Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 7767–7771, October 1990
Immunology

Characterization of a complete immunoglobulin heavy-chain variable region germ-line gene of rainbow trout

(immunoglobulin gene evolution)

TAKESHI MATSUNAGA*†, THOMAS CHEN‡, AND VUOKKO TÖRMÄNEN*

*Unit for Applied Cell and Molecular Biology, University of Umeå, Umeå, Sweden; and ‡Center of Marine Biotechnology, University of Maryland System, Baltimore, MD 21202

Communicated by Susumu Ohno, June 18, 1990

ABSTRACT **A germ-line heavy-chain variable region (V_H) gene (RTVH431) has been isolated from a rainbow trout (*Salmo gairdneri*) and characterized by complete nucleotide sequencing. It is characteristic of V_H, as shown by the conserved octamer and TATA motif in the 5′ region, the heptamer–nonamer recombination signal sequence in the 3′ region, and the 18-amino-acid-long hydrophobic leader interrupted by an intron. The 98-amino-acid-long V_H coding region has 50–70% nucleotide sequence homology and 40–60% amino acid sequence homology with V_Hs of various vertebrate species. We have also found unique or species-specific amino acid residues in the V_Hs of rainbow trout, amphibia (*Xenopus*), reptile (*Caiman*), and shark (*Heterodontus*) in our sequence analyses. The RTVH431 has an unusual amino acid in the conserved 34th position in complementarity-determining region 1 of V_H. Southern hybridization results suggest the presence of a large gene family related to RTVH431 in the trout genome. The complex evolution of antibody V genes is discussed.**

By mid-1980, molecular biology revealed genetic mechanisms to explain how the immune system generates an enormous repertoire of antibody specificity (1). The multiple germ-line gene elements and "somatic diversifiers" are exploited maximally by a variety of mechanisms: variable region heavy-chain–light-chain (V_H–V_L) as well as V–(D)–J (D, diversity; J, joining) recombination, joining imprecision, random insertion (N diversity), and somatic hypermutation (1–3). However, since early immunology was an offshoot of the field of infectious medicine, our knowledge about the century-old enigma of antibody repertoire (4, 5) has been derived almost exclusively from two mammalian species—mouse and human. Important questions remain: How have these multitudes of antibody genes and complex mechanisms emerged and evolved in vertebrate history, and how has the antibody repertoire been selected during evolution (6, 7)?

A few recent studies on two vertebrate species from divergent phyla—bird (8) and shark (9)—showed remarkable features not seen in mammals with regard to immunoglobulin gene organization and the mechanism for antibody repertoire generation. It also has been pointed out previously that cold-blooded vertebrates, such as amphibia and fish, have an antibody repertoire of small size (10) and that somatic hypermutation in the strict sense may be a phenomenon of only warm-blooded vertebrates (11). The teleost, although still primitive in vertebrate phylogeny, represents an advanced form of bony fish, many of which have been an important food resource since the prehistoric era. Thus, knowledge of the teleost antibody gene will contribute not only to our understanding of evolution of the immune system, but it should be helpful in an eradication program for infectious diseases that have been detrimental to modern aquatic culture (12). In this article, we describe the complete nucleotide sequence of a V_H germ-line gene isolated from rainbow trout, *Salmo gairdneri* (or *Oncorhynchus mykiss*), and the complexity of this V_H family in the genome.§ We also discuss some aspects of immunoglobulin V gene evolution in vertebrates.

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.

MATERIALS AND METHODS

Isolation of Immunoglobulin V_H Genomic Clones and DNA Sequencing. The trout genomic library was constructed with testis DNA from a single rainbow trout (13). The DNA was partially digested by *Eco*RI and cloned in Charon 4A phage. The DNA probe used for screening the library was mouse V_H S107 cDNA and a 1.2-kilobase (kb) V_H pseudogene clone from another teleost, *Elops saurus* (a kind gift from G. Litman). The probe DNA was labeled with ^{32}P by using random primers (14). Clones were initially identified by hybridization under relaxed conditions (55°C, 6× SSC; ref. 15). The positive clones from the primary screening were then transferred to bacterial plates and the filters were rehybridized under more stringent conditions (65°C, 6× SSC).

