Phylogenetic Diversification of Immunoglobulin Genes and the Antibody Repertoire[1]

Gary W. Litman, Jonathan P. Rast,* Michael J. Shamblott,* Robert N. Haire,* Michele Hulst,* William Roess,† Ronda T. Litman,* Kristin R. Hinds-Frey,* Anna Zilch,* and Chris T. Amemiya‡*

*Department of Pediatrics, University of South Florida, †Eckerd College, and ‡Lawrence Livermore National Laboratory, Biomedical Sciences Division

Immunoglobulins are encoded by a large multigene system that undergoes somatic rearrangement and additional genetic change during the development of immunoglobulin-producing cells. Inducible antibody and antibody-like responses are found in all vertebrates. However, immunoglobulin possessing disulfide-bonded heavy and light chains and domain-type organization has been described only in representatives of the jawed vertebrates. High degrees of nucleotide and predicted amino acid sequence identity are evident when the segmental elements that constitute the immunoglobulin gene loci in phylogenetically divergent vertebrates are compared. However, the organization of gene loci and the manner in which the independent elements recombine (and diversify) vary markedly among different taxa. One striking pattern of gene organization is the "cluster type" that appears to be restricted to the chondrichthyes (cartilaginous fishes) and limits segmental rearrangement to closely linked elements. This type of gene organization is associated with both heavy- and light-chain gene loci. In some cases, the clusters are "joined" or "partially joined" in the germ line, in effect predetermining or partially predetermining, respectively, the encoded specificities (the assumption being that these are expressed) of the individual loci. By relating the sequences of transcribed gene products to their respective germ-line genes, it is evident that, in some cases, joined-type genes are expressed. This raises a question about the existence and/or nature of allelic exclusion in these species. The extensive variation in gene organization found throughout the vertebrate species may relate directly to the role of intersegmental (V↔D↔J) distances in the commitment of the individual antibody-producing cell to a particular genetic specificity. Thus, the evolution of this locus, perhaps more so than that of others, may reflect the interrelationships between genetic organization and function.

Introduction

In all jawed vertebrate species, antibody diversity is specified by a heterodimeric immunoglobulin molecule consisting of two heavy chains and two light chains, which form a functionally bivalent monomer. Regional sequence variation in both heavy and light chains led to the classification of variable (V) and constant (C) regions, which later were shown to be encoded by separate exons. Immunoglobulin classes are

1. Key words: immunoglobulin, gene, evolution, segmental organization, gene rearrangement, diversity.

Address for correspondence and reprints: Gary W. Litman, All Children's Hospital, 801 Sixth Street South, Department 968, St. Petersburg, Florida 33701.

Mol. Biol. Evol. 10(1):60–72. 1993.

0737-4038/93/1001-0005$02.00

distinguished by differences in protein primary structure in constant regions and by different polymeric configurations of the basic monomer. Immunoglobulins are products of lymphocytes differentiating along the B cell developmental lineage. The genetic basis for the extensive variation that antibodies exhibit has an inherited (germ line) as well as a somatic basis. The immunoglobulin heavy-chain gene is encoded at loci consisting of four different segmental elements: V_H (variable), D_H (diversity), J_H (joining), and C_H (constant). Three different segmental elements encode the light chain: V_L, J_L, and C_L. The number of segmental elements constituting different heavy- and light-chain gene families vary in different species. This discussion largely will be based on the heavy-chain locus, which, in mammals, generally consists of several hundred different V_H elements, 20–30 D_H elements, and fewer <10 J_H elements. V_H-D_H-J_H joining occurs in B cells and involves rearrangement of these elements, in a true combinatorial sense. However, many of these (as high as ~50% in humans) are pseudogenes, and not all rearrangements produce functional molecules, i.e., only one in three will retain the correct reading frame. An example of a mammalian-type germ-line gene and a somatically rearranged form are shown in figure 1 (*top*). DNA between

