1 \) gene (see Fig. 6 C) can activate transcription of a reporter gene construct, the \( X \) promoter driving the bacterial chloramphenicol acetyl transferase gene, in B cells expressing their endogenous \( X \) genes. We have not found similar activity in analogous regions from the human \( X \) locus, where we have tested fragments including as much as 2.1 kb 3' of \( X7 \) (see Discussion).

In further experiments, we analyzed DNA fragments from DNase I-treated U266 nuclei to investigate the chromatin structure in the vicinity of the active \( \lambda \) locus. DNase I-sensitive chromatin regions have been shown to correlate strongly with regions having cis-regulatory function (36, 37). The chromatin in the vicinity of DNase I hypersensitive sites, found in association with enhancer sequences in the \( X \) locus and other genes, is apparently in an "open configuration" either to promote the binding of transcription factors or because such factors are bound. Several DNase I-hypersensitive sites were found in the vicinity of the functionally rearranged U266 \( \lambda \) gene (data not shown). Because we could not demonstrate

Figure 7. DNA sequence of the long ORF overlapping \( C_{\beta} \). The reverse complement of the \( C_{\beta} \) region is underlined and its ends are marked. Potential splice acceptor and donor sequences (27) are also underlined, and the first and last base of the exons that would result from splicing into these sequences are marked with vertical lines.

Figure 8. Restriction map and sequencing strategy for the active \( \lambda \) gene from the U266 human myeloma cell line. \( V \) regions are shaded, and the \( J \) and \( C \) regions are black. The rearranged \( V \) region, U266A, is shown rearranged into the \( J \) region of \( X2 \), and a pseudogene, OVX1, is shown 3.7 kb 5' of the active gene. The restriction sites shown are EcoRI, E; XbaI, X; BamHI, B; and BglII, Bg. The short arrows indicate some of the individual sequencing gel runs, and the scale in kilobases is shown by the ruler on the bottom. These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number X51754.
expression of λ genes containing these regions, we analyzed the DNA sequence of the λ promoter to detect sequence patterns that would indicate the presence of potential enhancer sequences. We have performed extensive computer searches for enhancer-like sequences in DNA surrounding the Cλ complex and the functionally rearranged U266λ gene. While, as expected in a sequence of this size, these searches have revealed numerous enhancer core-like and potential transcription factor binding sites, their relevance to expression, if any, is difficult to evaluate.

**Mapping of the λ Transcription Start Site.** While the mechanism of λ gene activation is still unknown, some features of the λ promoter are similar to analogous regions in the κ and IgH loci. The λ promoter contains the conserved octamer sequence, ATTTGCGAT, and a TATAA box (Fig. 9), often GATAA in λ genes (38, 39). To map the exact transcription start site in the U266λ gene, we annealed synthetic oligonucleotide primers complementary to the leader coding region of the λ message with total cellular RNA from U266 cells and used AMV reverse transcriptase to synthesize cDNA molecules corresponding to the 5' end of the λ mRNA. Fig. 10 shows a DNA sequencing gel with adjacent lanes containing the cDNA and a DNA sequencing ladder. These data indicate that the U266λ gene transcript begins at a G nucleotide 29 bp downstream of the TATAA box and 41 bp upstream of the translation initiation codon (see Fig. 9).

**A λV Pseudogene 3.7 kb Upstream of the Active Gene.** While searching for enhancer sequences in the λ promoter region, we discovered a λ pseudogene, ψVλ1, 3.7 kb upstream of U266λ (Fig. 8). The DNA sequences of ψVλ1 and U266λ are compared in Fig. 9. ψVλ1 has most of the sequence characteristics of active genes: upstream octamer and TATAA (GATAA) sequences, conserved mRNA splice donor and acceptor sequences, and appropriately spaced heptamer and nonamer sequences for Vλ-Jλ rearrangement. However, it also has a 10-bp deletion, from position 135 to 144 in the Vλ-coding region, that would render it inactive. Interestingly, ψVλ1 and U266λ are members of different Vλ region families. The U266λ protein is 94% similar to the Mcg Vλ region, a subgroup II protein, and the predicted amino acid sequence of the hypothetical ψVλ1 protein is 83% similar to the DEL Vλ region of subgroup III (40). These two