DNA clones that gave a consistent hybridization signal were further digested by several restriction enzymes, and a short DNA fragment that hybridizes with the probe was subcloned into plasmid pBluescript (purchased from Stratagene) for DNA sequencing. DNA sequencing was done according to the method of Sanger (16), and using oligonucleotides as primers.

Southern Blot Hybridization. About 20 μg of high molecular weight genomic DNA isolated from livers of eight individual rainbow trout was digested to completion by *Eco*RI restriction enzymes, electrophoresed in 0.8% agarose gel, and blotted onto nylon filters. The RTVH431 insert was cut out and labeled with ^{32}P by using random primers, and hybridization was performed under relatively stringent conditions (probe, 5×10^6 cpm per ml of hybridization solution; hybridization, 65°C, 6× SSC; washing, 60°C, 2× SSC). The filters were exposed to x-ray films for several days.

RESULTS AND DISCUSSION

Isolation and Some Features of a Rainbow Trout V_H Germ-Line Gene. In our initial screening of ≈200,000 plaques of the rainbow trout library by mouse V_H probe (S107 cDNA), we failed to isolate DNA clones that gave a consistent hybrid-

Abbreviations: V_H, variable region of immunoglobulin heavy chain; V, variable region of immunoglobulin; D, diversity region of immunoglobulin; J, joining region of immunoglobulin.

‡To whom reprint requests should be addressed at: Unit for Applied Cell and Molecular Biology, University of Umeå, Umeå, S 901 87, Sweden.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M37206).

ization signal. However, we could obtain several recombinant phage DNA clones that hybridize with the *Elops* V_H probe from screening ≈190,000 plaques. One of them (RTVH431) has been fully characterized by nucleotide sequencing (see legend to Fig. 1).

Fig. 1 shows the DNA sequence of RTVH431, a germ-line V_H gene of a rainbow trout. An open reading frame encodes the 18-amino-acid-long hydrophobic leader peptide, which is interrupted by a 116-base-pair (bp) intron, and the region that specifies the 98-amino-acid-long V_H protein. The 3′ end of the V_H-coding region is flanked by the heptamer–22-bp spacer–nonamer, a signal used for recombination of V_H and probably D_H. The TATA motif is found in the region 5′ to the leader-encoding region. These structural features are common to all vertebrate V_H genes (1). The conserved octamer (17) ATGCAAAG (prototype, ATGCAAAT; one variant ATGCAAAG is cited in a human V_H; see ref. 17) found in the 5′ to TATA motif is the enhancer–promotor signal for B-cell-specific V_H expression (18, 19). The conserved octamer is found in V_H genes not only of mice and humans, but also of reptiles (*Caiman crocodylus*) (20), amphibia (*Xenopus*) (21), and two teleosts, *Elops* (22) and goldfish (*Carassius*) (23). The sole exception is the V_H of horned shark (*Heterodontus*) (24), a class of elasmobranch. *Heterodontus* organizes its V_H genes as a V_H–D_H–J_H–C_H (C, constant) multiple unit in the genome (9) rather than clustering V, D, and J segments and rearranging to the unique constant region gene, as seen in mammals (1). It is curious to note that the conserved octamer is always present in the primitive vertebrate species that have "the mammalian type" V_H organization; i.e., *Xenopus* (21), *Elops* (22), and probably also *Caiman* (20). The octamer binding proteins are perhaps mechanistically involved in DNA rearrangement in the mammalian-type immunoglobulin organization. The rainbow trout, which also has the octamer, may therefore have this type of V_H organization. Very recently, catfish (*Ictalurus punctatus*), another teleost, has been shown to have the mammalian-type V_H organization (25). The mammalian type might have evolved from the shark-type organization after elasmobranch–teleost divergence and it has become the major type in vertebrates (24). Another possibility is that the mammalian-type immunoglobulin organization originated from the T-cell antigen receptor α/β gene, and the shark-type organization is a secondary adaptation (26).

