

Immunoglobulin Gene Rearrangements in Normal Mouse B Cells

P. EARLY,^{1†*} C. NOTTENBURG,² I. WEISSMAN,² AND L. HOOD¹

Division of Biology, California Institute of Technology, Pasadena, California 92115,¹ and Department of Pathology, Stanford Medical School, Stanford, California 94305²

Received 19 November 1981/Accepted 19 March 1982

We have analyzed the structure of rearranged μ heavy-chain genes obtained from the genomic DNA of normal BALB/c mouse spleen cells expressing surface immunoglobulin M. Examples were found of two types of nonproductive rearrangements, which may be responsible for allelic exclusion in normal B cells. In one of these rearrangements, a germ line *D* gene segment has joined to the *J_{H4}* gene segment but no *V/D* joining has occurred. We present evidence that *D* gene segments lie as a cluster between *V* and *J* gene segments in the germ line. A comparison of conserved sequences in *V* and *D* gene segments suggests that the *D* gene segments, which are found only in the heavy-chain gene family, may have evolved from *V* gene segments similar to the *V_κ* family.

Vertebrate immune systems are able to respond specifically to an almost unlimited number of different foreign molecules, or antigens. A major component of this response is the production of immunoglobulins binding specifically to any given antigen. Immunoglobulins are synthesized by the B-cell branch of the immune system, which in mammals consists of cells that originate in bone marrow and then migrate to the spleen and peripheral lymphoid organs. Each clone of B cells synthesizes a unique species of immunoglobulin. Thus, the population of B cells obtained from a spleen may contain almost as many different immunoglobulin species as cells. It has been estimated that B cells synthesize on the order of 10^7 different immunoglobulins (3).

The genes encoding immunoglobulins display several complex features associated with the need to generate such a large repertoire of specificities. There are three unlinked immunoglobulin gene families: the κ and λ gene families encode light chains, and a single gene family encodes heavy chains. An immunoglobulin molecule contains one species of light chain and one species of heavy chain in a basic H_2L_2 dimeric structure. The amino-terminal 110 to 125 amino acid residues of each pair of heavy and light chains form a variable region (*V*) domain, which binds specifically to antigen. The remainder of the two chains comprise the constant region (*C*) domains, which are involved in interactions with other components of the immune system. Heavy chains may contain any one of eight different constant regions, the most frequent being μ

(immunoglobulins with μ heavy chains are termed immunoglobulin M [IgM]).

Variable regions, as the name implies, show the most diversity when two species of immunoglobulin molecules are compared. Logically, this is due to the role of variable regions in recognizing a broad spectrum of diverse antigens. At least one specific genetic mechanism exists to create variable region diversity: the process of DNA joining. Light-chain variable regions are encoded by two gene segments, *V* and *J*, which are separated in germ line DNA (2). Heavy-chain variable regions are encoded by three such gene segments, *V*, *D*, and *J* (8, 14, 23). The *J* gene segments are located a short distance 5' to constant region gene segments in germ line DNA, so that *V-J* and *V-D-J* joining creates a complete immunoglobulin gene including both variable and constant region sequences.

The process of *V-J* or *V-D-J* DNA joining appears to involve specific sequences at the boundaries of the germ line gene segments which probably serve as recognition sites for "joining enzymes." The same recognition sequences appear in all the immunoglobulin gene families (and in both humans and mice) and are of the general form GGTTTTGT-(spacer)-CACTGTG when they occur 5' to the coding sequence of a gene segment, and the inverse complement of this sequence when they occur 3' to a gene segment (8, 14, 16, 20, 22, 23). The spacer lengths are either about 12 base pairs (bp) (roughly one turn of the DNA helix) or 23 bp (roughly two turns). Gene segments that can join to one another always have different spacer lengths: "one turn" for *V_κ* gene segments and

† Present address: Department of Molecular Biology, University of California, Berkeley, California 94720

"two turns" for J_{κ} gene segments, for example (8).

Although the sequences associated with DNA joining are fairly precise, the process itself is not and can result in junction points ranging over one or two codons (sometimes more) for any given gene segment (11). This variability in joining may contribute to the functional diversity of immunoglobulins, since $V-J$ and $V-D-J$ joining occur at a part of the variable region which is quite important for antigen binding specificity. Joining can also result in genes with frameshifts or termination codons.