**Figure 9.** Comparison of the U266λ and ψVλ1 V region sequences. The entire sequence of U266λ is shown but only the differences in ψVλ1 are shown. The deduced amino acid sequence, numbered as in Kabat et al. (40), of the active gene is indicated below the DNA sequences. The octamer sequence 67 bp upstream of the transcription start site, and the "TATAA" box (GATAA in λ genes) 28 bp upstream, are doubly underlined. The transcription start is indicated with an arrow 41 bp upstream of the translation initiation codon. The heptamer and nonamer Vλ-Jλ rearrangement sequences are underlined in the pseudogene sequence and the Jλ2 sequences, into which the active gene is rearranged, are shown.
genes are adjacent and yet they share a nucleotide similarity of only 66%.

Discussion

The complexity of the \( \lambda \) proteins is reflected in the complexity of the \( \lambda \) gene family. While the \( \lambda \) genes share a number of structural similarities with the IgH and \( \kappa \) genes, they exhibit differences as well. The human \( \lambda \) locus consists of many \( V_\lambda \) regions, as do the \( \kappa \) and IgH loci, but, unlike the other Ig loci, it encodes a cluster of \( J_\lambda-C_\lambda \) pairs. Although the mouse \( \lambda \) locus has only two functional \( V_\lambda \) regions, it also contains three functional \( J_\lambda-C_\lambda \) pairs. This conservation of multiple \( J_\lambda-C_\lambda \) pairs over evolutionary time, also seen in the shark (41), suggests that it has an important role in the immune response. Obviously, these multiple genes contribute to \( L \) chain diversity, a function that is achieved with more economy by multiple \( J \) regions in the \( \kappa \) and IgH loci in the absence of duplicated \( C \) regions.

The structure of the \( C_\lambda \) complex allows us to assign the four \( \lambda \) isotypes to genes within the main \( \lambda \) complex. This complex contains a tandem array of seven \( J_\lambda-C_\lambda \) gene pairs that we call \( \lambda 1-\lambda 7 \), 5' to 3'. In accordance with previous studies (8), we find that the first three genes are functional and code for the most common isotypic forms of the \( C_\lambda \) region, \( Mc\kappa, Ke^-Oz^- \), and \( Ke^-Oz^+ \), respectively. We also find a new \( \lambda \) gene, \( \lambda 7 \), that appears to be functional and is capable of producing a protein similar to the \( Ke^-Oz^- \) chain that is present in all human sera (4). We have found that \( \lambda 4, \lambda 5 \), and \( \lambda 6 \) are pseudogenes, in contrast to the report by Dariavach et al. (9), in which they describe \( \lambda 6 \) as the \( Ke^-Oz^- \) gene. It is possible, though unlikely, that this conflict results from a polymorphism. The sequence of \( C_\lambda \)6 presented by Dariavach et al. (9) resembles a combination of our sequences for \( C_\lambda \)6 and \( C_\lambda \)7. Such a hybrid gene may have resulted from recombination due to unequal crossing over between \( \lambda 6 \) and \( \lambda 7 \), deleting the intervening 3 kb. However, we have found three alleles of \( \lambda 6 \) and \( \lambda 7 \) to contain sequences identical to those shown in Figs. 3 and 4 (data not shown). Furthermore, Taub et al. (42) examined >100 individuals by Southern blot without detecting any alleles containing such a 3-kb deletion. Thus, the result of Dariavach et al. (9) may be due to cloning artifacts or sequencing errors, or they may have cloned a rare, deleted allele.

