AMINO ACIDS AT THE SITE OF V_κ-J_κ RECOMBINATION NOT ENCODED BY GERMLINE SEQUENCES

BY MARY HELLER, JAMES D. OWENS, J. FREDERIC MUSHINSKI, AND STUART RUDIKOFF

From the Laboratory of Genetics, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

The Ig light chain is encoded by three gene segments coding, respectively, for the variable (V_κ), joining (J_κ), and constant (C_κ) regions (1, 2). In inbred mice, a given V_κ gene can presumably recombine with any of four functional J_κ genes. The site of this recombination corresponds to the last amino acid of the third complementarity determining region (CDR)[1] in the mature protein and thus, the recombination event can alter one amino acid that might potentially be involved in antigen binding.

In previous studies (3, 4) involving a series of hybridoma and myeloma proteins with binding specificity for β (1,6)-galactan, we observed that 16 of 17 light chains expressed the amino acid Ile at position 96, the site of V_κ-J_κ recombination. These are the only murine light chains described to date with Ile at position 96 (5). Furthermore, codons for Ile are not readily generated by recombination between sequences commonly found at the 3′ end of other germline V_κ genes and the 5′ end of J_κ genes. To examine the mechanisms contributing to the generation of this unusual amino acid at the recombination site, we have cloned and sequenced a cDNA encoding the functionally rearranged antigalactan light chain as well as the corresponding germline gene.

Materials and Methods

Nucleic Acid Procedures. Blotting was performed as described by Southern (6). Filters were hybridized at 65°C in 3× SSC, 10% dextran sulfate followed by three washes in 0.1× SSC, 0.1% SDS at 37°C, and a final wash in 0.1× SSC, 0.1% SDS at 65°C. In some experiments hybridization was at 37°C in 50% formamide followed by three washes at 37°C in 0.2× SSC, 0.2% SDS and a final wash in 0.2× SSC, 0.2% SDS at 65°C. cDNA clones were constructed as previously described (7). The BALB/c myeloma library in Charon 4A was kindly provided by Dr. Laurel Eckhardt, Columbia University, New York. C_κ and J_κ probes were a gift of Dr. Ed Max, National Institute of Allergy and Infectious Diseases, Bethesda, MD. Sequencing was performed either by Maxam-Gilbert chemical modification (8) or M13 subcloning and didexoxy chain termination (9).

Results

cDNA Clones. To determine the nucleotide sequence encoding expressed antigalactan light chains, cDNA clones were prepared from poly $(A)^+$ mRNA prepared from XRPC24 (X24) tumor tissues (7). cDNAs were C-tailed, inserted

[1] *Abbreviation used in this paper:* CDR, complementarity determining region.