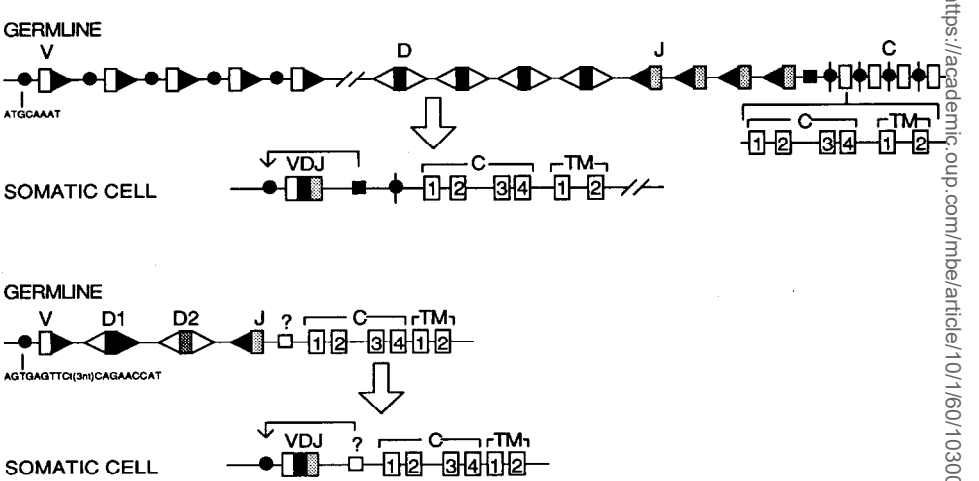

FIG. 1.—Comparison of germ-line organization of hypothetical mammalian-type (*top*) vs. cartilaginous fish (i.e., "cluster type") (*bottom*) immunoglobulin heavy-chain gene loci and their specific rearrangement product that would occur in a somatic cell (B lymphocyte). Recombination within the mammalian-type locus involves the rearrangement of individual variable (V), diversity (D), and joining (J) segments in a true combinatorial sense, whereas rearrangement of the cluster type apparently is not associated with combinatorial joining. Blackened triangles (▶ and ◀) denote 22 or 23 nucleotide spacers in recombination signal sequences; and unblackened triangles (▷ and ◁) denote 12 nucleotide spacers in recombination signal sequences. Constant regions (C) are associated with switch-region sequences (◆), repetitive sequences that facilitate class switching. An expanded view of one constant region consisting of six exons is shown; two of these are transmembrane segments (TM). Different 5′ regulatory elements are associated with the different types of gene organization. In the case of the mammalian-type locus, an enhancer sequence (■) located between the J segment and C regions is brought into close association with the ATGCAAAT promoter sequence (●) found upstream of each gene, leading to enhanced transcription of the rearranged gene product. The role of enhancer (□)/promoter relationships in cluster-type genes is not understood at this time (see text). A decamer-(spacer)-nonamer sequence, closely resembling that associated with mammalian T-cell antigen receptor Vβ, is designated in uppercase letters for the bases identical to those of the mammalian consensus and in a lowercase letter (t) for the single base that differs. The spacer distance is 3 nucleotides (3nt). The overall length of the mammalian V_H locus is $\sim 10^2$ times greater than that of the cartilaginous fish locus, and the distance between the most proximal rearranging V and D segments is $>10^2$ times greater for the mammalian locus vs. a cartilaginous fish locus.

the recombination elements is excised and eliminated as circular episomes. The deletion and insertion of nucleotides at the joining junctions, along with somatic mutation (hypermutation) of rearranged genes, further diversify the antibody diversity, resulting in an extraordinarily complex repertoire of specificities.

During the past several years, our laboratory has identified immunoglobulin genes in species that represent major phylogenetic groups of vertebrates (Litman et al. 1989). Taken together, the descriptions of immunoglobulin genes in bony fishes (Amemiya and Litman 1990, 1991), amphibians (Schwager et al. 1988, 1989; Haire et al. 1990, 1991), avians (Reynaud et al. 1987, 1989), and mammals (Blackwell and Alt 1988), as well as ongoing studies in our laboratory that involve other phylogenetically important species, are revealing the overall patterns in the evolution of immunoglobulin gene structure and diversity. In all of these species, segmental organization and DNA sequence–mediated selective rearrangement in somatic tissues, the most distinctive features of the immunoglobulin gene system, are highly conserved. The nucleotide and predicted peptide structures of individual V_H, V_L, D_H, J_H, J_L, and some C_H elements, in most instances, are highly conserved, as are the recombination signal sequences (RSSs) flanking the V_H, V_L, D_H, J_H, and J_L segmental elements (Litman et al. 1985*a;* Hinds and Litman 1986; Kokubu et al. 1988*b*). In addition, the exon-intron organization of the C_H elements in *Heterodontus francisci* (horned shark), a phylogenetically primitive vertebrate, is equivalent to that described in mammals (Kokubu et al. 1988*a*). To a certain degree, some of these findings are not unexpected, since in many cases the genes in lower vertebrates were detected by cross-hybridization with a mammalian immunoglobulin heavy-chain variable-region gene-specific probe (Litman et al. 1983, 1985*a*). Typically, any gene detected by using this procedure would have to exhibit ≥60% nucleotide sequence similarity. Although it is informative to compare the sequences of these genes with one another, it must be emphasized that the V_H elements are members of extensively diversified multigene families and that few guidelines can be applied to determine whether similarities and/or differences in gene sequence reflect orthologous evolutionary relationships (Litman et al. 1985*b*). For this reason, the emphasis in the paper will be on the organization of these gene segments and on the relationships of gene organization to function.

