Amino acid sequence of heavy chain from *Xenopus laevis* IgM deduced from cDNA sequence: Implications for evolution of immunoglobulin domains

(amphibian/antibody/variable region/constant region/nucleotide sequence)

JOSEPH SCHWAGER*, CAROLE A. MIKORYAK†, AND LISA A. STEINER‡

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Communicated by Alfred Nisonoff, November 16, 1987 (received for review August 29, 1987)

ABSTRACT Present understanding of the evolution of immunoglobulins is derived almost entirely from studies of a few mammalian species. To obtain information about immunoglobulin genes in *Xenopus laevis*, a cDNA library was prepared in the expression vector λ gt11 from mitogen-stimulated splenocytes of this species. Of $\approx 50,000$ clones screened, 18 were found to express IgM epitopes. One of these, λ XIg14, hybridized with RNA of ≈ 2 kilobases from splenocytes. The insert of this clone appears to encode a variable region and part of a μ constant region; that of another clone, λ XIg8, appears to encode a variable region and a complete μ constant region. Both inserts contain sequence corresponding to the three gene segments (V_H , D_H , and J_H) that encode heavy-chain variable regions. The heavy-chain constant region (C_H) encoded by λ XIg8 has the characteristic features of C_μ , including a four-domain structure and a carboxyl-terminal tail. The amino acid sequences of two μ -chain peptides agree with the cDNA sequence. The identity in amino acid sequence between the corresponding *Xenopus* and mouse C_μ domains ranges from 31 to 47%. The C_μ domains vary in the extent to which their sequences resemble the sequences of other immunoglobulins, consistent with previous suggestions that the immunoglobulin domains have an independent evolutionary history.

Present understanding of the evolution of immunoglobulins (Ig) rests almost entirely on information derived from studies of a few mammalian species. Relatively little is known about antibodies or the genes encoding them in other vertebrates. Molecules with the general features of IgM are found in all major vertebrate groups; additional Ig classes, with uncertain relation to mammalian Ig, have been identified rarely in fish (1) and more frequently in amphibians, reptiles, and birds (2). Antibody functions in lower vertebrates, including the ability to bind a variety of antigens and to fix complement, seem similar to those of antibodies in mammals, although the diversity of the antibody repertoire appears to be relatively restricted (3).

Recent studies of Ig genes in nonmammalian vertebrates have revealed both similarities and differences from mammalian Ig genes. In sharks, recombination of individual V_H may be restricted to particular D_H and J_H gene segments (4); in chickens, gene conversion appears to account for the diversification of light-chain variable regions (5).

Information about the sequence of constant regions of Ig in nonmammalian vertebrates is limited to that of a chicken (6) and a frog (*Rana catesbeiana*) (7) light chain and most of a chicken μ chain (8). We report here the sequence of a cDNA clone encoding the variable and constant regions of a μ chain from another amphibian, *Xenopus laevis*, as well as the sequence corresponding to the variable region of a second

clone.[§] The results are interpreted in terms of the evolution of Ig domains.

MATERIALS AND METHODS

IgM and Peptides. A high-molecular weight protein pool, obtained by gel filtration of *Xenopus* plasma on Bio-Gel A-1.5m, was mildly reduced (9), alkylated with iodo[1-¹⁴C]acetic acid, and rechromatographed on the same column; the second peak contained only μ and light chains. After chain separation (9), the μ chains were completely reduced and alkylated, rechromatographed, citraconylated (to block lysines), and digested with trypsin (9). Methods for isolation of peptides and determination of amino acid composition and sequence have been described (7, 9).

Antisera. Rabbits were injected twice with periodate-treated (10) *Xenopus* IgM in complete Freund's adjuvant. Antiserum to *Xenopus* IgY (a second class of *Xenopus* Ig) and affinity-purified antibodies to light chains were donated by I. Hadji-Azimi.

Preparation of RNA and of cDNA Library. *X. laevis* were obtained from Xenopus I (Ann Arbor, MI). Total RNA was prepared from ovaries as described (11) and from liver and splenocytes by extraction with guanidinium isothiocyanate, followed by CsCl centrifugation (12). Spleen cells from individual frogs were cultured for 6–7 days in Iscove's modified Dulbecco's medium, diluted to 70% (amphibian osmolarity), in the presence of *Escherichia coli* lipopolysaccharides (150 μ g/ml), pokeweed mitogen, or anti-IgM (13). Poly(A)⁺ RNA (12) from splenocytes of 11 frogs was used to prepare cDNA as described (14); after addition of linkers [d(GGAATTCC)], the cDNA was inserted into λ gt11 (15).