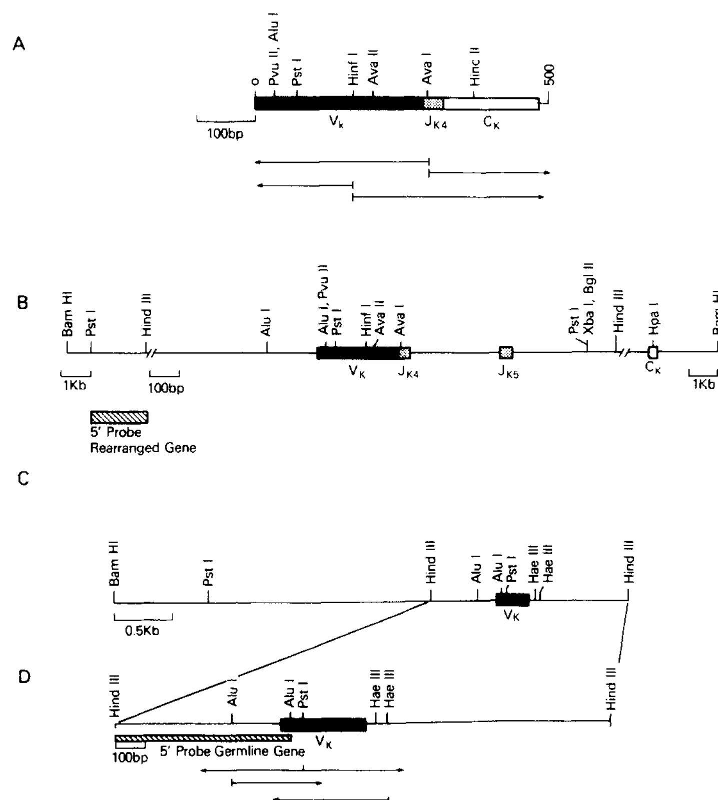

FIGURE 1. Restriction maps of X24 V_κ gene segments. Coding regions are designated by solid or stipled boxes. (*A*) cDNA clone of the expressed *X24* κ chain. Arrows indicate strategy for Maxam-Gilbert sequencing. (*B*) Partial map of the phage clone containing the rearranged *X24* V_κ gene. The 5′ probe used to isolate the germline counterpart is defined by the hatched box. (*C*) Partial map of phage clone containing the germline *X24* V_κ gene. (*D*) Expanded section of the germline clone. Arrows indicate strategy for Sanger chain termination sequencing. 5′ probe used for Southern blot analysis is defined by hatched box.

in G-tailed Pst I–digested pBR322, and transformed bacterial colonies were screened with a 260-bp Hinc II fragment from the C_κ constant region. DNA was prepared from positive colonies and subjected to restriction endonuclease mapping to determine the size of inserts. The plasmid X24-K3 containing the largest insert, ~500 bp, was used for subsequent sequence determination (Fig. 1*A*).

Germline Genes. Germline genes potentially encoding antigalactan light chains were initially characterized by Southern blot analysis of genomic BALB/c liver DNA using a 260-bp Pvu II–Ava I fragment from the X24-K3 cDNA corresponding to V_κ region amino acid positions 12–99. As can be seen in Fig. 2*A*, ~20 hybridizing bands are identified with this probe, indicating a relatively complex family. Attempts to determine which germline band was rearranged in antigalactan myeloma or hybridoma DNA were unsuccessful due to this complexity. Therefore, we proceeded to clone the rearranged gene from the X24 myeloma, which had previously been shown to be present on an 8.7-kb Bam HI fragment by hybridization with a J_κ probe detecting κ chain rearrangements (4).