Patterns of Gene Organization in Lower Vertebrates

As indicated above, the individual segmental elements of immunoglobulins in different vertebrate species are remarkably similar at the DNA and predicted peptide-sequence levels, even in cases where the segmental elements, e.g., D_H and J_H, have not been selected by cross-hybridization directly but, rather, have been detected by sequencing of an adjacent DNA segment(s), i.e., the gene segment(s) is in close chromosomal proximity to a gene that was localized by cross-hybridization or that has been identified in a cDNA (detected with a probe complementing different segmental elements) (Hinds and Litman 1986; Harding et al. 1990*a,* 1990*b;* Amemiya and Litman 1991). However, the genomic organization of these genes varies markedly in different vertebrates. In *Heterodontus,* immunoglobulin heavy-chain gene segmental elements (regions) are in ~16,000-bp linkage "clusters" consisting of a single V_H, two D_H, one J_H, and one C_H region, encoded by six exons (Hinds and Litman 1986; Kokubu et al. 1988*a,* 1988*b*). A diagrammatic representation of such a cluster and its somatically rearranged form is shown in figure 1 (*bottom*). We have resolved the entire sequence of a cluster in order to identify and localize regulatory sequences (K. R. Hinds-Frey and G. W. Litman, unpublished data). Hundreds of these clusters

are present in the genome of an individual, contrasting markedly with the organization of mammalian immunoglobulin genes, which form a single locus. Preliminary studies using in situ hybridization have shown immunoglobulin gene clusters in both *Heterodontus* and *Raja eglanteria* (skate) to be present at different chromosomal locations (C. T. Amemiya, J. P. Rast, and G. W. Litman, unpublished data). In *Heterodontus,* antibody gene diversity presumably is generated by the same rearrangement mechanism that is employed by higher vertebrate species, but recombination does not occur between segmental elements in different clusters. First detected in *Heterodontus,* this form of gene organization also has been found in Batoids, representative of a different elasmobranch order (Harding et al. 1990*b*), as well as in a chimera, *Hydrolagus colliei* (ratfish), a species that represents an independent evolutionary line of the chondrichthyes (J. P. Rast, C. T. Amemiya, and G. W. Litman, unpublished data).

A second markedly different heavy-chain constant-region isotype has been described in *R. kenojei* (Kobayashi et al. 1984), which we have characterized at the gene level (Harding et al. 1990*a*). This gene, designated "IgX," has the same pattern of segmental organization as do the other IgM (μ)-type elasmobranch heavy-chain genes (M. Hulst and G. W. Litman, unpublished data) and, likewise, is present on multiple chromosomes (and presumably at different sites on single chromosomes). In early development, both IgM and IgX (encoded at separate loci) are expressed on single lymphocytes, contrasting with allelic exclusion that is associated with immunoglobulin gene expression in higher vertebrates. These findings suggest that in one major group of vertebrates, a unique form of immunoglobulin gene organization is found, and this has not been observed in any other vertebrate group thus far examined. The relationship between this type of gene organization and the generation of antibody diversity will be discussed later in this paper. Rigorous comparisons of sequences of the framework (FR) segments of genes detected in cartilaginous fishes and other vertebrates have shown that the genes from cartilaginous fishes form a monophyletic group, correlating well with the collective observations regarding gene organization (T. Ota and M. Nei, personal communication).