All of the current information on the structure of rearranged immunoglobulin genes is the result of work with lymphoid tumor lines, particularly murine plasmacytomas or myelomas. Although they are convenient for study, such tumor lines are known to be restricted in the types of immunoglobulins they produce, and they may not be entirely representative of the processes leading to the diversity of the B-cell immunoglobulin repertoire. The present study was undertaken to look directly at variable region gene rearrangements in the pool of normal B cells synthesizing IgM, which is the first heavy-chain class synthesized after $V-J$ and $V-D-J$ rearrangement.

MATERIALS AND METHODS

The procedure for purification of BALB/c B lymphocytes has been described elsewhere (18). Briefly, spleen cells were stained with rabbit anti-mouse μ serum and fluoresceinated goat anti-rabbit serum before separation with a fluorescence-activated cell sorter. Greater than 97% of the cells collected stained positively for both surface μ and surface δ chains. A library of 12×10^6 recombinant phage was generated by ligating *EcoRI* partial digests of DNA from the sorted cells to DNA from the phage vector Charon 4A (1), followed by in vitro packaging (10).

The preparation of the sperm DNA library and procedures for screening the libraries have been described (5, 6). The Maxam-Gilbert protocol for DNA sequencing was used throughout (17). Procedures for handling recombinant DNA materials conformed to National Institutes of Health guidelines.

RESULTS

Productive rearrangements. Eleven recombinant Charon 4A phage containing μ heavy-chain genes derived from IgM⁺ BALB/c spleen cells were screened for the presence of J_H rearrangements. One phage contained the 6.5-kilobase (kb) *EcoRI* fragment characteristic of the un-rearranged germ line J_H configuration (the top line in Fig. 1), whereas the other phage either did not include the J_H locus (four cases) or displayed an altered J_H fragment consistent with DNA rearrangement. Four phage with altered J_H fragments were chosen at random for further

study. More detailed restriction mapping showed that two of these phage contained rearrangements at the J_{H1} gene segment (B2 and B49 in Fig. 1) and two contained rearrangements at the J_{H4} gene segment (B104 and B38 in Fig. 1).

DNA sequences of the B2 and B49 J_{H1} rearrangements showed that both are the result of $V-D-J$ joining to create V_H genes from which heavy chains can apparently be produced. Figure 2 shows the complete sequence of the B2 rearranged gene. So far as they can be determined, the positions of the $V-D$ and $D-J$ junctions are shown beneath the nucleotide sequence. The protein sequence that can be translated from the B2 gene is shown immediately above its nucleotide sequence. It is numbered from what would be the first amino acid of the mature polypeptide, after leader peptide cleavage. The B2 gene displays the same division of leader and main body exons found in myeloma genes (24). The RNA splice sites for these exons are doubly underlined in Fig. 2. The protein sequence translated from the B2 gene is closely related to group II myeloma heavy-chain sequences, such as that of the dextran-binding myeloma protein M104E (13), the sequence of which is shown above the B2 protein sequence in Fig. 2. The 3' portion of the B2 V_H gene segment is also homologous to group III myeloma sequences and hybridizes with a cDNA probe from the S107 V_H gene (unpublished data).

About 90% of the B49 gene was sequenced (Fig. 2). The protein sequence translated from this gene is homologous to the group I myeloma protein sequence of M315, which is shown on the upper lines.

Although both the B2 and B49 normal B-cell genes contain V_H gene segments only distantly related to V_H gene segments of the group III phosphorylcholine-binding myelomas, they use very similar or identical D gene segments and the same J_{H1} gene segment in forming the complete variable region gene. This point is shown more clearly in Fig. 3, where V_H-D and $D-J_{H1}$ junction points for the normal B-cell genes (B2 and B49) and group III myeloma genes (M167, M603, and S107) are compared. The arrows under the germ line J_{H1} sequence show the various $D-J_{H1}$ junctions, and the full D segment sequence in each rearranged gene is shown below. Figure 3 also shows how differences in junction points can contribute to amino acid sequence diversity in the third hypervariable region of the completed heavy chain. A similar diversity of junction points is seen in the J_{H4} rearrangements (Fig. 4).