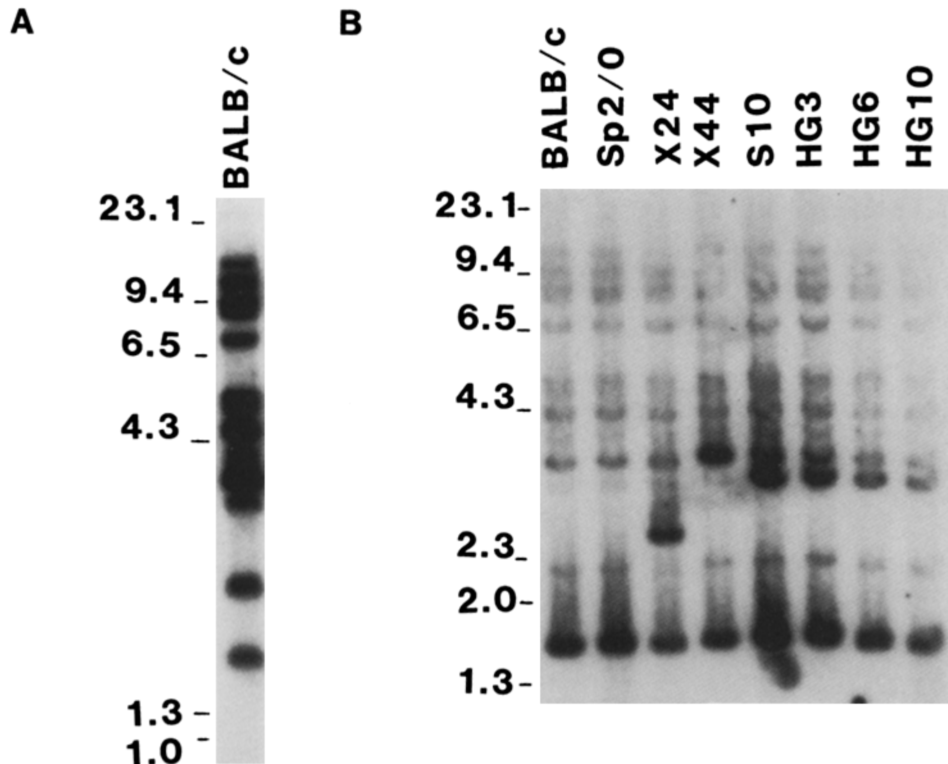

FIGURE 2. (*A*) Southern blot analysis of Hind III-digested BALB/c liver DNA. Filter was hybridized with a 260-bp Pvu II–Ava I cDNA fragment corresponding to variable region amino acids 12–99 at 37°C in 50% formamide. Final wash was at 65°C in 0.2% SDS, 0.2× SSC. (*B*) Southern blot analysis of Hind III-digested DNA from hybridoma and myeloma cell lines. Filter was hybridized with the 5′ germline probe (Fig. 1*D*) at 65°C in 3× SSC, 10% dextran sulfate. Final wash was at 65°C, 0.1% SDS, 0.1× SSC.

X24 DNA was cut with Bam HI and fractionated by sucrose gradient centrifugation. Fractions containing DNA fragments in the size range of 8–9 kb were screened with a J_κ probe, and DNA from positive fractions was ligated and packaged in the EMBL3 phage vector. This restricted library was then screened with the J_κ probe. Positive clones were rescreened with a C_κ probe, and detailed restriction mapping of one of the clones positive with both probes (Fig. 1*B*) clearly identified the rearranged gene. A 3.6-kb Ava I–Bam HI fragment containing 5′ flanking and coding region sequences was isolated and digested with Hind III. The resulting 2.7-kb Bam HI–Hind III fragment containing only 5′ sequences was subcloned into pUC8 and a 2.0-kb Pst I–Hind III fragment from this subclone was used to screen a BALB/c library. 5×10^5 recombinant phage were screened and 7 phage clones were identified containing the corresponding germline sequence. A 1.8-kb Hind III–Hind III fragment positive with the V region cDNA probe was isolated and used for sequencing studies as depicted in Fig. 1, *C* and *D*.

Sequence analysis of the cDNA and germline genes (Fig. 3) revealed complete identity throughout the entire variable region. Thus, the rearranged *X-24* light

```
                  1                                    10                                       20
                  Glu Ile Val Leu Thr Glu Ser Pro Ala Ile Thr Ala Ala Ser Leu Gly Glu Lys Val Thr Ile Thr Cys Ser
X24cDNA     AGGA  GAA ATT GTG CTC ACT CAG TCT CCA GCC ATC ACA GCT GCA TCT CTG GGG CAA AAG GTC ACC ATC ACC TGC AGT
X24 GL      ----  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                  30                                      40                                      50
            Ala Ser Ser Ser Val Ser Tyr Met His Trp Tyr Glu Glu Lys Ser Gly Thr Ser Pro Lys Pro Trp Ile Tyr Glu
X24cDNA     GCC AGC TCA AGT GTA AGT TAC ATG CAC TGG TAC CAG CAG AAG TCA GGC ACC TCC CCC AAA CCA TGG ATT TAT GAA
X24GL       --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                                60                                      70
            Ile Ser Lys Leu Ala Ser Gly Val Pro Ala Arg Phe Ser Gly Ser Gly Ser Gly Thr Ser Tyr Ser Leu Thr Ile
X24cDNA     ATA TCC AAA CTG GCT TCT GGA GTC CCA GCT CGC TTC AGT GGC AGT GGG TCT GGG ACC TCT TAC TCT CTC ACA ATC
X24GL       --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                80                                      90                          9 6
            Ser Ser Met Glu Ala Glu Asp Ala Ala Ile Tyr Tyr Cys Glu Glu Trp Asn Tyr Pro Leu I le Thr Phe Gly Se
Jκ4GL                                                                                           T TC ACG TTC GGC TC
X24cDNA     AGC AGC ATG GAG GCT GAA GAT GCT GCC ATT TAT TAC TGC CAG CAG TGG AAT TAT CCT CTT A TC ACG TTC GGC TC
X24GL       --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- - CACAGTG
                                                                                                  *******

            r Gly Thr Lys Leu Glu Ile Lys Arg
Jκ4GL       G GGG ACA AAG TTG GAA ATA AAA CGG
X24cDNA     G GGG ACA AAG TTG GAA ATA AAA CGG
```

