Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 4405–4409, July 1982
Immunology

Organization and evolution of immunoglobulin V_H gene subgroups

(deletion mapping/gene expansion/sequences of human V_HIII genes)

GIDEON RECHAVI, BRIGITTA BIENZ, DANIELA RAM, YINON BEN-NERIAH, JUSTUS B. COHEN, RINA ZAKUT, AND DAVID GIVOL

Department of Chemical Immunology, The Weizmann Institute of Science, Rehovot 76100, Israel

Communicated by C. B. Anfinsen, April 26, 1982

ABSTRACT **The organization and evolution of immunoglobulin variable region genes was studied by comparing human and mouse heavy chain variable region (V_H) genes. We show that a V_H gene subgroup constitutes a physically linked multigene family separated from another V_H subgroup. We mapped the V_HIII gene subgroup to be 3′ to the V_HII gene subgroup based on deletion of V_H genes after V–D–J rearrangement. The results indicate that the human V_HIII gene subgroup underwent a significant gene expansion as compared to the mouse V_HIII subgroup. Amino acid sequence data indicate that human V_HIII genes correspond to only a small subset of mouse V_HIII genes. Human V_HIII genes contain a shorter intron and are two codons shorter than most BALB/c mouse V_HIII genes. The nature of nucleotide substitutions between V_H genes within a species (human) is similar to that between genes of different species (human/mouse). Both contain approximately 50% silent substitutions.**

Both somatic and genetic ("germ line") factors operate in generating antibody diversity (1–3). The somatic processes probably act sequentially: first, by combinatorial joining of two or three gene segments, which is accompanied by junctional variation (4–6), and second, by somatic mutations, which take place after the switching process from IgM expression to other immunoglobulin classes (2, 7, 8). However, it is clear that the diversity of germ line heavy chain variable region and light chain variable region (V_H and V_L, respectively) genes is extensive and is accumulated throughout evolution (1, 9). Amino acid sequence data indicated that the variable regions of immunoglobulins can be divided into subgroups according to their extent of homology (10–13). A similar subdivision can be made on the level of the genes on the basis of cross hybridization of different V genes with various cDNA probes in which hybridization takes place with distinct sets of V genes (1, 14–17). Subgroups of genes defined by cross hybridization do not always correspond to subgroups defined by protein sequence (17), but in many cases—depending on the conditions of hybridization—there is a good correlation between the two (15).

The evolution of multigene families like the V gene subgroups may lead to gene expansion or contraction (11, 18), thereby modifying the germ line repertoire of V genes. Comparison of V_H gene subgroups from different species may increase our understanding of the evolutionary factors operating in generating antibody diversity. We studied some aspects of the organization and evolution of the V_HIII subgroup by comparing human and mouse V_H genes. We show by deletion mapping that V_HIII genes are located 3′ to V_HII genes on the mouse chromosome and that the two subgroups are not intermingled with each other. Our sequence analysis data of human V_HIII genes suggest that they correspond to a particular subset of mouse V_HIII genes. However, the V_HIII gene subgroup in human contains significantly more genes than does V_HIII in BALB/c mice. The sequence comparison also demonstrates that the proportion of silent substitutions in V_H genes within a species is similar to that present between genes of different species.

MATERIALS AND METHODS

Southern Blot Hybridization. DNA was isolated from BALB/c mouse liver or myeloma and from human placenta, as described by Blin and Stafford (19). DNA was digested to completion by *Eco*RI, fractionated on 0.7% agarose gels, transferred to nitrocellulose filters, and hybridized to ^{32}P-labeled probes as described (20, 21). Final wash after hybridization was in 150 mM NaCl/15 mM Na citrate/0.1% $NaDodSO_4$ at 65°C. Four probes were used: (*i*) *Pst* I fragments of recombinant plasmid pS107α.4 kindly donated by J. Adams (22). This insert contains sequences corresponding to V_H107 (mouse V_HIII subgroup) and to the constant region (*C*) gene; (*ii*) *Bam*HI fragment [0.92 kilobase (kb)] containing a mouse germ line V_HII gene denoted *pCh104* that was described previously (9); (*iii*) an *Eco*RI fragment (2.3 kb) containing a human germ line V_HIII gene denoted *H11*; (*iv*) a *Bam*HI fragment (1.6 kb) containing a human V_HII gene denoted *HA2*.