In addition to the seven \( \lambda \) genes in the \( C_\lambda \) complex, several unlinked DNA fragments crosshybridize with \( C_\lambda \) probes (8) (Fig. 6, A and B). Two of these fragments contain pseudogenes; one, \( \psi \lambda 1 \), is a processed pseudogene (12), located on chromosome 18 (O.W. McBride, personal communication) and the other, which has not been mapped to a specific chromosomal location, contains a defective \( C_\lambda \) region called \( \psi \lambda 18.1 \) (13). Two other dispersed \( \lambda \) sequences, called \( \lambda 14.1 \) and \( \lambda 16.1 \), may be functional genes; they are closely related to the \( \lambda \) genes in the main \( \lambda \) complex and share >86% nucleotide sequence similarity with the active Ig \( \lambda \) genes (13). Hollis et al. (43) showed that the \( \lambda 14.1 \) gene, and perhaps the \( \lambda 16.1 \) gene, are transcribed in pre-B cells. The structure of the \( \lambda 14.1/\lambda 16.1 \) genes, and their expression pattern, led Hollis et al. (43) to suggest that they are human analogues of the mouse \( \lambda 5 \) gene (Fig. 6 C). L chain–like proteins that associate with H chains in human (44) and mouse (45) pre-B cells may be encoded by the \( \lambda 14.1 \) and \( \lambda 5 \) genes, respectively.

The \( \lambda \) gene family appears increasingly complex; for, in addition to the seven \( \lambda \) genes in the \( C_\lambda \) complex we have described, three polymorphic forms of the \( \lambda \) locus may encode distinct \( \lambda \) proteins (42). These polymorphic alleles apparently result from duplication of the region containing \( \lambda 2 \) and \( \lambda 3 \) and concurrently increase the size of the 8.4-kb EcoR1 fragment on which they are normally found. Alleles carrying these polymorphic forms appear to have one, two, or three additional \( \lambda \) genes that have not yet been characterized in detail, though they are likely to be closely related to \( \lambda 2 \) and \( \lambda 3 \), as they share identical restriction maps. This amplification of the \( \lambda \) genes appears to be due to duplication of a 5.4-kb region containing \( \lambda 2 \) and \( \lambda 3 \) (Fig. 1). \( \lambda 2 \) and \( \lambda 3 \) are themselves located within two repeated sequences that share >95% sequence identity for 3.2 kb containing \( J_\lambda-C_\lambda 2 \) and \( J_\lambda-C_\lambda 3 \) (Fig. 1, shaded boxes); an additional 1.0 kb 5' of these regions share >85% identity. Since analogous positions on these repeats are 5.4 kb apart, mispairing between two chromosomes 22 and unequal crossing over or sister chromatid exchange could result in additional copies of the repeat unit. The resulting duplication would yield a 5.4-kb increase in the distance between the two EcoR1 sites flanking \( \lambda 2 \) and \( \lambda 3 \) at the expense of the sister chromatid. Both unequal crossing over, and precise excision of the direct repeats, could also result in the loss of \( \lambda \) gene copies, but \( \lambda \) loci with fewer than two genes in this region have not been found; this suggests that there may be specific selection against such alleles in the population. In any case, individuals in the population could have four to seven potentially active \( \lambda \) genes per haploid genome.

While \( \lambda 2 \) and \( \lambda 3 \) are the most similar of the \( \lambda \) genes, and the similarity between them is the most extensive, the other five \( \lambda \) genes also share a very high degree of DNA sequence similarity extending beyond the coding regions (Fig. 2). The similarity among the genes suggests a role for recent gene duplication events and/or exchange of genetic information by gene conversion. The lack of Alu sequences within the locus is consistent with this interpretation.

While the \( \lambda \) locus has a number of gross structural differ-
quences from the κ and IgH loci, it also differs in the more subtle structure of its transcriptional control elements. The IgH gene has a powerful, position- and orientation-independent, tissue-specific enhancer element in the Jα-Cα intron (46). The κ locus has a similar intron enhancer (47), although its activity appears to be not as strong.