Identification of Recombinants Expressing IgM Epitopes. The unamplified cDNA library was screened (16) with the IgG fraction of anti-IgM (50 μ g/ml). Lysogens in *E. coli* Y1089 were prepared (17), induced with isopropyl β -thiogalactopyranoside at 43°C (14), grown for 50 min at 38°C, and lysed in gel sample buffer. Lysates were subjected to NaDodSO₄/PAGE and processed for immunoblotting (18).

Abbreviations: A heavy-chain variable region is encoded by three gene segments, V_H (variable), D_H (diversity), and J_H (joining); the corresponding segments of the protein are denoted V_H , D_H , and J_H . A heavy-chain constant region, C_H , contains several domains, C_{H1} , C_{H2} , etc., each encoded by an exon C_{H1} , C_{H2} , etc. The constant region of the μ heavy chain is also called C_μ and its domains $C_{\mu1}$, $C_{\mu2}$, etc.; PIR, Protein Identification Resource.

*Present address: Basel Institute for Immunology, Grenzacherstrasse 487 CH-4058 Basel, Switzerland.

†Present address: University of Texas at Dallas, Richardson, TX 75080.

‡To whom reprint requests should be addressed.

[§]The sequences reported in this paper are being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession nos. J03631 and J03632).

Nucleotide Sequence. The cDNA inserts were subcloned into M13. Sequential deletion clones were generated as described (19). The sequences of both strands of the entire insert of clone λ XIg8 and of the segment encoding the variable region of λ XIg14 were determined (20).

Computer Programs. Programs (FASTP for data base comparisons and ALIGN for pairwise alignments) and data bases were from the Protein Identification Resource (PIR).[‡] For protein alignments, the 250 PAM mutation data matrix (bias, 6; gap penalty, 6) was used (21). The matrix assigns a value to each aligned pair of residues according to the frequency of the corresponding replacement in related proteins (21). Optimal alignment is that alignment for which the sum of these values is greatest. This maximum score is defined as the similarity score [called "total score" in ALIGN program (21)]. For nucleic acid alignments, a unitary matrix (bias, 2; gap penalty, 2) was used. Identities (determined from the optimum alignment) and similarity scores were normalized by dividing by the number of residues in the shorter sequence and multiplying by 100.

RESULTS

Isolation and Characterization of cDNA Clones Expressing IgM Epitopes. Of 50,000 recombinant clones screened, 18 clones reacted with antibodies to *Xenopus* IgM; none of the 18 clones reacted with antibodies to IgY or to light chains. Five of these clones expressed proteins that were larger than β -galactosidase and that contained IgM epitopes (Fig. 1a). Two clones containing the largest inserts, λ XIg8 ($\approx$ 2000 nucleotides, data not shown) and λ XIg14 ($\approx$ 1600 nucleotides), were analyzed as described below.

Blot Hybridization of RNA from Different Tissues. RNA was isolated from several tissues and, after gel electrophoresis, was hybridized with the cDNA insert of λ XIg14. An RNA species in mitogen-stimulated spleen cells, approximately the same size as the insert from λ XIg8, hybridized with the probe; there was no reaction with RNA from ovary and a barely detectable reaction with RNA from liver (Fig. 1b).

Nucleotide Sequence of cDNA Inserts λ XIg8 and λ XIg14. The sequence of λ XIg8 consists of 1920 nucleotides plus a poly(A) tract (Fig. 2). An open reading frame, extending from position 2 to position 1768, is followed by two termination codons and 146 untranslated nucleotides. However, the cDNA is not full-length, since the initiation codon is absent. All the segments expected in a complete μ chain are present: a signal sequence, a variable region, and a constant region consisting of four domains and a carboxyl-terminal tail. The sequence of the segment of λ XIg14 that encodes the variable region is also shown (Fig. 2a).

Amino Acid Sequence of μ Chain Peptides. The sequences of two peptides agreed with the deduced amino acid sequence of λ XIg8. One of these peptides lacked arginine and was therefore independently localized to the carboxyl terminus. The second peptide was in $C_{\mu}3$.

DISCUSSION

Variable-Region Sequences. The heavy-chain variable region of mouse myeloma protein MOPC 141 (32) was found, in the FASTP analysis, to be the best match for the variable regions encoded by λ XIg8 and λ XIg14; the three sequences are compared in Fig. 2a.