Isolation of Human V_HIII Genes. A human fetal liver gene library kindly donated by T. Maniatis (23) was screened by plaque hybridization to the mouse pS107α.4 (24). Cross hybridization between mouse and human V_H genes allowed the isolation of several recombinant phages containing V_H genes. Restriction endonuclease mapping of purified phages was performed with *Eco*RI, *Bam*HI, *Kpn* I, *Bgl* II, and *Hin*dIII and by filter hybridization with pS107α.4. Reasonably small fragments (2–3 kb) that gave positive hybridization were eluted from the gel and subjected to DNA sequence analysis.

DNA Sequence Analysis. This was performed mainly by cloning and sequencing in the M13 mp8 phage vector (25). DNA fragments eluted from agarose were subjected to sonication, site filling by DNA polymerase I (Klenow), size fractionation on agarose, subcloning by blunt end ligation into the *Sma* I site of M13 mp8, and transformation of JM101 bacteria. Recombinant M13 mp8 phages (white plaques) were plated onto a lawn of JM101 bacterial host, transferred to nitrocellulose, and hybridized to ^{32}P-labeled pS107α.4. Positive plaques were picked, added to 1 ml of JM101, and grown for 5 hr, and the DNA was prepared and subjected to sequence analysis as described (26). In some cases sequence analysis by the Maxam and Gilbert method (27) was also used.

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

Abbreviations: V_H, heavy chain variable region; V_L, light chain variable region; HV, hypervariable region; C, constant region; kb, kilobase(s); bp, base pair(s).

RESULTS

V_HIII Gene Subgroup Is Mapped 3′ to V_HII Gene Subgroup. The V–D–J recombination (6, 28) of heavy chain genes in antibody-producing cells results in a deletion of those V_H genes present between the two recombining gene segments (14, 29). This is not always easy to observe in a myeloma that is diploid for chromosome 12, which contains the heavy chain genes, because the V_H genes in the nonfunctional chromosome will show up in a Southern blot. We took advantage of the situation in mouse myeloma MOPC104E, which is haploid for chromosome 12 (30) and produces an IgM containing a V_HII gene product (31). Fig. 1 *Upper* shows a Southern blot of DNA from BALB/c liver and MOPC104E tumor analyzed with probes containing either a V_HII gene or V_HIII gene. It is shown that the entire group of V_HIII genes is missing in MOPC104E DNA. This suggests that all genes of the V_HIII subgroup are located at the 3′ side of the V_HII genes because they are deleted by the rearrangement of a V_HII gene. On the other hand, there is no clear difference between the hybridizing bands detected by the V_HII probe in liver and myeloma DNA, suggesting that most of the V_HII genes are still present in MOPC104E DNA after the recombination event. Hence, the expressed gene in this tumor is probably located at the 3′ side of this gene subgroup, as indicated in Fig. 1 *Lower*. The result of this blot suggests that V_HIII genes are separate from V_HII genes and that closely related V_H genes (a subgroup) are physically linked.

Gene Expansion of V_HIII Subgroup in Man. We have used either a mouse probe (pS107α.4) or a human probe (an *Eco*RI fragment of Ch H11, containing a germ line gene) to detect V_HIII genes in mouse and human DNA. The results (Fig. 2) show that there is no significant difference between the number of hybridizing bands detected by the two probes in each DNA. However, the number of bands detected in human DNA is significantly larger, suggesting a marked change in the number of germ line V_HIII genes between BALB/c mouse and man. Four of the genes located in the hybridizing bands in mouse DNA were isolated and subjected to sequence analysis (2), and they probably comprise most of the genes of this subgroup in mouse. The sequences showed that they are all similar to the *T15* V_H gene coding for the V_H of the antiphosphoryl choline antibody (2, 13). To compare human V_HIII genes to the mouse V_HIII we isolated several human genes using the plasmid S107α.4 insert as a probe and determined their sequences.