FIGURE 3. Sequence comparison of *X24* cDNA and the *X24* V_{κ} germline (GL) gene. The $J_{\kappa4}$ germline sequence is taken from Max et al. (1). Amino acid position 96 has been divided to reflect the recombination site at which the first nucleotide (A) of the codon is contributed by the V_{κ} gene and the remaining two (T,C) by the J_{κ} gene. The conserved heptamer recombination sequence is marked by asterisks. The X24 germline sequence is identical to a V_{κ} germline gene (H6) crossreacting with members of the family encoding light chains involved in the immune response to oxazolone (10).[2]

[2] These sequence data have been submitted to the EMBL/GenBank Data Libraries under accession number Y00632.

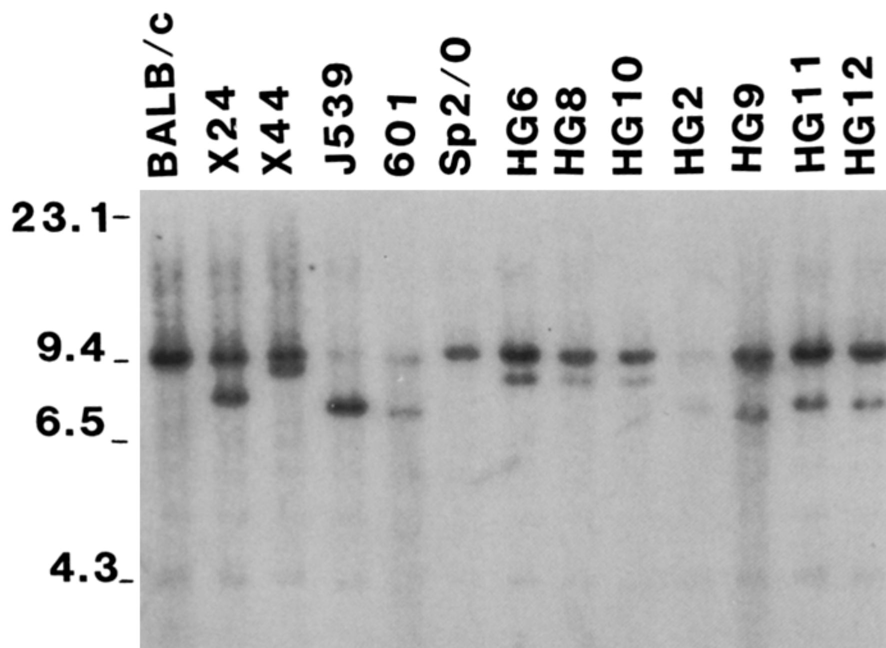

FIGURE 4. Southern blot analysis of Bam HI-digested hybridoma and myeloma DNAs. Probe and conditions were as in Fig. 2*B*.

chain gene lacks somatic mutations and reflects an unaltered germline sequence. The sequence between the nominal 3′ end of the germline variable region and the beginning of the consensus heptamer recombination sequence consists of a single nucleotide, Adenine. This base plays a key role in the generation of an Ile codon found at the site of V_κ-J_κ recombination in 16 of 17 antigalactan light chains.