In V_H (not including the signal sequence), only 18/288 nucleotides differ between λ XIg8 and λ XIg14, resulting in 11

FIG. 1. Characterization of cDNA clones by immunoblotting and RNA hybridization. (a) Immunoblots of lysates from lysogens carrying the recombinant clones (λ XIg18, λ XIg14, etc.) are indicated. Fusion proteins were detected by reaction with rabbit anti-*Xenopus* IgM, followed by peroxidase-labeled goat anti-rabbit IgG (Boehringer Mannheim) and peroxidase substrate. Clone λ XIg8 is not included because it reacted very weakly. Molecular size markers are expressed in kDa; β gal indicates the position of β -galactosidase. (b) Blot hybridization of total cellular RNA, denatured with glyoxal, from mitogen-stimulated spleen cells (S), liver (L), and ovary (O). RNA was electrophoresed through a 1.2% agarose gel, blotted to a nylon filter (Zetabind AMF Cuno), and hybridized (22) with the 32 P-labeled (23) cDNA insert of λ XIg14. mRNA amount applied to each lane, as estimated by reprobing with *Xenopus* fibronectin cDNA, was similar for spleen and liver and several-fold less for ovary. λ 8 is the denatured cDNA insert of λ XIg8. Markers are 18S and 28S mouse ribosomal RNA.

amino acid replacements. The similarity between these V_H suggests that they belong to the same gene family (26). However, the existence of at least one other V_H family is suggested by comparison with the coding sequence of a genomic *Xenopus* V_H (27), which is only 59% identical to the V_H of λ XIg8. Amino-terminal sequences of pooled heavy chains from *Xenopus* serum Ig (28, 29) are very similar to the sequence encoded by the genomic V_H .

Most of the conserved features of mammalian V_H (24) are found in the predicted *Xenopus* amino acid sequences. In addition to the two invariant cysteines, highly conserved tryptophans are present at positions 36 and 47. A number of other residues characteristic of V_H are also found in the *Xenopus* sequence (positions 4, 7, 8, 25, 26, 37, 38, 45, 46, 66, 80, 86–88, 90, and 91).

Table 1 summarizes results of pairwise comparisons of amino acid and nucleotide sequences of seven different V_H : three from *X. laevis*, one each from horned shark and caiman (a reptile), and two from mice. The degree of divergence among these sequences is not correlated with the species.

Much of the third complementarity-determining region of mammalian heavy-chain variable regions is encoded by a D_H gene segment. In this region, the two *Xenopus* proteins are very different from one another and from the mouse protein (Fig. 2a). However, an alternative alignment of the two *Xenopus* DNA sequences reveals a stretch of 16 identical nucleotides, consistent with use of the same D_H gene segment; the two sequences are translated in different reading frames, perhaps due to the insertion of "N nucleotides" (38) at the V_H - D_H and D_H - J_H junctions.

The remainder (carboxyl-terminal section) of the heavy-chain variable region is encoded by a J_H gene segment, which is relatively short in MOPC 141 (32). The *Xenopus* J_H may begin several nucleotides 5' to the position indicated in Fig. 2a. Amino acid residues at positions 103, 104, 106, 107, 110, and 111 of the *Xenopus* sequences are identical to residues encoded at these positions by the four germ-line J_H of mice (24), the six germ-line J_H of humans (24), and a germ-line J_H of sharks (4).

[‡]Protein Identification Resource (Natl. Biomed. Res. Found., Washington, DC), Protein Sequence Database, Release 12.0 (1987) and Nucleic Acid Sequence Database, Release 30.0 (1987).

Table 1. Comparison of V_H domains

	Similarity score: protein						
	λXlg8	λXlg14	XEN VH1	SHARK	CAIMAN	MOPC 141	93G7
XENOPUS λXlg8	1103 100(100)	1050	880	856	881	943	836
XENOPUS λXlg14	89(94)	1098 100(100)	911	873	899	949	871
XENOPUS VH1	44(61)	52(65)	1104 100(100)	915	944	897	889
SHARK HXIA	41(57)	45(58)	53(62)	1116 100(100)	907	860	854
CAIMAN C3	42(55)	47(57)	60(65)	48(62)	1114 100(100)	910	889
MOUSE MOPC 14	60(62)	61(62)	47(60)	44(52)	48(62)	1095 100(100)	886
MOUSE 93G7	32(53)	38(55)	44(56)	37(53)	45(57)	45(57)	1097 100(100)
Per cent identity: protein (DNA)							