Comparison of Human and Mouse V_HIII Gene Sequences. Several recombinant phages were isolated from the human gene library by screening with pS107α.4. Mapping of these clones showed that three of them contained two genes in their insert and three others contained only one gene. The mapping also showed that two phages contain an overlapping fragment and

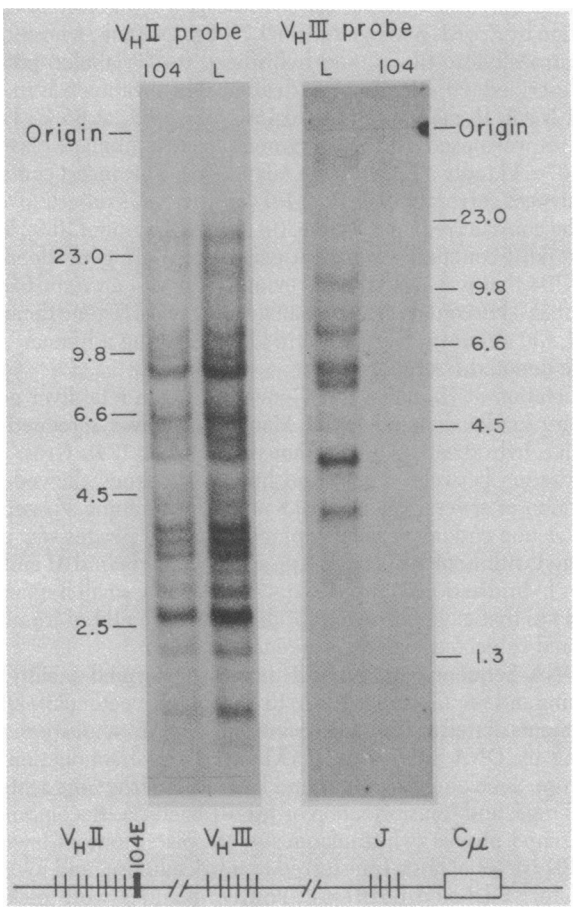

FIG. 1. Detection of V_HII and V_HIII genes in BALB/c mouse liver and plasmacytoma MOPC104E DNA. (*Upper*) Southern blots of *Eco*RI-digested DNA from plasmacytoma MOPC104E (104) or liver (L) were hybridized with the nick-translated insert of either HA2 (V_HII probe) or H11 (V_HIII probe). See *Materials and Methods* for the conditions of hybridization. Size markers in kb are *Eco*RI/*Bam*HI-digested Ch4A phage. (*Lower*) Schematic representation of the order of V_H gene subgroups II and III as deduced from this analysis (see text).