Usage of the X24 Germline V_κ Gene in Other Antigalactan Hybridomas and Myelomas. Rearrangement of the *X24* germline gene in a battery of antigalactan hybridoma and myeloma lines was examined in a series of Southern blot experiments. Hybridization of either Bam HI (Fig. 4) or Hind III (Fig. 2*B*) digests with an ~600-bp Hind III–Alu I fragment (Fig. 1*D*) containing 5′ sequences as well as 30 bp of the coding region, revealed a rearranged fragment in all lines except the fusion partner, Sp2/0. The variability in the size of the rearranged fragment in each line is consistent with recombination of a single variable region gene segment with different J_κ genes. These studies, in conjunction with previous protein sequencing (3) and Southern blot analyses (4), demonstrate that X44 uses the $J_{\kappa 1}$ segment, HyGals 3, 6, 8, 10, and S10 the $J_{\kappa 2}$ segment, and X24 the $J_{\kappa 4}$ segment. HyGals 6, 8, and 10 are clonally related (4) so that $J_{\kappa 2}$ has recombined with this V_κ gene on three independent occasions. The remaining 10 lines in this collection, HyGals 1, 2, 4, 7, 9, 11, 12, T191, J539, and T601, use the $J_{\kappa 5}$ segment.

TABLE I
Nucleotides Found 3′ to V_κ Germline Genes

Gene	Codon 95	3′ Sequence	Heptamer	Reference
$V_{\kappa21}$	Pro CCT	CC	CACAGTG	11
$V_{\kappa41}$	Pro CCT	CC	CACAGTG	1
L6	Pro CCT	CC	CACAGTG	12
L7	Pro CCA	AC	CACAGTG	12
M167	Pro CCT	C	CACAGTG	13
X24	Leu CTT	A	CACAGTG	This paper

Sequences in the table are contiguous in the germline. Codon 95 represents the last codon of the V_κ gene segment. The 3′ sequence indicates nucleotides found between this codon and the heptamer recombination signal.

Discussion

The gene encoding a functional Ig light chain is formed by recombination between a variable region gene segment that nominally encodes amino acids 1–95, one of several J_κ segments that encode amino acids 96–108, and the C_κ segment. The site of recombination between V_κ and J_κ has been shown to vary considerably. On the 5′ side, recombination may occur either within the last variable region codon (95) or between the 3′ end of this codon and the beginning of the conserved heptamer recombination sequence. The region 3′ of the variable gene segment usually consists of one or two bases that are predominantly C nucleotides, as illustrated in Table I. The actual site of recombination is thus involved in the determination of the last amino acid in the third CDR. Interestingly, in the mouse the length of the third CDR is highly conserved, with very few variants having been reported (14–18), although, from the three-dimensional structure of the murine κ chain (19), there is no obvious constraint that would predict such a conservation. In fact, considerable size variation is observed in this region of rabbit κ chains (5).

The maintenance of murine CDR-3 size is accomplished by compensation in the location of the 3′ and 5′ recombination sites, so that in nearly all cases, a single codon is generated during this event. This is somewhat surprising in that there is considerable variation in length at the sites of heavy chain V_H-D-J_H assembly due to the apparent random addition of nucleotides between the encoded segments (20, 21). It is thus curious that Ig light and heavy chain gene segments use the same heptamer/nonomer recombination sequences, yet unlike heavy chain V_H-D-J_H recombination, no such variation in length is observed at the site of V_κ-J_κ joining. In fact, to our knowledge there are no reported examples of V_κ-J_κ recombination in which *de novo* addition of nucleotides occur. All documented recombination events involve, with the exception of somatic point mutations, only sequences encoded in V_κ and J_κ germline genes. The antigalactan light chains appear to be an exception. We have previously reported that 16 of