Another unusual feature of immunoglobulin genes in cartilaginous fishes is the germ-line joining of segmental elements of approximately one-half of the clusters in *Heterodontus* (Kokubu et al. 1988*b;* Harding et al. 1990*b;* fig. 2). In all cases thus far examined, fully joined (VDJ) segments maintain a correct reading frame (Kokubu et al. 1988*a,* and unpublished data); if joining were random, only one sequence in three would be expected to correctly encode the J_H element, which has a well-conserved and thus recognizable peptide structure. For V_HD_H-J_H-joined elements in *Heterodontus* (Kokubu et al. 1988*b*) and V_HD_1-D_2J_H-joined elements in *R. erinacea* (little skate) (Harding et al. 1990*b;* see below), the RSSs, found at the 3′ of D_H in *Heterodontus* and at the 3′ of D_1 and 5′ of D_2 in *Raja,* are similar to those found in the nonjoined clusters that have been shown to be expressed, suggesting that these can potentially rearrange; none of these contains stop codons, frameshift mutations, etc. The V_H coding segments of joined genes are not related appreciably more to one another than to the nonjoined clusters (Kokubu et al. 1988*b;* K. R. Hinds-Frey and G. W. Litman, unpublished data). However, the D regions of joined genes, by comparison with D elements found in the unjoined clusters, differ markedly, suggesting that the former may encode "predetermined" specificities that could be important components of host defense in these species. The joined form of organization may represent a "regression" from the segmental rearrangement system and may be of decided host survival advantage in species where the kinetics of antigen-driven cell selection/proliferation

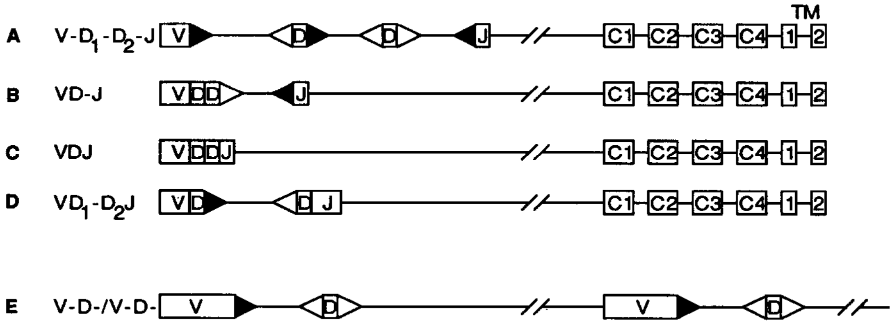

FIG. 2.—Different patterns of cluster-type gene organization in the germ-line cells of cartilaginous fish and *Latimeria,* living coelacanth. Shown (from top to bottom) are (A) V-D_1-D_2-J-type, ex. *Heterodontus;* (B) VD-J-type, ex. *Heterodontus;* (C) VDJ-type, ex. *Heterodontus;* (D) VD-DJ-type, ex. *Raja;* and (E) V-D-//-V-D-type, ex. *Latimeria.* V = variable; D = diversity; J = joining; C (1–4) = constant-region exons 1–4; and TM (1 and 2) = constant-region transmembrane exons. Blackened triangles (▶ and ◀) denote recombination signal sequences with 22 or 23 nucleotide spacers; and unblackened triangles (▷ and ◁) denote recombination signal sequences with 12 nucleotide spacers. Intersegmental distances are not to scale.

are not as efficient or otherwise differ from that observed in higher vertebrates. The chimera also exhibit germ-line joining, and, in one case, unequivocal evidence has been obtained, by using polymerase chain reaction (PCR), that a cDNA transcript of a joined gene is expressed (J. P. Rast, C. T. Amemiya, and G. W. Litman, unpublished data). Similarly, all of the members of a major family of light-chain genes, recovered from a *Raja* genomic DNA library, are "joined" (M. J. Shamblott, M. Hulst, and G. W. Litman, unpublished data), in marked contrast to the major expressed family of light-chain genes in *Heterodontus,* which are organized in a cluster configuration (V_L-J_L-C_L) but show no evidence of "germ-line joining," despite an exhaustive effort to detect such configurations (Shamblott and Litman 1989). As of yet, we have been unable to detect an unmodified transcript of the *Raja* germ-line-joined light-chain genes, suggesting that either there are other members of this family or the joined elements are targets of somatic change. On the basis of comparisons of light-chain peptide sequences and the sequences of cDNAs that are related to the germ-line-joined gene, it is likely that *Raja* possesses additional light-chain gene families (M. Hulst, M. J. Shamblott, and G. W. Litman, unpublished data), as does *Heterodontus* (J. P. Rast, M. J. Shamblott, and G. W. Litman, unpublished data).