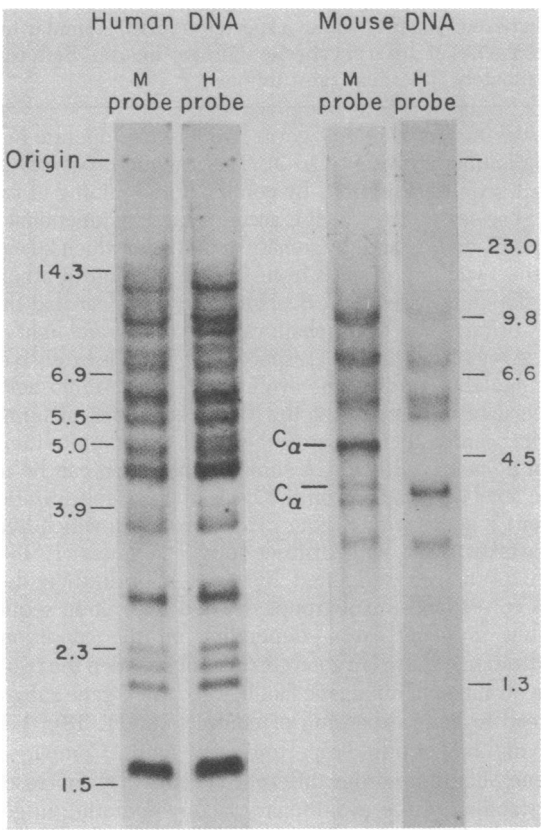

FIG. 2. Detection of V_HIII genes in either BALB/c mouse liver DNA or human placenta DNA. Southern blots of *Eco*RI-digested DNA were hybridized with a nick-translated *Pst* I fragment of mouse pS107α.4 (M probe containing the V_HIII sequence) or with a human germ line V_HIII gene H11 (H probe). See *Materials and Methods* for conditions of hybridization. The bands corresponding to the C_α genes detected by pS107α.4 were identified by a separate hybridization with a *Pvu* II fragment of Ps107α.4 that contains only a constant region sequence (22). Size markers in kb were *Eco*RI/*Bam*HI digested Ch4A (shown on left) or *Hin*dIII-digested phage DNA (shown on right).

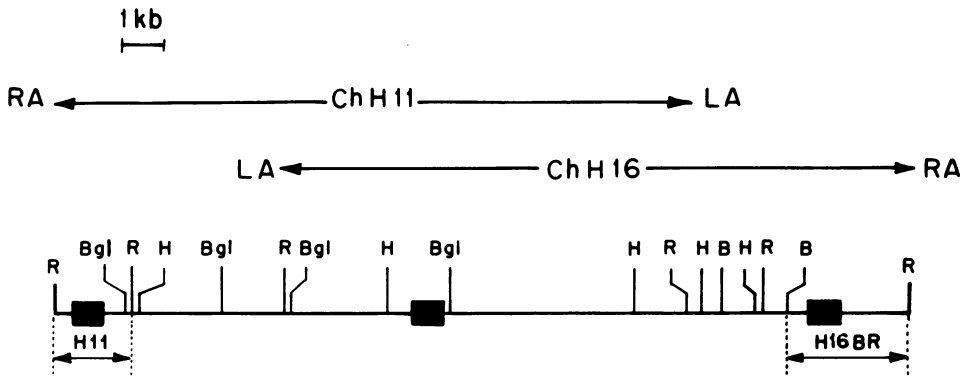

FIG. 3. Restriction endonuclease and linkage map of two recombinant phages containing $V_H III$ genes. The restriction maps were derived by a combination of single- and double-restriction enzyme digestion and Southern filter hybridization. Both phages contain an identical 8.5-kb *Eco*RI fragment. The *Eco*RI fragment H11 and the *Bam*HI/*Eco*RI fragment H16BR were selected for sequence analysis of their V_H genes. RA and LA are the right and left arms, respectively, of the recombinant phages. R, *Eco*RI; Bgl, *Bgl* II; H, *Hin*dIII; B, *Bam*HI. The black rectangles indicate the positions of V_H genes.

constitute a physically linked group of three V_H genes separated by spacers of approximately 8 kb (Fig. 3). Two of these V_H genes, *H11* located on an *Eco*RI fragment of 2.3 kb and *H16BR* located on a *Bam*HI/*Eco*RI fragment of 2.0 kb (Fig. 3), were subjected to sequence analysis.