To develop a functional assay for the mechanisms of transcriptional control of the λ genes, we cloned the active λ gene (rarranged into λ2) from the U266 human myeloma cell line. We and others (33, 34, 48) have found that cloned λ genes are only weakly transcribed in transfected tissue culture cells, even in cells that are actively expressing endogenous λ genes. Furthermore, Neuberger et al. (49) and Hagman et al. (50) found that expression of certain mouse λ clones in transgenic mice required attachment of IgH chain enhancer sequences. However, we have found that the 8.5-kb U266X clone, with 5.2 kb of 5' flanking sequences, is specifically transcribed at high levels in the lymphoid tissues of three transgenic mouse lines, without addition of heterologous enhancer sequences (Vasicek, et al., manuscript in preparation). This result suggests that sequences required for tissue-specific activation of the human λ gene in transgenic mice are located within the 8.5-kb U266X clone. This does not rule out the possibility that additional enhancer elements might be located further outside of the structural genes, but it suggests that, to become activated, the gene must be present in the cell during some stage of development. Such conditions are obviously difficult, if not impossible, to recapitulate in tissue culture cells.

Bich-Thuy and Queen (35) recently found that several nonoverlapping restriction fragments up to 4 kb downstream of the mouse λ7 gene (the most 3' λ gene; see Fig. 6 C) are capable of activating the λ promoter and other promoters. Furthermore, Meyer and Neuberger (51) recently demonstrated the presence of a second, more powerful enhancer 9 kb downstream of the mouse κ C region. Since V region rearrangement into one of the λ genes deletes the upstream sequences, one might predict a downstream position for the λ enhancer. We have not detected enhancer activity in sequences as much as 2.1 kb downstream of λ7 (data not shown). We cannot rule out the possibility that the human λ locus contains additional enhancer sequences further 3' of the Cκ complex, but it seems that the 8.5-kb U266X clone contains the necessary sequences for high level, tissue-specific expression in vivo. Since these sequences are only active in transgenic mice, however, further characterization will require utilization of this in vivo model.

We thank Peter Gentile for oligonucleotide synthesis, and Cynthia Morton and Jacob Sarid for the 49a-4 cosmid. We are also grateful to Judy Swain and Robert Replogle for their early efforts in the project.

A portion of this work was supported by grants from E.I. Dupont, Inc., the Markey Foundation, and the American Business Foundation for Cancer Research.

Address correspondence to Philip Leder, Department of Genetics, Harvard Medical School, 25 Shattuck Street, Boston, MA 02115.

Received for publication 22 February 1990 and in revised form 16 May 1990.

References

1. Max, E.E. 1989. Immunoglobulins: molecular genetics. In Fundamental Immunology, 2nd ed. W.E. Paul, editor. Raven Press Ltd., New York. 235–290.

2. Hood, L., W.R. Gray, B.G. Sanders, and W.J. Dreyer. 1967. Light chain evolution: antibodies. Cold Spring Harbor Symp. Quant. Biol. 32:133.

3. Ein, D. 1968. Nonallelic behavior of the Oγ groups in human immunoglobulin chains. Proc. Natl. Acad. Sci. USA. 60:982.

4. Hess, M., N. Hilschmann, L. Rivat, C. Rivat, and C. Ropartz. 1971. Isotypes in human immunoglobulin λ chains. Nature (Lond.). 234:58.

5. Fets, J.W., and H.F. Deutsch. 1975. A new λ chain gene. Immunochernistry. 12:643.

6. Hilschmann, N., and L.C. Craig. 1965. Amino acid sequence studies with Bence-Jones proteins. Proc. Natl. Acad. Sci. USA. 53:1403.

7. Frangione, B., T. Moloshok, F. Prelli, and A. Solomon. 1985. Human λ light chain constant region gene Cλκ: the primary structure of λ VI Bence Jones protein. Proc. Natl. Acad. Sci. USA. 82:3415.

8. Hieter, P.A., G.F. Hollis, S.J. Korsmeyer, T.A. Waldmann, and P. Leder. 1981. Clustered arrangement of immunoglobulin λ constant region genes in man. Nature (Lond.). 294:536.

9. Dariavach, P., G. Lefranc, and M.-P. Lefranc. 1987. Human immunoglobulin Cα6 gene encodes the Kern "Oγ-" λ chain, and Cα4 and Cα5 are pseudogenes. Proc. Natl. Acad. Sci. USA. 84:9074.