Nonproductive rearrangements. In contrast to the two J_{H1} rearrangements we analyzed, neither of the two J_{H4} rearrangements contains a functional $V-D-J$ gene. This is shown in Fig. 4,

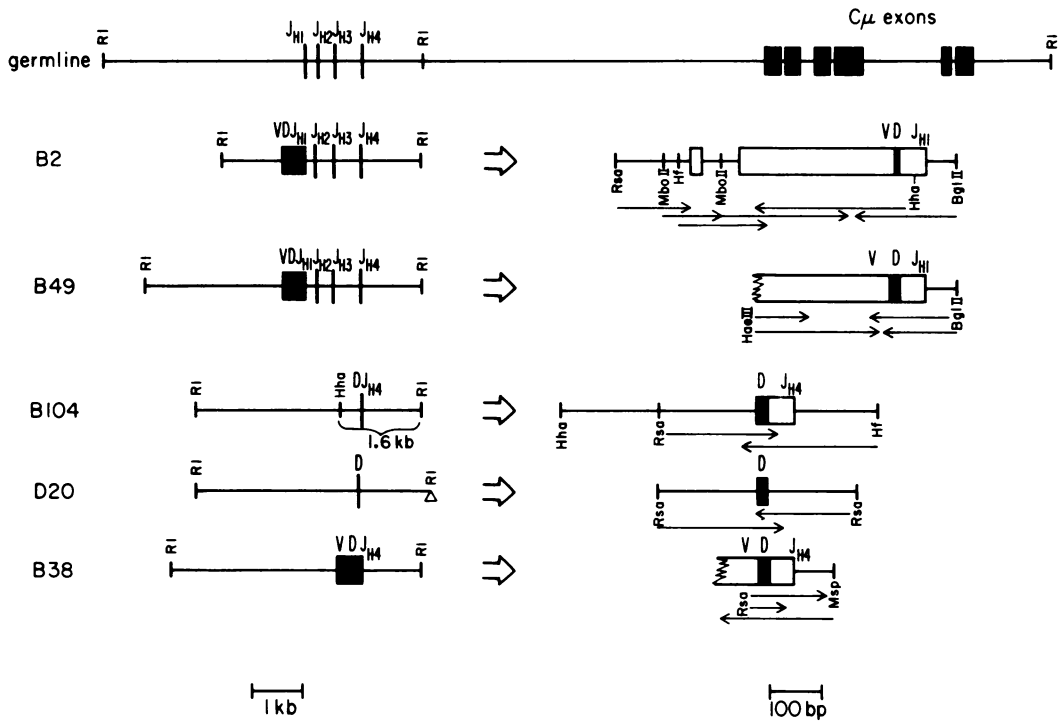


FIG. 1. Maps of cloned DNAs and DNA sequencing strategies. The map of the germ line J_H-C_μ locus has been reported elsewhere (8, 15, 24). All clones were derived from the IgM⁺ B cell library produced by partial digestion with *EcoRI*, with the exception of the germ line D20 clone, which came from a *HaeIII-AluI* linker library (5) (the triangle indicates a possible linker *EcoRI* site in D20). Only relevant portions of the total insert DNAs in each recombinant phage are depicted. The diagram to the right shows the Maxam-Gilbert sequencing strategy used for each clone.

where the germ line J_{H4} sequence is displayed on the top line and the sequences of the two normal B-cell rearrangements, B38 and B104, are shown on the next two lines. The arrows under the germ line J_{H4} sequence show the D - J junction points of B38 and B104, and the arrows above the J_{H4} sequence show junction points for various myeloma genes (24). The B38 sequence is split at the V - D junction to allow a reader comparison with the D segments beneath it. B38 is a V - D - J rearrangement, but the V_H and J_H gene segments are in different reading frames, creating termination codons in the D segment (doubly underlined). Although only about a quarter of its sequence was determined, the V_H segment of B38 displays homology to the B49 V_H gene segment and group I myeloma variable regions.