Fig. 4 compares these sequences with that of $V_H 26$, another human $V_H III$ gene (32), and two mouse $V_H III$ genes—*T15* (2) and $V_H 441$ (33)—that were recently reported. The comparison shows that the three human genes have a similar structure which differs from that of the mouse *T15* in two aspects: (*i*) *T15* has two additional codons, AAA and GCT, between codons 53 and 54 in the second hypervariable region (HV2) coding for Lys-Ala. (*ii*) *T15* V_H gene has a 56-base pair (bp) longer intron between the leader and the V_H segment. On the other hand, the mouse $V_H 441$ resembles the human $V_H III$ in its general structure because its intron (101 bp) is similar in size to that of the human $V_H III$ genes (102 bp) and it lacks the two additional codons in HV2. In spite of this structural homology, there is no significant difference in the percentage of homology between human *H11* and either *T15* (75%) or $V_H 441$ (76%). The V_H sequences also show that *H16BR* contains a termination codon TGA at position 9. In addition, two codons for cysteine are present at unusual positions: TGC at position 38 and TGT at position 53. Because of the termination codon, the gene *H16BR* may be considered a pseudogene. We previously showed that the immunoglobulin V_H gene repertoire is rich in pseudogenes (9). From four different human and five different mouse V_H genes that we subjected to sequence analysis at random, two and two, respectively, contain termination codons and may be considered pseudogenes. However, they preserve the general structure of a V_H gene and might be corrected by recombination with other V_H genes (3). It is also possible that a pseudogene will drift by accumulating further mutations. The two cysteine codons shown here were not found in any of the expressed V_H proteins and may be an example of such mutation. The comparison of substitutions between human and mouse genes is shown in Table 1. It is shown that the percentage of silent substitutions in genes within a species and between species is quite similar and is close to 50%.

DISCUSSION

We studied some aspects of the organization and evolution of $V_H III$ genes by comparing human and mouse genes. Our results indicate that a V_H gene subgroup is a multigene family that shows a physical linkage. Three human $V_H III$ genes on two separate recombinant phages were mapped and found to be linked with spacers of approximately 8 kb (Fig. 3). Based on deletion of V_H genes upon V–D–J recombination (6, 29), we show that the two V_H subgroups, $V_H II$ and $V_H III$, are physically separated and that their genes are not intermingled with one another. This is concluded because after V–D–J rearrangement in MOPC104E that expressed a $V_H II$ gene, almost all subgroup $V_H II$ genes are present in the myeloma genome, whereas all subgroup $V_H III$ genes are deleted. Our results also locate $V_H III$ genes 3′ to $V_H II$ genes. Similar results were obtained by Adams *et al.* (14), who used other lymphoid cell lines and other probes. The deletion mapping approach may be useful to map the entire V_H gene locus.

It was shown earlier that $V_H III$ genes in the mouse—as detected by the pS107α.4 probe—comprise a very limited group of genes (2). We demonstrated that the same probe can detect at least 3-fold more genes in human DNA. Such gene expansion, or contraction, was suggested to be an important factor in the evolution of multigene families (11, 18). This may contribute to antibody diversity as it provides a mechanism that changes the repertoire of V genes. It is noteworthy that the proportion of $V_H III$ in serum immunoglobulins varies considerably in mammals (34), and it will be interesting to analyze the number of $V_H III$ genes in ruminants in which no immunoglobulins that contain $V_H III$ region were found (34).

Four of the mouse $V_H III$ genes, which constitute most of this subgroup in the BALB/c mouse, were subjected to sequence analysis and found to have a general structure like the *T15* V_H gene. The immunoglobulins of mouse $V_H III$ subgroup (13) include also a group of antigalactan antibodies (X44, X24, J539, and T601) and a group of antiinulin antibodies (ABPC47N, EPC109, ABPC4, and UPC61). One V_H gene, presumably coding for antigalactan antibodies ($V_H 441$), was recently subjected to sequence analysis (33) and it differs by 28% from the *T15* V_H gene. One can see from its structure that it may represent a subset of the $V_H III$ subgroup because it contains a shorter intron and lacks two codons in *HV2* (Fig. 4) in comparison with the *T15* V_H gene. The sequence comparison (Fig. 4) indicates that the human $V_H III$ genes isolated so far are more closely related to mouse $V_H 441$ than to *T15* V_H. Although their extent of homology to either $V_H 441$ or *T15* V_H is similar, their general structure is more similar to $V_H 441$ because the intron size is similar and the HV2 of the human $V_H III$ genes lacks the two additional codons. This may hold for most human $V_H III$ genes because of 18 human V_HIII proteins that were subjected to sequence analysis (13), none contains the 2 additional amino acids Lys52B-Ala52C present in HV2 of the *T15* genes. This suggests that most human