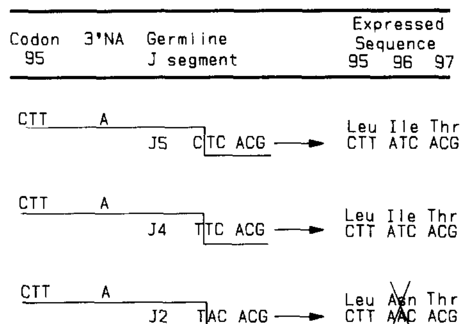

FIGURE 5. Mechanisms for generation of an Ile codon corresponding to amino acid position 96.

17 antigalactan light chains express Ile at position 96, the point of V_κ-J_κ recombination (Ile is not found at this position in any other murine V_κ region). Since this amino acid could not readily be generated by sequences usually found at the 3′ end of V_κ and the 5′ ends of J_κ genes, we have cloned the rearranged and germline genes used in this response to determine the origin of the Ile codon at position 96. Sequence comparison of the rearranged and germline gene encoding these light chains (Fig. 3) reveals a single nucleotide (A) between codon 95 of the V region and the consensus recombination sequence. From Fig. 5, it can thus be seen that light chains using $J_{\kappa 4}$ or $J_{\kappa 5}$ can readily generate the Ile codon, ATC, by combining the A nucleotide 3′ of the V_κ gene and the second and third nucleotides (T,C) of the first codon from either $J_{\kappa 4}$ or $J_{\kappa 5}$. However, an examination of the $J_{\kappa 2}$ sequence reveals that no recombination using germline sequences can generate an Ile codon (ATA, ATT or ATC). Instead, recombination in the same reading frame postulated for $J_{\kappa 4}$ and $J_{\kappa 5}$ (which generates the Ile codon), in fact, would produce an Asn codon that is not found in any of the antigalactan light chains. Southern blot analysis (Figs. 2*B* and 4) using a flanking region probe (Fig. 1*D*) from the germline antigalactan light chain revealed that all antigalactan hybridoma and myeloma lines used the same V_κ gene. A number of these, including HyGal3, HyGal6, and SAPC10 expressed Ile at position 96 (3, 4) in conjunction with a $J_{\kappa 2}$ segment. Although it is possible that a second V_κ gene exists that is nearly identical to the one we have cloned and has additional nucleotides 3′ to the V_κ segment that might encode an Ile, we have found no evidence for such a second gene in Southern blot analyses using a variety of enzymes. Since Ile cannot be generated by the germline V_κ sequence we have isolated and the germline sequence of $J_{\kappa 2}$ (Fig. 5), the Ile in these chains must result either from somatic mutation occurring within this trinucleotide to produce an Ile codon or from the *de novo* addition of nucleotides that form an Ile codon. Either of these events would presumably be followed by strong selection. In the putative case of somatic mutation, alterations would have had to occur in both nucleotides contributed to this codon by the J_κ gene on three independent occasions, assuming the A nucleotide 3′ of the V_κ gene is the first nucleotide of this codon. Alternatively, a deletion of the second nucleotide in the first codon of $J_{\kappa 2}$ (TAC) could also generate an Ile codon. In any case, the mechanism used to derive the Ile codon at position 96 appears to be novel in terms of documented V_κ-J_κ recombination and may represent an example of *de novo* addition of nucleotides, as seen in heavy chain gene segment recombination. It is of further

interest that none of the light chains expressing Ile at position 96 use $J_{\kappa 1}$. As in the case of $J_{\kappa 2}$, an Ile codon cannot be formed by recombination of the galactan V_κ segment and the germline $J_{\kappa 1}$. Only X44, which has Trp at position 96, uses $J_{\kappa 1}$.