Since the translational reading frame is established by the V gene segment, the J gene segment in a rearranged gene must be in the same reading frame as the V gene segment. Otherwise, in-frame translation of the constant region which follows the J segment could not occur. There is no such constraint within the D gene

segment, and the use of alternative reading frames for germ line D gene segments may contribute to immunoglobulin diversity. However, the D gene segment in B38 contains termination codons in two translational reading frames: one of these frames is that of the V_H gene segment, and the other is the same as the J_{H4} gene segment. As far as can be determined, the germ line V_H , D , and J_H gene segments in B38 are each functional, and the lack of function of the B38 gene is solely the result of the nonproductive rearrangement.

The second J_{H4} rearrangement, B104, does not contain a V_H gene segment. Instead, an in-frame D - J_{H4} junction has taken place without a V_H - D junction. As a result, B104 retains the germ line sequence 5' to the D gene segment, including the one-turn (12-bp spacer) joining sequence originally predicted for germ line D gene segments (8). (This and other joining sequences are underlined once in Fig. 4.) Although the points of D - J_{H4} joining differ, the B38 and B104 D gene segments are identical where they overlap, and both may be derived from the same germ line D gene segment. If this is the case, it

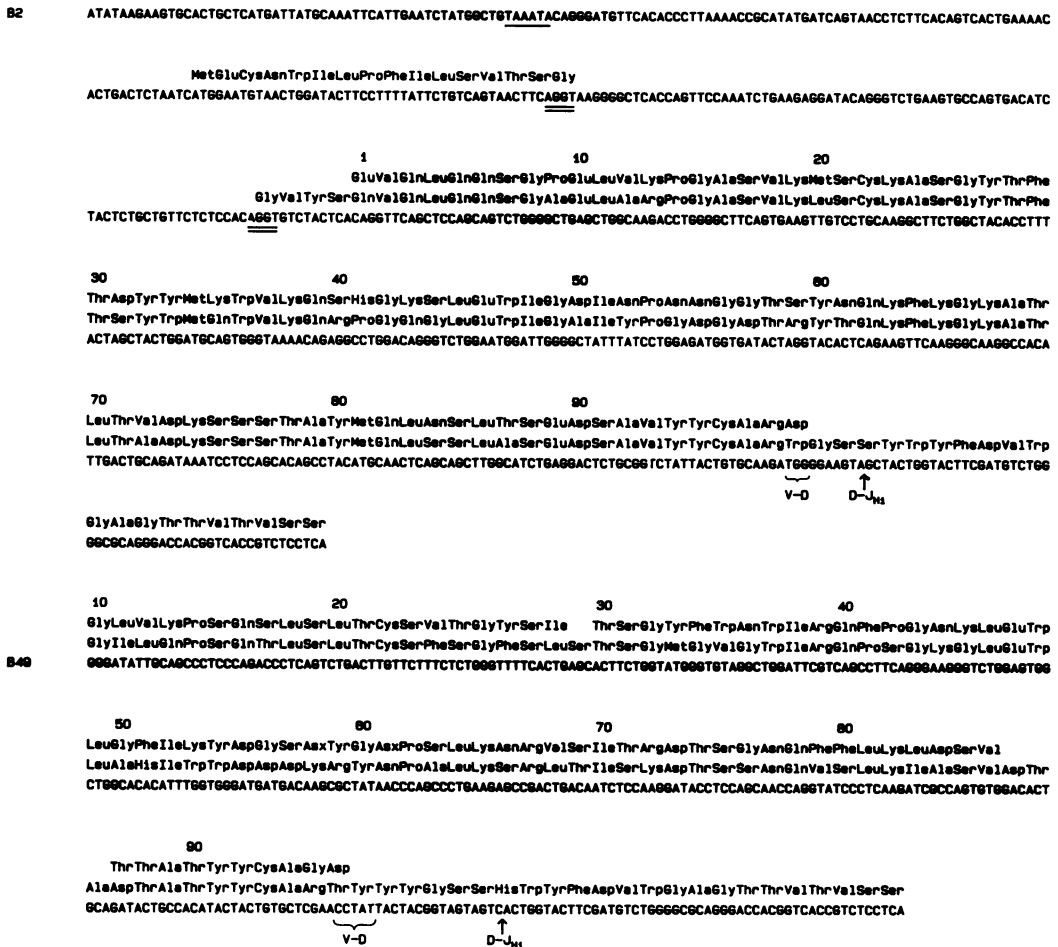


FIG. 2. Sequences of productive *V-D-J* rearrangements from normal B cells. The complete B2 gene is shown, as well as a portion of the B49 gene beginning approximately at the codon for the 10th amino acid of the mature polypeptide. Positions of *V-D* junctions cannot be determined accurately without the germ line gene segment sequences. The amino acid sequences translated from these genes are shown above the DNA sequences. Above the translated amino acid sequences are the sequences of the M104E (B2) and M315 (B49) myeloma proteins (13). Numbering is from the first amino acid residue of the mature polypeptide chain. Gaps were introduced in the M315 sequence to maximize homology; the larger of these gaps may indicate an error in the protein sequence since it involves a repeated amino acid sequence. Many of the amino acid differences between B49 and M315 are conservative changes. The double underlines indicate RNA splice junctions between the signal peptide and main body exons, and the single underline indicates a potential "TATAAATA box" for the B2 gene (M. Goldberg, Ph.D. thesis, Stanford University, Stanford, Calif., 1979).

demonstrates the diverse consequences that two independent joining events with the same *D* and *J_H* gene segments can have: two junction points on both the *D* and the *J_H* gene segments, with one junction in frame and one out of frame, and one joining event accompanied (not necessarily simultaneously) by *V_H-D* joining, whereas the other has occurred without *V_H-D* joining.