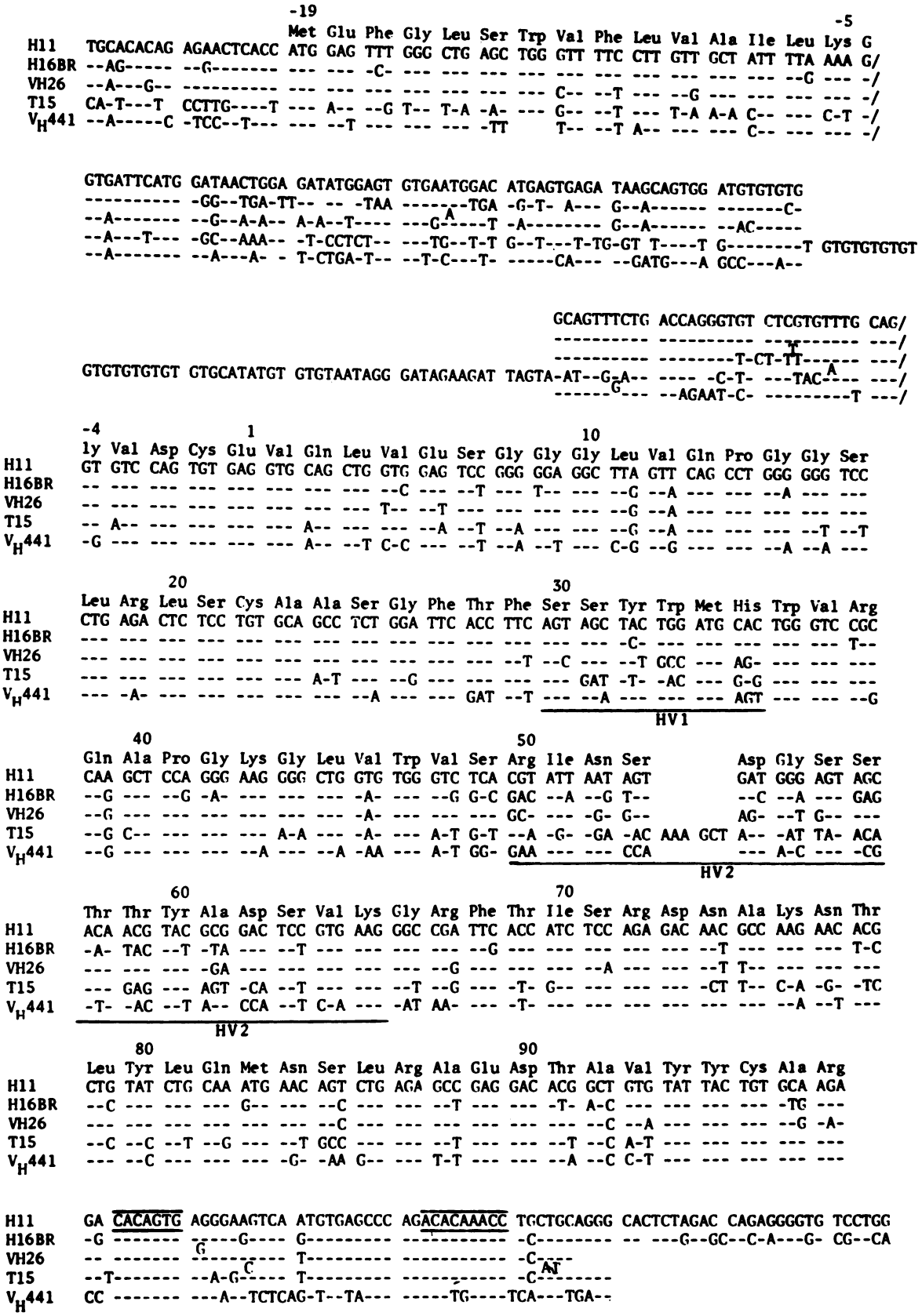

FIG. 4. Comparison of human and mouse $V_H III$ gene sequences. Human genes are *H11, H16BR* (this work), and $V_H 26$ (32). Mouse genes are *T15* V_H (2) and $V_H 441$ (33). Only the coding strand sequence is shown and the predicted amino acids of *H11* are given above the line. Sequences in noncoding regions were aligned to maximize homology and include a small number of deletions (empty space) or insertions (a letter above the line). *T15* contains a longer intron and two additional codons between codon 53 and 54 (AAA and GCT). Hypervariable regions HV1 and HV2 are underlined. Recombination signals at the end of V_H are boxed.

Table 1. Substitutions in human and mouse V_H genes

Genes compared	Total substitution, %	Silent substitutions, % of total
Human *H16BR*/human *H11*	15.4	51.0
Human V_H*26*/human *H11*	10.8	45.5
Mouse *T15*/human *H11*	25.0	45.5
Mouse V_H*441*/human *H11*	24.0	50.5
Mouse *T15*/mouse V_H*441*	28.0	40.0

* Substitutions were compared in the coding region (codons −4 to 98).

V_H*III* genes and the mouse V_H*441* evolved from a common ancestor in the past which duplicated in the human genome more extensively than in the mouse genome. On the other hand, the mouse *T15* V_H-like genes may have evolved from a different ancestry, the counterpart of which in human genes has not yet been found. In another study we also compared human and mouse V_H*II* genes and found that they contain the same size intron (unpublished). This supports our suggestion that in general the intron size in V_H genes is being preserved within a subgroup.