The occurrence of Ile at position 96 in 16 of 17 antigalactan light chains suggests strong selection for this particular residue, most likely at the level of antigen-driven clonal selection. The nearly exclusive use of Ile is somewhat surprising in that the single protein in this collection that does not have an Ile at position 96 is the myeloma XRPC44 (X44). In X44, position 96 is Trp, which is directly encoded in the germline $J_{\kappa 1}$ segment found in this protein. A number of structural studies (22, 23) have failed to demonstrate significant binding differences between the X44 protein and other members of the collection. Thus, Trp appears to be functionally interchangeable for Ile, at least in terms of the assays previously used. While both amino acids are hydrophobic in nature, a significant difference in side chain volume exists between the two, suggesting that the preference for Ile must, in some manner, reflect a corresponding functional difference. The presence of Ile at the V_κ-J_κ junction in the antigalactan light chains appears similar to the situation among antiarsonate light chains in which an essentially invariant Arg is found at this same position (24, 25). Like Ile, Arg cannot be generated from known germline sequences and the repeated occurrence of this amino acid at the site of recombination may similarly reflect novel recombinational processes not previously seen in murine κ chains.

Summary

Murine V_κ-J_κ recombination is characterized by a maintenance of size at the site of recombination and the use of nucleic acids found only in germline sequences. This is in contrast to heavy chain V_H-D-J_H assembly where random nucleotides are added at the recombination sites to produce considerable size variation, even though the heptamer/nonomer recombination sequences are identical in both κ and heavy chain genes. We have examined the origin of an unusual amino acid, Ile, found at the site of V_κ-J_κ recombination in antigalactan antibodies, by sequence analysis of the corresponding rearranged and germline genes. Results indicate that the Ile codon can be generated by use of a single nucleotide 3′ of the V_κ segment in combination with the second and third nucleotides of the first codon of $J_{\kappa 5}$ or $J_{\kappa 4}$. However, several antigalactan antibodies express Ile in combination with $J_{\kappa 2}$. An Ile codon cannot be generated by recombination in any reading frame between germline V_κ and $J_{\kappa 2}$ segments. These results suggest that the origin of the Ile codon in lines using $J_{\kappa 2}$ may represent a novel event in murine light chain assembly, possibly similar to the *de novo* addition of nucleotides observed in heavy chain gene recombination.

Received for publication 31 March 1987 and in revised form 26 May 1987.

References

1. Max, E. E., J. G. Seidman, and P. Leder. 1979. Sequence of five potential recombination sites encoded close to an immunoglobulin κ constant region gene. *Proc. Natl. Acad. Sci. USA.* 76:3450.