Germ line locus of *D* gene segments. Since the B104 gene retains germ line sequences 5' to the *D* gene segment, it was possible to use this gene

as a probe for germ line *D* gene segments. A 1.6-kb *HhaI-EcoRI* fragment of B104 (Fig. 1) containing both *D* and *J_H* sequences was used to probe a phage library containing germ line BALB/c DNA. Several phage were isolated by this procedure, and the sequence of one of them D20, is shown in Fig. 4. Although the nucleotide sequence is not completely identical to the B104 *D* gene segment, the structure of the D20 gene segment confirms the identification of B104 as a *D-J_{H4}* junction. The D20 germ line *D* gene

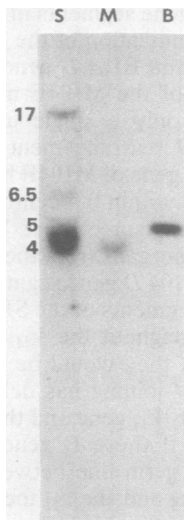


FIG. 5. Hybridization of the B104 *D-J* probe to *EcoRI*-digested genomic DNAs from BALB/c sperm (S) or M104E myeloma tumors (M), and to *EcoRI*-digested B104 DNA (B). Sizes in kilobases were determined from a parallel lane of *HindIII*-digested λ DNA.

gene segments seen by others (14). In the M104E myeloma (lane M), in contrast, only a single band of 4 kb is visible, which is the fragment containing J_H sequences in the rearranged *V-D-J* gene of this myeloma (D. Livant, unpublished data). The disappearance of all the *D* gene segment bands indicates that this group of *D* gene segments, at least, is located between the M104E V_H gene segment and the J_H locus in the germ line. Although one *D* gene segment unrelated to the SP2.1 family is located less than 1 kb 5' to J_{H1} (23), none of the *D* gene segments hybridizing to the B104 probe is within 8 kb of the J_H locus (unpublished data). Additionally, only a single germ line *D* gene segment homologous to B104 was found on the 4.7-kb *EcoRI* fragment from D20 (Fig. 1). The *D* gene segments thus appear to be much more widely dispersed than the *J* gene segments.

DISCUSSION

Gene rearrangement in normal B cells. Although a great deal of information has been gathered on the structure of rearranged immunoglobulin genes from myeloma tumors, it was not known how representative this was of the processes occurring in normal B cells. In the present study we have shown that the V_H , *D*, and J_H gene segments which are joined in four examples of normal B cells are drawn from the same pool of germ line gene segments used in

myelomas. In addition to functional *V-D-J* rearrangements, normal B cells contain nonfunctional rearrangements of at least two types: out-of-frame rearrangements, in which a termination codon or frameshift between *V* and *J* interrupts polypeptide translation, and incomplete rearrangements, in which a *D* and a J_H gene segment are joined without a V_H-D junction. The *D-J_H* rearrangement may be a transient stage which can later be completed by adding a V_H gene segment. Perhaps a heavy-chain chromosome could simultaneously contain $V_{H2}-D_1$ and D_2-J_{H1} rearrangements which may progress to a *V-D-J* rearrangement (in or out of frame) by "leap-frogging" from V_{H1} to D_2-J_{H1} or from $V_{H2}-D$ to J_{H2} .