The nature of nucleotide substitutions in V_H genes is shown in Table 1. The homology between the human V_H genes is 85–90%, whereas between mouse *T15* V_H and V_H*441* it is only 72%. This homology is lower than that between mouse and human genes (75–76%) in this subgroup, supporting our suggestion that the mouse V_H*III* gene subgroup can be subdivided into subsets. Table 1 shows that the percentage of silent substitutions in V_H genes either within a species (human) or between species (human/mouse) is very similar and is close to 50%. This perhaps suggests a long evolutionary process and indicates some selection against replacement substitutions. On a random basis, 25% of the substitutions should be silent, but homologous genes in different species show a much higher percentage of silent substitutions due to selection. It was argued recently that there is an unusually high percentage of replacement substitutions in immunoglobulin constant region genes as compared to other genes (35). Our results show that this is not the case for the V_H genes. This may suggest that during evolution the structure of a *V* gene is more preserved than that of the *C* genes. A possible explanation for this may be that selection operates on *V* genes to preserve complementarity to antigen. Such structural requirements may not differ significantly between mouse and human. On the other hand, *C* genes diverge with respect to effector functions which may differ considerably between species.

The main conclusion from our study is that V_H genes of the same subgroup constitute a physically linked multigene family that preserves significant sequence homology and other characteristics such as the size of the intron. The linkage between three genes (Fig. 3) and the results of deletion mapping suggest that *V* gene subgroups evolve by gene duplication and keep their linkage throughout evolution.

We thank Dr. Stan Fields for his help in the M13 cloning sequencing technique, Dr. T. Maniatis for the human gene library, and Dr. J. Adams for plasmid pS107. The excellent technical assistance of Mrs. S. Wylder and Mrs. S. Hazum is gratefully acknowledged.