Pairs of V_H domains (excluding signal sequences) were aligned; similarity scores (upper right triangle) and percent identities (lower left triangle) were determined. Nucleic acid sequences encoding these domains were also aligned, and these percent identities are shown in parentheses. The *Xenopus* λXlg8 and λXlg14 sequences are from Fig. 2a; *Xenopus* VH1 is from ref. 27. Other sequences are from the PIR protein (nucleic acid) data bases; PIR codes and positions are as follows: shark HXIA, HVRK 20–116 (HVRK 186–476); caiman C3, HVCQ 20–117 (HVCQC3 299–592); mouse MOPC 141, G2MS14 20–116 (HVMS14 269–559); mouse hybridoma 93G7, HVMSG7 20–117 (HVMSG7 75–368). See Table 2 for scores obtained with randomized protein sequences. For alignments of the randomized nucleic acid sequences, mean percent identity was 38.

Constant-Region Sequences. An alignment of the *Xenopus* and mouse C_μ is shown in Fig. 2b. The domain boundaries indicated are based on the exons of the mouse C_μ gene (25). Comparison with the splice sites of mouse (25) and human (30) C_μ genes and inspection of preferred nucleotides at the exon ends of known splice junctions (31) suggests that the *Xenopus* C_H1–C_H2 boundary might be shifted in the 3' direction by one codon.

Each of the predicted *Xenopus* C_H domains has the characteristic features of an Ig domain, including a pair of cysteines that could form a disulfide bridge spanning about 60 residues and a conserved tryptophan lying between these cysteines (24). For the different domains, identities between the *Xenopus* and mouse amino acid sequences range from 31% (C_μ2) to 47% (C_μ3) (Table 2). For the total constant region, the two protein sequences are 39% identical (degree of identity for the nucleic acid sequences is 51%). The carboxyl-terminal tail of the *Xenopus* μ chain is four residues shorter than that of the mouse μ chain; the tyrosine residue that is invariably found at the carboxyl terminus of mammalian and chicken μ chains is absent from the *Xenopus* sequence (Fig. 2b). Comparison with the genes for secreted and membrane forms of mouse μ chain (39) suggests that the second two nucleotides in the codon for Gly-557 (GT) could be the beginning of an intron that separates *Xenopus* C_μ4 from the first membrane exon.

The four cysteine residues that usually form interchain disulfide bridges in mouse and/or human IgM (positions 140, 337, 414, and 575) are conserved in the *Xenopus* sequence. The carboxyl-terminal cysteine (position 575) may be the site for attachment of J chain, which is present in *Xenopus* IgM (33). The *Xenopus* μ chain has eight acceptor sites for N-linked oligosaccharides, five of which are in the same location as acceptor sites in mouse μ.

In C_μ4, the *Xenopus* sequence has an insertion of four residues between positions 505 and 506, relative to the mouse sequence. In addition, this domain contains replace-

Table 2. Comparison of domains of *Xenopus laevis* and mouse C_μ with domains of other Ig and Ig-like proteins

	C _μ 1		C _μ 2		C _μ 3		C _μ 4	
	XEN	MO	XEN	MO	XEN	MO	XEN	MO
MO μ	798 32(4)	1106 100(0)	733 31(8)	1083 100(0)	825 47(3)	1077 100(0)	883 42(1)	1103 100(0)
RAB μ	803 35(5)	955 56(1)	770 33(4)	935 59(1)	794 34(5)	927 61(0)	885 44(1)	1032 85(0)
CHI μ	—	—	718 30(7)	711 24(7)	819 44(3)	816 41(1)	876 38(1)	881 45(0)
MO γ2a	804 ¹ 38(6)	754 ¹ 23(4)	742 ³ 30(3)	741 ³ 30(4)	709 ² 25(4)	758 ² 28(3)	854 ³ 40(4)	817 ³ 33(3)
MO α	738 ¹ 28(5)	747 ¹ 30(6)	750 ³ 29(5)	726 ³ 27(6)	721 ² 28(5)	745 ² 28(4)	855 ³ 40(2)	873 ³ 43(1)
HU δ	728 ¹ 25(5)	729 ¹ 29(7)	729 ¹ 22(6)	699 ³ 31(5)	728 ¹ 23(5)	706 ² 25(4)	744 ¹ 20(4)	753 ¹ 24(3)
HU ε	741 31(5)	711 23(5)	715 19(4)	774 33(4)	715 26(2)	777 31(3)	805 35(3)	794 35(5)
MO κ	709 19(1)	650 14(5)	748 30(3)	737 30(4)	677 22(7)	689 18(5)	737 24(5)	732 26(3)
MO λ1	730 25(2)	692 23(5)	798 32(3)	771 30(3)	690 25(7)	689 19(7)	765 30(5)	771 31(5)
H-2 Aβ ^d	673 15(2)	701 24(6)	772 28(3)	755 29(5)	712 21(8)	668 18(7)	738 27(7)	748 35(5)
TCR-β _{C2}	734 30(8)	708 21(4)	731 27(5)	717 30(5)	678 25(8)	677 23(5)	699 24(9)	724 25(4)
MO μ random (16 x)	598 15(6)	588 16(6)	593 16(5)	588 15(5)	603 15(6)	587 15(6)	596 17(6)	583 15(5)