The structures of the *D* gene segments isolated in the course of this study confirm the predictions of our "one-turn/two-turns" model of *V-D-J* joining (8). This was also shown to be the case by Tonegawa and co-workers, using *D-J_H* rearrangements from lymphoid tumors (14, 23). *D* gene segments related to B104 and D20 are evidently clustered between the M104E V_H gene segment and the J_H locus, supporting a straightforward picture of *V-D-J* joining between gene segments which are originally in the order V_H-D-J_H in the germ line.

Analyses of the pool of IgM⁺ BALB/c spleen cells by genomic Southern blots to the J_H locus have shown that virtually all heavy-chain chromosomes contain rearrangements at the J_H locus (or at a *D* gene segment located less than 1 kb 5' to the J_{H1} gene segment) (12, 18). Since only one of two chromosome homologs normally produces heavy chains ("allelic exclusion"; 4, 19), almost all normal B cells may contain a nonproductively rearranged heavy-chain chromosome, probably one of the types analyzed in this report. The close relationship between nonproductive rearrangements and the phenomenon of allelic exclusion has been reviewed elsewhere (7).

Evolution of *D* gene segments. In comparing the B104 *D-J* rearrangement and the D20 germ line *D* gene segment with various *V-D-J* rearranged genes, it became apparent that V_H gene segments contain a sequence near the 3' end of the coding region which is identical to part of the one-turn joining sequence of *D* gene segments. This is shown in Fig. 6 by comparing the underlined part of the coding sequence in B2, B38, B49, and S107 to the underlined one-turn joining sequence of the *D* gene segments directly below (B104, D20). The comparison of *D* gene segments with V_κ gene segments (which also have one-turn joining sequences 3' to the coding sequence) is even more striking: both the 7- and 9-bp conserved elements of the one-turn joining sequence 5' to *D* gene segments are closely