Nucleotide Sequence Comparison. A nucleotide homology search in the EMBO/GenBank data base (CDEMB 19A) scored the best-matched 25 sequences of 22,130 sequences listed, and all 25 sequences turned out to be members of V_H: 12 human, 6 mouse, 1 rabbit, 2 *Xenopus*, 2 *Caiman*, and 2 goldfish. The 5′ upstream region (310–320 bp long) between the octamer and the −1 position of the leader peptide is composed largely of introns (Fig. 1). This region has a moderate but significant homology with the corresponding region of other vertebrate V_Hs; e.g., 49.8% human V_H HG3 (27), 48.0% *Caiman* (20), 54.5% goldfish 5A, 54.1% goldfish 3 (23).

DNA sequence identity within the V_H coding region (+1 to 98) is around 60–70%; e.g., human germ-line V_H H11 subgroup III (33), 68.2%; HG3 subgroup I (27), 62.5%; 71.4 subgroup IV (28), 63.2%; mouse germ-line V_H 283 subgroup IIID (29), 67.7%; B10.P T15 subgroup III (30), 64.3%; rabbit germ-line V_H clone P26.9B1 (31), 64.6%; *Caiman* V_H (20), 64.5%; *Xenopus* V_H (21), 71.0%; goldfish V_H 3 (23), 62.1%. *Heterodontus* and catfish V_H NG70 (25) are less homologous—56.6% and 52.6%, respectively. The *Elops* V_H4 pseudogene used as a probe to isolate RTVH431 has 68.8% identity, whereas the mouse probe S107 (32) has ≈56% identity. Thus, the varied phylogenetic distance among the V_Hs compared seems reflected neither in the difference of V_H nucleotide sequences nor in amino acid sequence differences (see below), nor is there closer homology to particular subgroups of human V_H.

Amino Acid Sequence Conservation and Unique Residues. Fig. 2 shows predicted V_H amino acid sequence of rainbow trout aligned with other vertebrate V_Hs [numbering system according to Kabat (34)]. The overall amino acid identity of rainbow trout RTVH431 with other V_Hs in Fig. 2 is 57.3% (catfish), 52.5% (goldfish 3), 45.5% (*Heterodontus*), 48.5% (*Xenopus*), 54.5% (*Caiman*), 58.6% (human, H11), and 56.6% (mouse, V_H 283).

The conservation of many residues is clear in the framework regions, which compose a β-pleated structure for the "immunoglobulin fold" (35). Residue positions that are buried in such a β-sheet in mouse and human V_Hs (35, 36) are occupied by residues such as L and M (position 4); E and Q (position 6); G, A, V, and M (position 12); V and L (position 18); L, I, and M (position 20); invariant C (position 22); G, A, S, T, and V (position 24); V, L, I, and M (position 34); invariant W (position 36); R and K (position 38); V, L, I, and W (position 48); G and A (position 49); G, A, V, I, M, and F (position 69); A, L, F, and Y (position 78); L and M (position 80); L, I, and M (position 82); G, A, and S (position 88); Y (position 90); invariant C (position 92). Residue positions involved in V_H–V_L interdomain contact are occupied by V at position 37, Q at position 39, L at position 45, W at position 47, F and Y at position 91, and A and L at position 93. These residues are indicated in boldface letters in Fig. 2.

The rainbow trout RTVH431 has the same or similar residues in most of these positions, suggesting conformational conservation. However, we came to note that Y at position 24, G at position 28, and N at position 34 are absolutely unique; they occur neither in Kabat's recent compilation of 488 V_H sequences (34) nor in the recently isolated 9 *Heterodontus* V_H (25) or 3 *Xenopus* V_H (21). If G at position 28 was aligned at position 29, it occurred only once in the Kabat compilation. Position 34 is one of the conserved