Each constant domain of the *Xenopus* (XEN) or mouse (MO) μ chain was compared (ALIGN program) with a domain from a chain shown in the left column. For chains with four constant domains (μ and ε), alignment was with the corresponding domain (e.g., *Xenopus* C_μ2 and chicken C_μ2); for chains with three constant domains (γ, α, and δ), alignment was with the best-matching domain, as indicated by the superscript (e.g., *Xenopus* C_μ2 and C_H3 of mouse α); for chains having a single constant domain (κ, λ1, H-2 Aβ^d, and T-cell receptor β chain), alignment was with that domain. In each group, the upper number is the similarity score, the lower number on left is the percent identity, and the number in parentheses is the number of gaps in alignment. Domains correspond to exons where known or were assigned based on alignment with mouse C_μ. The XEN sequences are from Fig. 2b. For other constant regions, the PIR protein data base codes (in brackets below) and residue positions are as follows: MO (mouse) μ [MHMS] C_H1:1–105, C_H2:106–218, C_H3:219–324, C_H4:325–437; RAB (rabbit) μ [MHRB] C_H1:1–106, C_H2:107–221, C_H3:222–327, C_H4:328–440; CHI (chicken) μ [MHCH] C_H2:23–128, C_H3:129–236, C_H4:237–349; MO γ2a [G2MSA] C_H1:1–97, C_H2:114–223, C_H3:224–330; MO α [AHMS] C_H1:1–101 [based on sequences in Kabat *et al.* (24), Ser-18 was replaced by cysteine and Asn-67 by serine], C_H2:114–213 (residues 102–113 encoded at the beginning of the C_H2 exon, corresponding to all or part of a hinge region, were omitted), C_H3:214–326; HU (human) δ [DHHU] C_H1:1–102, C_H2:160–267, C_H3:268–375; human ε [EHHU] C_H1:1–103, C_H2:104–210, C_H3:211–318, C_H4:319–428; MO κ [KIMS]; MO λ [LIMS]; H-2 class II Aβ^d chain (mouse) [HLSAB, 123–216]; T-cell receptor β2 constant region (mouse) [RWMSBC, 1–125]. MO μ random (16 x): each sequence in the indicated pair (e.g., XEN C_μ1 and MO C_μ1, MO C_μ1 and MO C_μ1, etc.) was randomized before alignment; each entry represents the mean of values from 16 such alignments.

ments at positions 502, 505, 521, and 523, which are invariant in all other known four-domain constant regions (μ and ε) (24). However, it appears that these variations can be accommodated within the overall structure of the Ig fold (E. A. Padlan, personal communication).

The individual *Xenopus* C_μ domains were compared to the PIR protein data base by the FASTP program; unexpectedly, a significantly different set of "best matches" was obtained for each domain. To quantitate these comparisons, pairwise alignments of each *Xenopus* domain with domains of other Ig and Ig-like proteins were done. The results (see