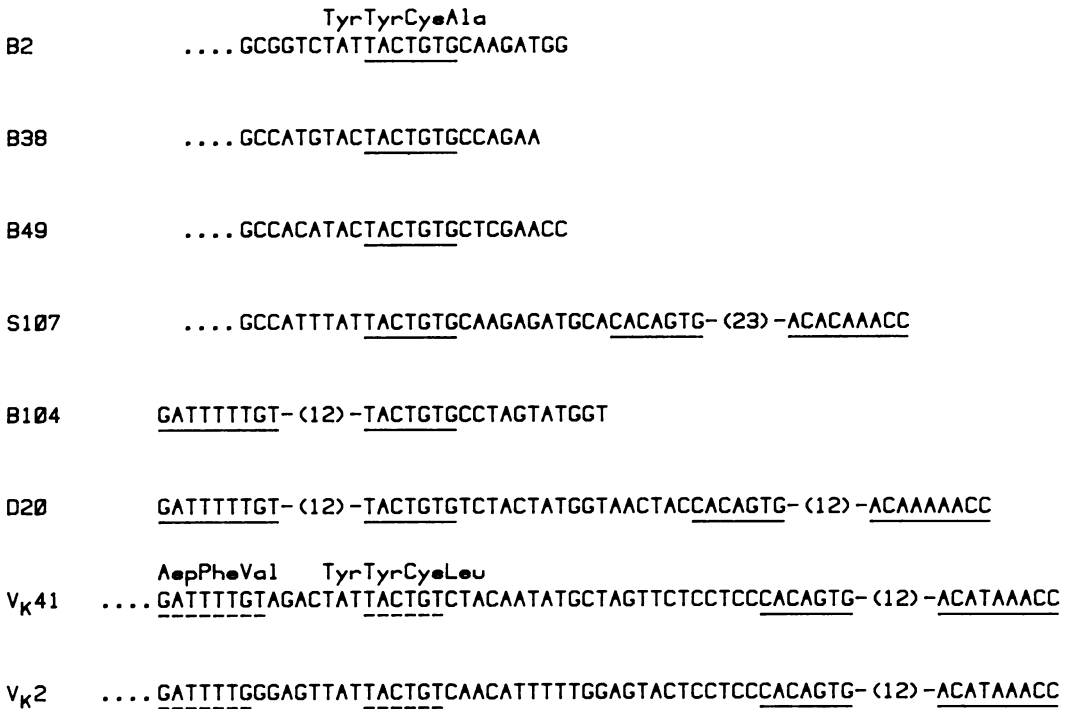


FIG. 6. Comparison of B104 and D20 *D* gene segments with *V_H* (B2, B38, B49, S107) and *V_K* (*V_K41*, *V_K2*) gene segments. The S107 *V_H* sequence is from reference 8, and the *V_K* sequences are from reference 16. Joining sequence in *V* and *D* gene segments and regions of identity (solid lines) or close homology (dotted lines) of *V* gene segments with the joining sequences 5' to *D* gene segments are underlined.

approximated by the *V_K* sequences with dotted underlines in Fig. 6.

One possible explanation for these observations is that *D* gene segments arose from *V* gene segments located on what is now the chromosome carrying the heavy-chain gene family. This model is illustrated in Fig. 7. A set of *V* gene segments with two-turn joining sequences (the arrow with a 2 above) might have been situated 5' to a gene family with a one-turn *V* gene segment and two-turn *J* gene segments. Occasionally, a two-turn *V* gene segment may have joined to an internal sequence (dotted arrow with a 1 below) in a one-turn *V* gene segment which resembled a one-turn site, like the internal sequences of the *V_K* gene segments in Fig. 6. This is shown on the top line of Fig. 7. "D"-*J* joining at the usual site in the one-turn *V* gene segment (second line in Fig. 7) could then have created a tripartite variable region like those found in heavy chains at present (bottom line in Fig. 7). Eventually, the original set of one-turn *V* gene segments may have evolved into a set of *D* gene segments which cannot form a functional gene without *V-D* joining. This view of the origin of the heavy-chain gene family may explain why

the short *D* gene segments appear to be widely dispersed in the genome, like *V* gene segments, rather than tightly clustered like *J* gene segments.

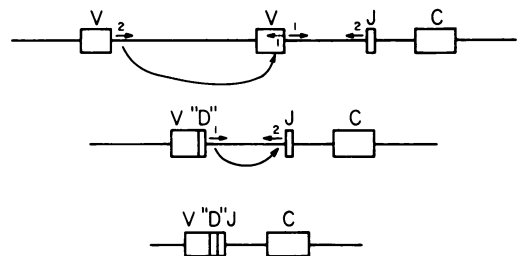


FIG. 7. Model for evolution of *D* gene segments. For simplicity, only a single gene segment of each type is shown. Joining sequences and their orientation are represented by short arrows, with spacer lengths (one or two turns) above the arrow. Joining might have occurred in this case between a two-turn *V* gene segment and an internal sequence in a one-turn *V* gene segment. If accompanied by joining between the one-turn *V* gene segment and a two-turn *J* gene segment, this may have generated the ancestor of the present tripartite *V_H* gene.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grants AI 16913 (to L.H.) GM 07276 (to I.W.), and AI 09072 (to I.W.) from the National Institutes of Health and by American Cancer Society grant IM 56 to I.W. P.E. was supported by a National Institutes of Health postdoctoral fellowship.

We thank Bob Coffman for antisera and Francis Assisi for FACS assistance.