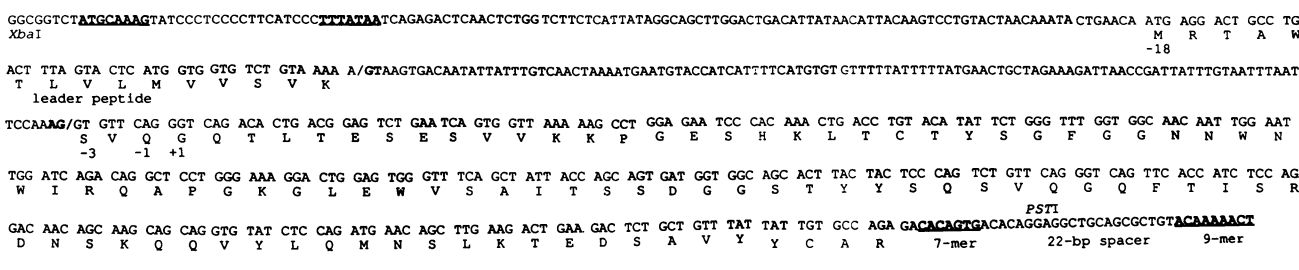

FIG. 1. The clone RTVH431 has an ≈4.5-kb *Eco*RI insert and it generated a single hybridization fragment (619-bp *Xba* I/*Pst* I) with the *Elops* V_H probe (65°C, 5× SSC). The *Xba* I/*Pst* I fragment was subsequently subcloned into the pBluescript SK M13 (+) plasmid and both DNA strands were sequenced. The site 3′ to *Pst* I was sequenced only in one direction. The conserved octamer and the TATA box in the 5′ upstream region and the heptamer–nonamer recombination signal are indicated by underlined boldface letters. The GT–AG boundaries for intron splicing in the leader peptide region is indicated by slashes adjacent to the boldface GT and AG. The single-letter code for amino acids is indicated below the codons.