A third unique feature of the organization of heavy-chain genes in the cartilaginous fishes is the absence of the regulatory octamer (ATGCAAAT or ATTTGCAT) associated with the B cell–specific immunoglobulin promoter, an invariant component of immunoglobulin gene expression in teleosts and tetrapods (see below). The absence of this sequence motif at its characteristic position 5′ of the "TATA"-equivalent sequence in cartilaginous fishes suggests that the tissue-specific regulation of immunoglobulin gene expression may differ in these species (fig. 1). A decamer/nonamer sequence that is associated with regulation of the β T cell antigen receptor (TCR) (Lee and Davis 1988) has been detected upstream of the initiation codon in *Heterodontus* immunoglobulin heavy-chain genes (Kokubu et al. 1988*b;* E. Davidson, personal communication). However, *Heterodontus* light-chain genes possess the octamer sequence in a 5′ position that corresponds to the location of this regulatory sequence in tetrapods (Shamblott and Litman 1989), whereas in *Raja* at least some light-chain

genes lack an octamer (M. J. Shamblott and G. W. Litman, unpublished data). In summary, there are major differences in the organization and regulation of immunoglobulin heavy- (and light-) chain genes in the most phylogenetically primitive jawed vertebrates. High interspecific sequence similarity in V_H, the exon organization of C_H, and the differential processing of 3′ secretory and transmembrane sequences are characteristics of higher-vertebrate immunoglobulin and, specifically, of μ-type heavy chains. However, (1) the close linkage of segmental elements, (2) presence of two D_H elements, (3) absence of an immunoglobulin octamer with the presence of a TCR nonamer/decamer (in *Heterodontus*), and (4) utilization of D_H in all three reading frames (K. R. Hinds-Frey and G. W. Litman, unpublished data) suggest that the rearranging genes in the cartilaginous fish also are related closely (in terms of overall organization) to TCRs. It appears that the heavy-chain locus reflects properties of both TCR and immunoglobulin gene loci. These relationships are shown in figure 3. In the absence of evidence either for or against the presence of TCRs in cartilaginous fishes, one interpretation of this chimeric gene structure/organization is that the cluster-type immunoglobulin heavy-chain gene(s) in *Heterodontus* reflects the structure of the ancestral genes from which the immunoglobulin and TCR systems evolved.

Immunoglobulin Gene Homologues in the Jawless Vertebrates

Two important findings that may relate to the nature of immunoglobulin gene structure in the more primitive jawless vertebrates emerge from these studies: (1) at least four distinct sequence regions—variable regions FR2 and FR3, J_H, and the constant-region secretory/transmembrane regions—are highly conserved, and (2) close segmental linkage most likely is ancestral to the extended-locus form of gene organization. Thus, gene structure and organization in the primitive vertebrates may be homologous to the patterns noted in the cartilaginous fishes. Alternatively, the gene structure and organization may be more closely akin to that observed in the bony