2. Sakano, H., K. Huppi, G. Heinrich, and S. Tonegawa. 1979. Sequence at the somatic recombination sites of immunoglobulin light chain genes. *Nature (Lond.).* 280:288.
3. Pawlita, M., M. Potter, and S. Rudikoff. 1982. κ-chain restriction in anti-galactan antibodies. *J. Immunol.* 129:615.
4. Rudikoff, S., M. Pawlita, J. Pumphrey, and M. Heller. 1984. Somatic diversification of immunoglobulins. *Proc. Natl. Acad. Sci. USA.* 81:2162.
5. Kabat, E. A., T. T. Wu, H. Bilofsky, M. Reid-Miller, and H. Perry. 1983. Sequences of Proteins of Immunological Interest. U.S. Department of Health, Education and Welfare, Public Health Service, National Institutes of Health, Bethesda. 323.
6. Southern, E. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503.
7. Mushinski, J. F., F. R. Blattner, J. D. Owens, F. D. Finkelman, S. W. Kesler, L. Fitzmaurice, M. Potter, and P. W. Tucker. 1980. Mouse immunoglobulin D: Construction and characterization of a cloned delta chain cDNA. *Proc. Natl. Acad. Sci. USA.* 77:7405.
8. Maxam, A., and W. Gilbert. 1977. A new method for DNA sequencing. *Proc. Natl. Acad. Sci. USA.* 74:560.
9. Sanger, F., S. Nickles, and A. Colson. 1977. DNA sequencing with chain terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* 74:5463.
10. Even, J., G. M. Griffiths, C. Berek, and C. Milstein. 1985. Light chain germline genes and the immune response to 2-phenyloxazolone. *EMBO (Eur. Mol. Biol. Organ.) J.* 4:3439.
11. Heinrich, G., A. Traunecker, and S. Tonegawa. 1984. Somatic mutation creates diversity in the major group of mouse immunoglobulin κ light chains. *J. Exp. Med.* 159:417.
12. Pech, M., J. Hochtl, H. Schnell, and H. G. Zachau. 1981. Differences between germline and rearranged immunoglobulin V_{κ}. Coding sequences suggest a localized mutation mechanism. *Nature (Lond.).* 291:668.
13. Selsing, E., and U. Storb. 1981. Somatic mutation of immunoglobulin light chain variable region genes. *Cell.* 25:47.
14. Weigert, M., L. Gatmaitan, E. Loh, J. Schilling, and L. Hood. 1978. Rearrangement of genetic information may produce immunoglobulin diversity. *Nature (Lond.).* 276:785.
15. Lazure, C., W. T. Hum, and D. M. Gibson. 1981. Sequence diversity within a subgroup of mouse immunoglobulin kappa chains controlled by the *Igk-Ef2* locus. *J. Exp. Med.* 154:146.
16. Dzierzak, E. A., P. Brodeur, T. Marion, C. A. Janeway, Jr., and A. Bothwell. 1985. Molecular characterization of antibodies bearing Id-460. II. Molecular basis for Id-460 expression. *J. Exp. Med.* 62:1494.
17. Smith-Gill, S. J., P. A. Hamel, M. H. Klein, S. Rudikoff, and K. J. Dorrington. 1986. Contribution of the Vκ4 light chain to antibody specificity for lysozyme and β (1,6)-D-galactan. *Mol. Immunol.* 23:919.
18. Kaartinen, M., and O. Makela. 1987. Functional analogues of the Vκ OX1 gene in different strains of mice: evolutionary conservation but diversity based on V-J joining. *J. Immunol.* 138:1613.
19. Segal, D. M., E. A. Padlan, G. H. Cohen, S. Rudikoff, M. Potter, and D. R. Davies. 1974. The three dimensional structure of a phosphorylcholine-binding mouse immunoglobulin and the nature of the antigen binding site. *Proc. Natl. Acad. Sci. USA.* 71:4298.
20. Alt, F. W., and D. Baltimore. 1982. Joining of immunoglobulin heavy chain gene

segments: implications from a chromosome with evidence of three D-J_H fusions. *Proc. Natl. Acad. Sci. USA.* 79:4118.

21. Clarke, S. H., J. L. Claflin, M. Potter, and S. Rudikoff. 1983. Polymorphisms in anti-phosphorylcholine antibodies reflecting evolution of immunoglobulin families. *J. Exp. Med.* 157:98.
22. Jolley, M. E., S. Rudikoff, M. Potter, and C. P. J. Glaudemans. 1973. Spectral changes on binding of oligosaccharides to murine immunoglobulin A myeloma proteins. *Biochemistry.* 12:3039.
23. Jolley, M. E., C. P. J. Glaudemans, S. Rudikoff, and M. Potter. 1974. Structural requirements for the binding of derivatives of D-galactose to two homogeneous murine immunoglobulins. *Biochemistry.* 13:3179.
24. Siegelman, M., and J. D. Capra. 1981. Complete amino acid sequence of light chain variable regions derived from five monoclonal anti-*p*-azophenyl-arsonate antibodies differing with respect to a cross-reactive idiotype. *Proc. Natl. Acad. Sci. USA.* 78:7679.
25. Juszczak, E., R. I. Near, M. L. Gefter, and M. N. Margolies. 1984. Complete heavy and light chain variable region sequence of anti-arsonate monoclonal antibodies from BALB/c and A/J mice sharing the 36-60 idiotype are highly homologous. *J. Immunol.* 133:2603.