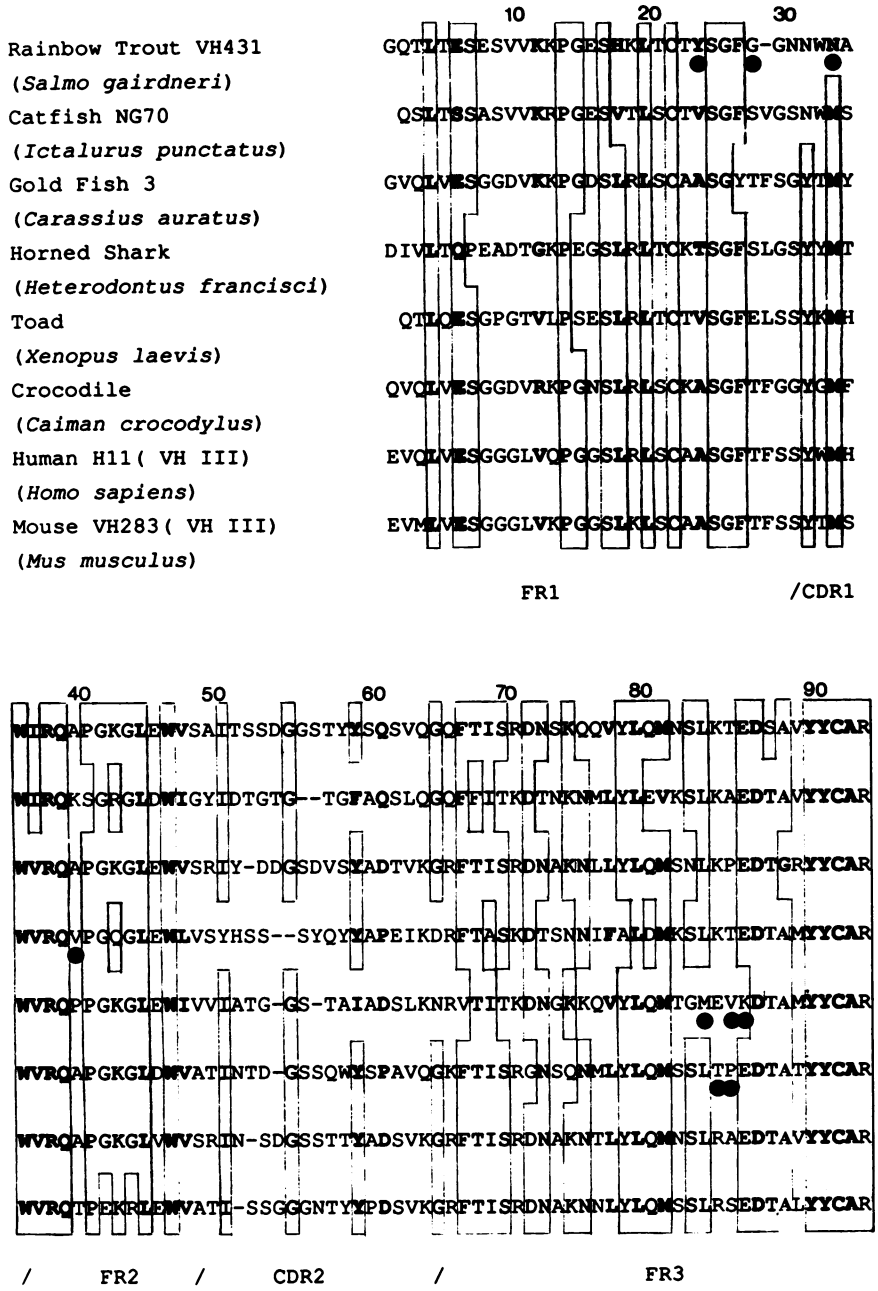

FIG. 2. The amino acid sequence of V_Hs of various vertebrate species is indicated by the single-letter code. The numbering system of Kabat (34) is used. The sequences, which are buried in the β-sheet, as well as residues involved in V_H–V_L contact are shown in boldface letters (see text). Conserved residues that appear more than four times are boxed. Unique residues are marked by a solid circle below the residue. Data are from the following sources: catfish NG70, ref. 25; goldfish 3, ref. 23; horned shark, ref. 24; crocodile, ref. 20; human H11, ref. 33; mouse V_H283, ref. 29.

residues within complementarity-determining regions (37), and it is occupied mostly by M (66.1%) or I (15.5%) or other hydrophobic residues (34). Furthermore, residue 34 is identified as buried in the β-sheet (36), yet N is usually found at the exterior of the protein. It is possible that the combining site of rainbow trout V_H complementarity-determining region 1 has structural idiosyncracy.

Having noticed several unique residues of RTVH431, we further analyzed V_Hs of other species for unique residues, because they are not only very informative in understanding the evolution of V genes (38, 39), but they also provide useful species-specific markers for V_H molecules. We found the following residues to be unique: V at position 40 of framework region in *Heterodontus*, M at position 82 (C), V at position 84, K at position 85 of framework region 3 in *Xenopus*, T at position 83, P at position 84 in framework region 3 in *Caiman* (see Fig. 2). These unique residues are not an unusual minority within a species; V at position 40 occurred in 8 of 9 *Heterodontus* sequences (24); M at position 82 (C), V at position 84, and K at position 85 occurred in 3 of 5 *Xenopus* sequences (21). The unique residues may be specific to the V_H subgroup of a species.

Southern Blot Hybridization. To study the size of the RTVH431 gene family in the rainbow trout genome, we have performed Southern hybridization under rather stringent conditions. As shown in Fig. 3, 10–20 hybridization bands were detected in the genomic DNA isolated from 8 local rainbow trout, and many such bands overlap among DNAs of different individuals. Thus, we would propose that there exist $\approx$20 V_H, which are closely related to the RTVH431 gene. Our

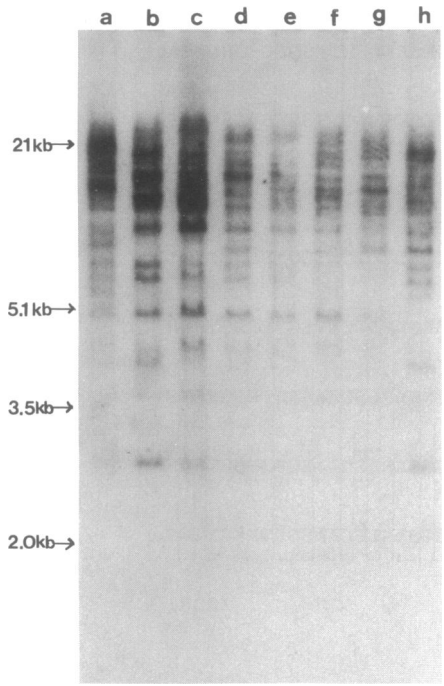

FIG. 3. Southern hybridization of *Eco*RI-digested rainbow trout genomic DNA from eight individual fish hybridized with RTVH431 probe. The RTVH probe was a 619-bp *Xba* I/*Pst* I fragment of genomic clone RTVH431 (see Fig. 1).

result happens to agree with the estimated size of the NG70 V_H family in catfish (25), but it stands in sharp contrast with goldfish, in which such hybridization bands have not been observed in many individuals (23).