Table 2) confirm that the degree of similarity between the *Xenopus* μ chain and the other sequences varies markedly from domain to domain—for example, (i) the best matches for *Xenopus* $C_{\mu}3$ are $C_{\mu}3$ domains of the other species [similarity scores 825 (mouse), 794 (rabbit), and 819 (chicken); other scores for alignments of *Xenopus* $C_{\mu}3$ are 728 or less]; (ii) *Xenopus* $C_{\mu}1$ matches C_{H1} of mouse $\gamma 2a$ (score 804) about as well as it matches the other $C_{\mu}1$ domains [scores 798 (mouse) and 803 (rabbit)]; (iii) *Xenopus* $C_{\mu}4$ matches the C_{H3} domains of both the mouse α and $\gamma 2a$ chains (scores 855 and 854) almost as well as it matches other $C_{\mu}4$ domains (scores 883, 885, and 876); (iv) *Xenopus* $C_{\mu}2$ actually matches some of the non-heavy-chain domains about as well as—or even better than—it matches the $C_{\mu}2$ domains of the other species (scores with constant domains of λ chain, H-2 class II, and T-cell receptor are 798, 772, and 731, respectively; scores with other $C_{\mu}2$ are 733, 770, and 718). In fact, the scores for *Xenopus* $C_{\mu}2$ are reasonably uniform for the entire set of comparisons; however, the best matches involving the $C_{\mu}2$ domain are not as good as the best matches involving the other C_{μ} domains.

The finding that the *Xenopus* C_{μ} domains, with the exception of $C_{\mu}3$, are nearly as similar to domains of mouse $\gamma 2a$ as to other C_{μ} domains can be interpreted to mean that the divergence of the μ and γ chains did not occur long before the separation of the lineages that led to *X. laevis* and to mammals—i.e., so that at the time of species divergence, the distinction between μ and γ had not yet become firmly established. Only scores for the $C_{\mu}3$ domain suggest that the *Xenopus* chain should be regarded as a member of the μ -chain family. Yet, many of its features—the four-domain structure of the constant region, the presence of a carboxyl-terminal tail and, especially, the correspondence of the cDNA sequence to the sequence of two peptides isolated from the heavy chain of an IgM-like serum protein—establish that it is indeed a μ chain. The results do illustrate, however, the potential difficulty of determining, from the sequence alone, the phylogenetic relation between a chain derived from an Ig of a lower vertebrate and a mammalian Ig chain.

The entire set of alignments was repeated using the constant domains of mouse μ chain instead of those of the *Xenopus* chain (Table 2). Although, for mouse C_{μ} domains the greatest similarity is to other mammalian C_{μ} , some features noted with *Xenopus* domain comparisons are again evident. Thus, C_{H3} of the α chain is a reasonably good match for mouse $C_{\mu}4$ (similarity score 873) and mouse $C_{\mu}2$, like *Xenopus* $C_{\mu}2$, is broadly similar to the other domains with which it was compared. Overall, domains of the other mammalian heavy chains are no more similar to the mouse than to the *Xenopus* C_{μ} domains (e.g., the score of C_{H1} of mouse $\gamma 2a$ vs. mouse $C_{\mu}1$ is 754 and vs. *Xenopus* $C_{\mu}1$ is 804). A possible interpretation of this finding is that the gene duplications responsible for the different classes of mammalian heavy chains did not occur significantly after divergence of the lineages leading to *X. laevis* and to mammals. However, variation from domain to domain (e.g., for C_{H2} of $\gamma 2a$ vs. $C_{\mu}3$ of mouse and *Xenopus*, the comparable numbers are in the opposite direction: 758 and 709) makes interpretation difficult and complicates attempts to account for the evolution of the multidomain Ig chains.

These data provide additional support for the suggestion of Putnam and coworkers (34), based on their analysis of sequences of human heavy chains, that “individual domains of heavy chains have evolved independently rather than through a line of descent in which one whole chain is the ancestor of another.” Genetic exchange among exons (35) could contribute to this process. The evolutionary origin of Ig domains has been explored by the construction of phylogenetic trees (see ref. 36). Incorporating into such trees the

sequences of lower-vertebrate Ig now becoming available should prove instructive.

Note. Kokubu *et al.* (37) have recently reported the nucleotide sequences of the $C_{\mu}1$ exons of five shark heavy-chain germ-line genes. We have aligned the translated sequences with the domains listed in Table 2. For shark sequence 801, normalized similarity scores are as follows: mouse $C_{\mu}1$, 767; rabbit $C_{\mu}1$, 783; mouse $\gamma 2a$ C_{H1} , 771; and other domains, 731 or less.

We thank I. Hadji-Azimi for gifts of antisera, D. A. Melton for the fibronectin cDNA probe, E. Hsu, E. S. Lander, and E. A. Padlan for discussions, and W. A. Gilbert for help with computer programs. This work was supported by Grant AI-08054 from the National Institutes of Health and fellowships to J.S. from the Swiss National Science Foundation and the Fogarty International Center.