1. Seidman, J. G., Leder, A., Nau, M., Norman, B. & Leder, P. (1978) *Science* **202**, 11–17.
2. Crews, S., Griffin, J., Huang, H., Calame, K. & Hood, L. (1981) *Cell* **25**, 59–66.
3. Baltimore, D. (1981) *Cell* **26**, 295–296.
4. Brack, C., Hirama, M., Lenhard-Schuller, R. & Tonegawa, S. (1978) *Cell* **15**, 1–14.
5. Seidman, J. G., Max, E. E. & Leder, P. (1979) *Nature (London)* **280**, 370–375.
6. Early, P., Huang, M., Davis, M., Calame, K. & Hood, L. (1980) *Cell* **19**, 981–992.
7. Bothwell, A. L. M., Paskind, M., Reth, M., Imanishi-Kari, T., Rajewsky, K. & Baltimore, D. (1981) *Cell* **24**, 625–637.
8. Kim, S., Davis, M., Sinn, E., Patten, P. & Hood, L. (1981) *Cell* **27**, 573–581.
9. Givol, D., Zakut, R., Effron, K., Rechavi, G., Ram, D. & Cohen, J. B. (1981) *Nature (London)* **292**, 426–430.
10. Milstein, C. (1967) *Nature (London)* **216**, 330–332.
11. Milstein, C. & Pink, J. R. L. (1970) *Prog. Biophys. Mol. Biol.* **21**, 211–263.
12. Smith, G. P., Hood, L. & Fitch, W. M. A. (1971) *Annu. Rev. Biochem.* **40**, 969–1012.
13. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1979) *Sequences of Immunoglobulin Chains* (Natl. Inst. Health, Bethesda, MD), Publ. 80-2008.
14. Adams, J. M., Kemp, D. J., Bernard, O., Gough, N., Webb, E., Tyler, B., Gerondakis, S. & Cory, S. (1981) *Immunol. Rev.* **59**, 5–32.
15. Cory, S., Tyler, B. M. & Adams, J. M. (1981) *J. Mol. Appl. Genet.* **1**, 103–116.
16. Rabbitts, T. H., Matthyssens, G. & Hamlyn, P. (1980) *Nature (London)* **284**, 238–243.
17. Bentley, D. L. & Rabbitts, T. H. (1981) *Cell* **24**, 613–623.
18. Hood, L., Campbell, J. H. & Elgin, S. C. R. (1975) *Am. Rev. Genet.* **9**, 305–353.
19. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3**, 2303–2308.
20. Southern, J. (1975) *J. Mol. Biol.* **98**, 503–517.
21. Ben-Neriah, Y., Cohen, J. B., Rechavi, G., Zakut, R. & Givol, D. (1981) *Eur. J. Immunol.* **11**, 1017–1022.
22. Adams, J. M., Gough, N. M., Webb, E. A., Tyler, B. M., Jackson, J. & Cory, S. (1980) *Biochemistry* **19**, 2711–2719.
23. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connel, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15**, 687–701.
24. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
25. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Rol, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
27. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
28. Sakano, H., Kurosawa, Y., Weigert, M. & Tonegawa, S. (1981) *Nature (London)* **290**, 562–565.
29. Cory, S. & Adams, J. W. (1980) *Cell* **19**, 37–51.
30. Coleclough, C., Perry, R. P., Karjalainen, K. & Weigert, M. (1981) *Nature (London)* **290**, 372–378.
31. Zakut, R., Cohen, J. B. & Givol, D. (1980) *Nucleic Acids Res.* **8**, 3591–3601.
32. Matthyssens, G. & Rabbitts, T. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6561–6565.
33. Ollo, R., Auffray, C., Sikorav, J.-L. & Rougon, F. (1981) *Nucleic Acids Res.* **9**, 4099–4107.
34. Capra, D. J., Wasserman, R. L. & Kehoe, M. J. (1973) *J. Exp. Med.* **138**, 410–427.
35. Sheppard, H. W. & Gutman, G. A. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7064–7068.