Evolution of Vertebrate V_H. In the evolution of antibody V_H genes, the pattern in the evolution of single or a few copy genes is not immediately clear; sequence differences in hemoglobin or cytochrome *c* molecules in various species reflect approximate divergence points (e.g., see ref. 40) in evolution. On the other hand, since adaptive fitness of each individual antibody V gene is considered null or near null (6, 7, 41), the degree of sequence conservation of V_H over 400–500 million years is indeed impressive.

In the evolution of multigene families, genetic processes such as homologous-but-unequal crossing-over (42) or gene conversion (39) have been proposed to account for the occurrence of species-specific residues or sequences. Our analysis showed that species-specific residues in immunoglobulin molecules, originally found in mammalian light chains (43), also occur in V_H genes of primitive vertebrates. One may realize, however, that, while these proposed mechanisms can readily explain the process for homogenizing sequences among multigene family members (e.g., rRNA genes; ref. 44), they also provide an explanation for the conservation of multiple V_H genes as described above.

How then do antibody genes diversify sequences in evolution? It is possible that the homogenization process in antibody genes is slow enough to allow sequence diversity in complementarity-determining regions by classical mutation, drift, and selection (45). In accordance with this notion, our Southern hybridization (unpublished data) shows a remarkable degree of conservation of the RTVH431 family within several species of Salmonidae (e.g., Atlantic salmon, char, sea trout) and northern pike (*Esox lucius*), which is distantly related to the Salmonidae but might have diverged >70 million years ago (46); in these species, there are 10–20 hybridization bands observed under stringent conditions, whereas several other fish species (e.g., goldfish, plaice, burbot, skate) showed no clear hybridization bands under the same conditions.

Note Added in Proof. The second rainbow trout germ-line V_H gene, which has been sequenced recently, also has species-specific residues identical to those of RTVH431.

We thank Drs. G. Litman for the *Elops* V_H probe, D. Holmberg for the S107 V_H probe, T. Ny and P. Sideras for technical advice, and C. Korch for help with the computer sequence search. This work was supported by the Swedish Science Research Council (NFR, B-BU 8445-300).