LITERATURE CITED

- Blattner, F. R., B. G. Williams, A. E. Blechl, K. Denniston-Thompson, M. E. Faber, L.-A. Furlong, D. J. Grunwald, D. O. Kiefer, D. D. Moore, J. W. Schumm, E. L. Sheldon, and O. Smithies. 1977. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. *Science* **196**:161-169.
- Brack, C., M. Hiram, R. Lenhard-Schuller, and S. Tonegawa. 1978. A complete immunoglobulin gene is created by somatic recombination. *Cell* **15**:1-14.
- Cancro, M. P., W. Gerhard, and N. R. Klinman. 1978. The diversity of the influenza-specific primary B-cell repertoire in BALB/c mice. *J. Exp. Med.* **147**:776-787.
- Cebra, J. J., J. E. Colberg, and S. Dray. 1966. Rabbit lymphoid cells differentiated with respect to α -, γ -, and μ -heavy polypeptide chains and to allotypic markers A_{A1} and A_{A2}. *J. Exp. Med.* **123**:547-558.
- Davis, M. M., K. Calame, P. W. Early, D. L. Livant, R. Joho, I. L. Weissman, and L. Hood. 1980. An immunoglobulin heavy chain gene is formed by at least two recombinational events. *Nature (London)* **283**:733-739.
- Early, P. W., M. M. Davis, D. B. Kaback, N. Davidson, and L. Hood. 1979. Immunoglobulin heavy chain gene organization in mice: analysis of a myeloma genomic clone containing variable and α constant regions. *Proc. Natl. Acad. Sci. U.S.A.* **76**:857-861.
- Early, P., and L. Hood. 1981. Allelic exclusion and nonproductive immunoglobulin gene rearrangements. *Cell* **24**:1-3.
- Early, P., H. Huang, M. Davis, K. Calame, and L. Hood. 1980. An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: V_H, D and J_H. *Cell* **19**:981-992.
- 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 μ gene by alternative RNA processing pathways. *Cell* **20**:313-319.
- Hohn, B., and K. Murray. 1977. Packaging recombinant DNA molecules into bacteriophage particles *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **74**:3259-3263.
- Hood, L., M. Davis, P. Early, K. Calame, S. Kim, S. Crews, and H. Huang. 1981. Two types of DNA rearrangements in immunoglobulin genes. *Cold Spring Harbor Symp. Quant. Biol.* **45**:887-898.
- Hurwitz, J. L., C. Coleclough, and J. J. Cebra. 1980. C_H gene rearrangements in IgM-bearing B cells and in the normal splenic DNA component of hybridomas making different isotypes of antibody. *Cell* **22**:349-359.
- Kabat, E. A., T. T. Wu, and H. Bilofsky. 1979. Sequence of immunoglobulin chains. Publication no. 80-2008. National Institutes of Health, Bethesda, Md.
- Kurosawa, Y., H. von Boehmer, W. Haas, H. Sakano, A. Trauneker, and S. Tonegawa. 1981. Identification of D segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes. *Nature (London)* **290**:565-570.
- Liu, C.-P., P. W. Tucker, J. F. Mushinski, and F. R. Blattner. 1980. Mapping of heavy chain genes for mouse immunoglobulins M and D. *Science* **209**:1348-1353.
- Max, E. E., J. G. Seidman, and P. Leder. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin K constant region gene. *Proc. Natl. Acad. Sci. U.S.A.* **76**:3450-3454.
- Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
- Nottenberg, C., and I. L. Weissman. 1981. C _{μ} gene rearrangement of mouse immunoglobulin genes in normal B cells occurs on both the expressed and nonexpressed chromosomes. *Proc. Natl. Acad. Sci. U.S.A.* **78**:484-488.
- Pernis, B., G. Chiappino, A. S. Kelus, and P. G. H. Gell. 1965. Cellular localization of immunoglobulins with different allotypic specificities in rabbit lymphoid tissues. *J. Exp. Med.* **122**:853-876.
- Rabbitts, T. H., D. L. Bentley, W. Dunnick, A. Forster, G. E. A. R. Matthysseus, and C. Milstein. 1981. Immunoglobulin genes undergo multiple sequence rearrangements during differentiation. *Cold Spring Harbor Symp. Quant. Biol.* **45**:867-878.
- Riblet, R. J. 1977. The organization of antibody genes. *ICN-UCLA Symp. Mol. Cell. Biol.* **6**:83-89.
- Sakano, H., K. Huppi, G. Heinrich, and S. Tonegawa. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature (London)* **280**:288-294.
- Sakano, H., Y. Kurosawa, M. Weigert, and S. Tonegawa. 1981. Identification and nucleotide sequence of a diversity DNA segment (D) of immunoglobulin heavy-chain genes. *Nature (London)* **290**:562-565.
- Sakano, H., R. Maki, Y. Kurosawa, W. Roedel, and S. Tonegawa. 1980. Two types of somatic recombination are necessary for the generation of complete immunoglobulin heavy-chain genes. *Nature (London)* **286**:676-683.