10. Erickson, J., J. Martinis, and C.M. Croce. 1981. Assignment of the genes for human λ immunoglobulin chains to chromosome 22. Nature (Lond.). 294:173.

11. McBride, O.W., P.A. Hieter, G.F. Hollis, D. Swan, M.C. Otye, and P. Leder. 1982. Chromosomal location of human κ and λ immunoglobulin light chain constant region genes. J. Exp. Med. 155:1480.

12. 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.

13. Chang, H., E. Dmitrovsky, P.A. Hieter, K. Mitchell, P. Leder, L. Tuoroczi, I.R. Kirsch, and G.F. Hollis. 1986. Identification
of three new Ig \( \lambda \)-like genes in man. J. Exp Med. 163:425.

14. Kudo, A., N. Sakaguchi, and F. Melchers. 1987. Organization of the murine Ig-related \( \lambda \) gene transcribed selectively in pre-B lymphocytes. EMBO (Eur. Mol. Biol. Organ.) J. 6:103.

15. Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. 545 pp.

16. Ish-Horowicz, D., and J.K. Burke. 1981. Rapid and efficient cosmid vector cloning. Nucleic Acids Res. 9:2989.

17. Nilsson, K., H. Bennich, S.G.O. Johansson, and J. Pontén. 1970. Established immunoglobulin producing myeloma (IgE) and lymphoblastoid (IgG) cell lines from an IgE myeloma patient. Clin. Exp. Immunol. 7:477.

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

19. Tabor, S., and C.C. Richardson. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA. 84:4767.

20. Sheen, J., and B. Seed. 1988. Electrolyte gradient gels for DNA sequencing. Biotechniques. 6:942.

21. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387.

22. Feng, D.-F., and R.F. Doolittle. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25:351.

23. Sharp, P.A. 1983. Conversion of RNA to DNA in mammals: Alu-like elements and pseudogenes. Nature (Lond.). 301:471.

24. Hess, J.F., M. Fox, C. Schmidt, and C.-K. J. Shen. 1983. Molecular evolution of the human adult \( \alpha \)-\( \beta \)-globin-like gene region: insertion and deletion of \( \alpha \)-like repeats and non-\( \alpha \) DNA sequences. Proc. Natl. Acad. Sci. USA. 80:5970.

25. Akira, S., K. Okazaki, and H. Sakano. 1987. Two pairs of recombination signals are sufficient to cause immunoglobulin V(D)-J joining. Science (Wash. DC). 238:1134.

26. Hesse, J.E., M.R. Lieber, K. Mizuuchi, and M. Gellett. 1989. V(D)J recombination: a functional definition of the joining signals. Genes & Dev. 3:1053.

27. Shapiro, M.B., and P. Senapathy. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15:7155.

28. Ponstising, V.H., M. Hess, and N. Hilschmann. 1968. Die vollständige aminosäure-sequence des Bence-Jones-Proteins dern eine neue untergruppe der immunoglobulin-L-ketten vom \( \lambda \)-typ. Hoppe-Seyler's Z. Physiol. Chem. 349:867.

29. Selsing, E., J. Miller, R. Wilson, and U. Storb. 1982. Evolution of mouse immunoglobulin \( \lambda \) genes. Proc. Natl. Acad. Sci. USA. 79:4681.

30. Flanagan, J.G., M.-P. Lefranc, and T.H. Rabitts. 1984. Mechanisms of divergence and convergence of the human immunoglobulin \( \alpha 1 \) and \( \alpha 2 \) constant region gene sequences. Cell. 36:681.

31. Udell, J.A., and B.B. Blumberg. 1988. Intergenic exchange maintains identity between two human \( \lambda \) light chain immunoglobulin gene intron sequences. Nucleic Acids Res. 16:2959.

32. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell. 44:283.

33. Cone, R.D., E.B. Reilly, H.N. Eisen, R.C. Mulligan. 1987. Tissue-specific expression of functionally rearranged \( \lambda 1 \) Ig gene through a retrovirus vector. Science (Wash. DC). 236:954.