1. Tonegawa, S. (1983) *Nature (London)* **302,** 575–581.
2. McKean, D., Hüppi, K., Bell, M., Staudt, L., Gerhard, W. & Weigert, M. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 3180–3184.
3. Wable, M., Burrows, P. O., von Gabin, A. & Steinberg, C. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 479–482.
4. Kindt, T. J. & Capra, J. D. (1984) *The Antibody Enigma* (Plenum, New York).
5. Silverstein, A. (1989) *A History of Immunology*, (Academic, New York).
6. Cohn, M. (1968) in *Nucleic Acids in Immunology*, eds. Plescia, O. J. & Braun, W. (Springer, New York), pp. 671–715.
7. Ohno, S. (1976) *Perspect. Biol. Med.* **19,** 527–532.
8. Reynaud, C. A., Anquez, V., Dahan, A. & Weill, J. C. (1985) *Cell* **40,** 238–291.
9. Hinds, K. R. & Litman, G. W. (1986) *Nature (London)* **320,** 546–549.
10. DuPasquier, L. (1982) *Nature (London)* **296,** 311–313.
11. Matsunaga, T. (1985) *Dev. Com. Immunol.* **9,** 585–596.
12. Manning, M. J. & Tatner, M. F., eds. (1985) *Fish Immunology* (Academic, New York).
13. Aggelon, L. B., Davies, S. L., Chen, T. T. & Powers, D. A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 5136–5140.
14. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132,** 6–13.
15. Benton, W. D. & Davies, R. D. (1977) *Science* **196,** 180–182.
16. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5643–5467.
17. Parslow, T. G., Blair, D. L., Murphy, W. J. & Granner, D. K. (1984) *Proc. Acad. Natl. Sci. USA* **81,** 2650–2654.
18. Mason, J. O., Williams, G. T. & Neuberber, M. S. (1985) *Cell* **41,** 479–487.
19. Gerster, Y., Matthias, P., Thali, M., Jirincy, J. & Schaffner, W. (1987) *EMBO J.* **6,** 1323–1330.
20. Litman, G. W., Murphy, K., Berger, L., Littman, R., Hinds, K. & Erikson, B. W. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 844–848.
21. Schwager, J., Grossberger, D. & Du Pasquier, L. (1988) *EMBO J.* **7,** 2409–2415.
22. Amemiya, C. T. & Litman, G. W. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 811–815.
23. Wilson, M. R., Middleton, D. & War, G. W. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1566–1570; *Proc. Natl. Acad. Sci. USA* (1989) **86,** 3276, correction.
24. Kokubu, F., Litman, R., Shamblott, M. J., Hinds, K. & Litman, G. W. (1988) *EMBO J.* **7,** 3143–3422.
25. Chaffari, S. H. & Lobb, C. J. (1989) *J. Immunol.* **143,** 2730–2739.
26. Matsunaga, T. & Törmanen, V. (1990) *Dev. Comp. Immunol.* **14,** 1–8.
27. Lee, K. H., Matsuda, F., Kinashi, T., Kodaira, M. & Honjo, T. *J. Mol. Biol.* **195,** 761–768.
28. Rechavi, G., Ram, D., Glazer, L., Zukat, R. & Givol, D. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 855–859.
29. Ollo, R., Sikov, V. & Rougeon, F. (1983) *Nucleic Acids Res.* **11,** 7887–7879.
30. Permitter, R. M., Berson, B. M., Griffin, J. A. & Hood, L. (1985) *J. Exp. Med.* **162,** 1998–2016.
31. Currier, S. J., Gallarda, J. L. & Knight, K. L. (1988) *J. Immunol.* **140,** 1651–1659.
32. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. (1980) Cell **19,** 981–992.
33. Rechavi, G., Bienz, B., Ram, D., Ben-Neriah, Y., Cohen, J. B., Zykut, R. & Givol, D. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4409–4411.

34. Kabat, E. A., Wu., T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1987) *Sequences and Proteins of Immunological Interest* (National Institutes of Health, Bethesda, MD).
35. Saul, F. A., Amzel, L. M. & Poljack, R. J. (1978) *J. Biol. Chem.* **253,** 585–597.
36. Chothia, C., Boswell, D. R. & Lesk, A. M. (1988) *EMBO J.* **7,** 3745–3755.
37. Ohno, S., Mori, N. & Matsunaga, T. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2945–2949.
38. Hood, L., Campbell, J. H. & Elgin, S. C. R. (1975) *Annu. Rev. Genet.* **9,** 305–353.
39. Dover, G. A. & Strachan, T. (1987) in *Evolution and Vertebrate Immunity*, eds. Kelsoe, G. & Schultze, D. H. (University of Texas Press, Austin, TX), pp. 15–33.
40. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, (Cambridge Univ. Press, New York).
41. Kelsoe, G., Miceli, R., Cerny, J. & Schultze, D. H. (1989) *Immunogenetics* **29,** 288–296.
42. Smith, G. P. (1973) *Cold Spring Harbor Symp. Quant. Biol.* **38,** 507–513.
43. Hood, L., Eichman, K., Lackland, H., Krause, R. M. & Ohms, J. J. (1979) *Nature (London)* **228,** 288–296.
44. Arnheim, N. (1983) in *Evolution of Genes and Proteins*, eds. Nei, M. & Koehn, R. K. (Sinauer, Sunderland, MA), pp. 39–61.
45. Gojobori, T. & Nei, M. (1984) *Mol. Biol. Evol.* **1,** 195–212.
46. Carrol, R. L. (1987) *Vertebrate Paleontology and Evolution* (Freeman, New York).