- Kobayashi, K., Tomonaga, S. & Kajii, T. (1984) *Mol. Immunol.* **21**, 397–404.
- Marchalonis, J. J. (1977) *Immunity in Evolution* (Harvard Univ. Press, Cambridge, MA).
- Du Pasquier, L. (1982) *Nature (London)* **296**, 311–313.
- Hinds, K. R. & Litman, G. W. (1986) *Nature (London)* **320**, 546–549.
- Reynaud, C.-A., Anquez, V., Grimal, H. & Weill, J.-C. (1987) *Cell* **48**, 379–388.
- Reynaud, C.-A., Dahan, A. & Weill, J.-C. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4099–4103.
- Mikoryak, C. A. & Steiner, L. A. (1988) *Mol. Immunol.*, in press.
- Dahan, A., Reynaud, C.-A. & Weill, J.-C. (1983) *Nucleic Acids Res.* **11**, 5381–5389.
- Elliott, B. W., Jr., Eisen, H. N. & Steiner, L. A. (1984) *J. Immunol.* **133**, 2757–2761.
- Mattes, M. J. & Steiner, L. A. (1978) *Nature (London)* **273**, 761–763.
- Long, E. O. & Dawid, I. B. (1979) *Cell* **18**, 1185–1196.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
- Schwager, J. & Hadji-Azimi, I. (1985) *Differentiation* **30**, 29–34.
- Schwarzbauer, J. E., Tamkun, J. W., Lemischka, I. R. & Hynes, R. O. (1983) *Cell* **35**, 421–431.
- Huynh, T. V., Young, R. A. & Davis, R. W. (1985) in *DNA Cloning: A Practical Approach*, ed. Glover, D. M. (IRL, Oxford), Vol. 1, pp. 49–78.
- Young, R. A., Bloom, B. R., Grosskinsky, C. M., Ivanyi, J., Thomas, D. & Davis, R. W. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2583–2587.
- Berman, M. L., Enquist, L. W. & Silhavy, T. J. (1982) *Advanced Bacterial Genetics* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
- Burnette, W. N. (1981) *Anal. Biochem.* **112**, 195–203.
- Dale, R. M. K., McClure, B. A. & Houchins, J. P. (1985) *Plasmid* **13**, 31–40.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545.
- Wahl, G. M., Stern, M. & Stark, G. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3683–3687.
- Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
- Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. (1987) *Sequences of Proteins of Immunological Interest* (HHS, National Institutes of Health, Bethesda, MD), 4th Ed.
- Kawakami, T., Takahashi, N. & Honjo, T. (1980) *Nucleic Acids Res.* **8**, 3933–3945.
- Brodeur, P. H. & Riblet, R. (1984) *Eur. J. Immunol.* **14**, 922–930.
- Yamawaki-Kataoka, Y. & Honjo, T. (1987) *Nucleic Acids Res.* **15**, 5888.
- Wang, A. C., Tung, E., Fudenberg, H. H. & Hadji-Azimi, I. (1978) *J. Immunogenet.* **5**, 355–364.
- Brandt, D. C., Griessen, M., Du Pasquier, L. & Jaton, J.-C. (1980) *Eur. J. Immunol.* **10**, 731–736.
- Rabbitts, T. H., Forster, A. & Milstein, C. P. (1981) *Nucleic Acids Res.* **9**, 4509–4524.
- Mount, S. M. (1982) *Nucleic Acids Res.* **10**, 459–472.
- Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980) *Nature (London)* **286**, 676–683.
- Hadji-Azimi, I. & Michea-Hamzehpour, M. (1976) *Immunology* **30**, 587–591.
- Liu, Y.-S. V., Low, T. L. K., Infante, A. & Putnam, F. W. (1976) *Science* **193**, 1017–1020.
- Gilbert, W. (1978) *Nature (London)* **271**, 501.
- Barker, W. C., Ketcham, L. K. & Dayhoff, M. O. (1980) *J. Mol. Evol.* **15**, 113–127.
- Kokubu, F., Hinds, K., Litman, R., Shambloot, M. J. & Litman, G. W. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5868–5872.
- Alt, F. W. & Baltimore, D. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4118–4122.
- Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. & Hood, L. (1980) *Cell* **20**, 313–319.