34. Picard, D., and W. Schaffner. 1983. Correct transcription of a cloned mouse immunoglobulin gene in vivo. Proc. Natl. Acad. Sci. USA. 80:417.

35. Bich-Thuy, L. t., and C. Queen. 1989. An enhancer associated with the mouse immunoglobulin \( \kappa \) gene is specific for light chain producing cells. Nucleic Acids Res. 17:5307.

36. Stalder, J., A. Larsen, J.D. Engel, M. Dolan, M. Groudine, and H. Weintraub. 1980. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAse I. Cell. 20:451.

37. Parslow, T.G., and D.K. Granner. 1983. Structure of a nucleosensitive region inside the immunoglobulin kappa gene: evidence for a role in gene regulation. Nucleic Acids Res. 11:4775.

38. Tabor, S., and C. C. Richardson. 1987. DNA sequence analysis with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 84:4767.

39. Sheen, J., and B. Seed. 1988. Electrolyte gradient gels for DNA sequencing. Biotechniques. 6:942.

40. Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12:387.

41. Feng, D.-F., and R.F. Doolittle. 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. 25:351.

42. Hesse, J.E., M.R. Lieber, K. Mizuuchi, and M. Gellett. 1989. V(D)J recombination: a functional definition of the joining signals. Genes & Dev. 3:1053.

43. Shapiro, M.B., and P. Senapathy. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. 15:7155.

44. Ponstising, V.H., M. Hess, and N. Hilschmann. 1968. Die vollständige aminosäure-sequence des Bence-Jones-Proteins dern eine neue untergruppe der immunoglobulin-L-ketten vom \( \lambda \)-typ. Hoppe-Seyler's Z. Physiol. Chem. 349:867.

45. Selsing, E., J. Miller, R. Wilson, and U. Storb. 1982. Evolution of mouse immunoglobulin \( \lambda \) genes. Proc. Natl. Acad. Sci. USA. 79:4681.

46. Flanagan, J.G., M.-P. Lefranc, and T.H. Rabitts. 1984. Mechanisms of divergence and convergence of the human immunoglobulin \( \alpha 1 \) and \( \alpha 2 \) constant region gene sequences. Cell. 36:681.

47. Udell, J.A., and B.B. Blumberg. 1988. Intergenic exchange maintains identity between two human \( \lambda \) light chain immunoglobulin gene intron sequences. Nucleic Acids Res. 16:2959.

48. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell. 44:283.

49. Cone, R.D., E.B. Reilly, H.N. Eisen, R.C. Mulligan. 1987. Tissue-specific expression of functionally rearranged \( \lambda 1 \) Ig gene through a retrovirus vector. Science (Wash. DC). 236:954.
52. Miller, J., E. Selsing, and U. Storb. 1982. Structural alterations in J regions of mouse immunoglobulin lambda chains are associated with differential gene expression. Nature (Lond.). 295:428.

53. Emanuel, B.S., P.C. Nowell, C. McKeon, C.M. Croce, and M.A. Israel. 1986. Translocation breakpoint mapping: molecular and cytogenetic studies of chromosomes 22. Cancer Genet. Cytogenet. 19:81.

54. Storb, U., D. Haasch, B. Arp, P. Sanchez, P.-A. Cazenave, and J. Miller. 1989. Physical linkage of mouse λ genes by pulsed-field gel electrophoresis suggests that the rearrangement process favors proximate target sequences. Mol. Cell. Biol. 9:711.

55. D'Eustachio, P., A.I.M. Bothwell, T.K. Takaro, D. Baltimore, and F.H. Ruddle. 1981. Chromosomal location of the structural genes encoding murine immunoglobulin λ light chains. J. Exp. Med. 153:793.

56. Kudo, A., and F. Melchers. 1987. A second gene, V_{μμ}, in the λ loci of the mouse, which appears to be selectively expressed in pre-B lymphocytes. EMBO (Eur. Mol. Biol. Organ.) J. 6:2267.