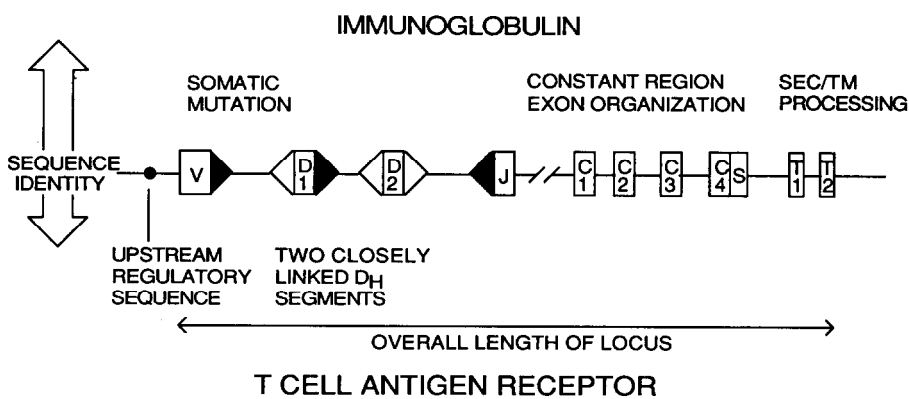

FIG. 3.—Relationships of "immunoglobulin" gene clusters in cartilaginous fish to mammalian immunoglobulin vs. T-cell antigen receptor. Notations above the horizontal axis of the figure imply immunoglobulin-like characteristics, and those below the figure imply T-cell antigen receptor-like characteristics. Overall, the V-D-J segmental elements share a greater degree of similarity with immunoglobulins than with T-cell antigen receptors, as indicated by the length of the respective arrows. The exons of a μ-type constant region (shown after — // —) are indicated as constant region (C) 1–4; C4 contains the secretory (S) sequence and two transmembrane exons (T1–T2). Blackened triangles (▶ and ◀) denote recombination signal sequences with 22 or 23 nucleotide spacers; and unblackened triangles (▷ and ◁) denote recombination signal sequences with 12 nucleotide spacers. Intersegmental distances are not to scale.

fishes and more advanced vertebrates or the "antibodies" will not be significantly (or will be only distantly) related to immunoglobulin, on the basis of the "shared" criteria indicated above. It would seem that relatedness to the cartilaginous fish pattern is the most likely. However, exhaustive efforts to detect immunoglobulin gene homologues in both lampreys and hagfishes, including the screening of cDNA libraries constructed from tissue sources in which histologically defined lymphoid-like elements are detected, have been uniformly unsuccessful with a variety of probes that complement the highly conserved regions indicated above. In addition, degenerate primer PCR analyses directed at sequence regions exhibiting elevated levels of nucleotide sequence identity also have proved unsuccessful, although these same methods yield appropriate products among a broad representation of jawed vertebrates. Although we (Varner et al. 1991) and others (Kobayashi et al. 1985; Hanley et al. 1990) have identified immunoglobulin-like heterodimers that exhibit limited amino acid sequence similarity with other vertebrate immunoglobulins, large numbers of peptides isolated from these heterodimers lack sequence similarity with either heavy- or light-chain genes (Varner et al. 1991). Recently, it has been shown that this molecule represents a complement polypeptide (Ishiguro et al. 1992).

These findings come as no surprise, in view of the peculiar nature of the "antibody" response in lamprey. Antibody is recovered consistently in an ~300–500-kD, highly labile serum fraction. Immunization with antigenetically complex human group "O" erythrocytes results in high-titer antibody being directed to the H surface antigen; no cross-reactivity with type-A or -B cells is evident. Similarly, high-titer antibody to a *Brucella abortus* vaccine does not crossreact with "O" erythrocytes and exhibits coincidental chromatographic behavior with the anti-"O" antibody. The physicochemical properties of the serum protein fraction associated with antibody are dissimilar to the properties associated with conventional antibody and lack the multimeric disulfide bonding and stability of antibody that are found in the most phylogenetically distant, modern forms of jawed vertebrates. While homologous forms of immunoglobulin may be present in basal protochordate species, the extant jawless vertebrates may lack "immunoglobulin" that is homologous to that seen in representatives of the more advanced vertebrates. A completely different system of inducible, specific immunity not involving members of the immunoglobulin gene superfamily, as presently recognized, may exist in these species.

Restriction in the V_H Repertoire in the Early Vertebrates

Estimates of the size of the V_H repertoire in lower vertebrate species are based on comparisons of both germ-line and rearranged, expressed (cDNA) immunoglobulin genes. Our initial analyses of V_H elements in *Heterodontus,* including genes that exhibit germ-line joining, indicate that, with a single exception, all V_H elements are members of a single family (type 1), i.e., the genes possess ≥70% absolute nucleotide similarity (typically V_H elements in *Heterodontus* show ~90% similarity) (Kokubu et al. 1988*b*). The single exceptional gene (type 2) has ~61% sequence similarity with the most closely related members of the type 1 *Heterodontus* V_H family (Kokubu et al. 1988*b*). The type 2 gene possesses an inverted intervening sequence between D_1 and D_2 that changes the polarity of the RSSs, permitting $V_HD_1J_H$ (or $V_HD_1D_2J_H$) joinings—unlike the type 1 cluster(s), in which $V_HD_2J_H$ (or $V_HD_1D_2J_H$) joinings are possible. Diagrammatic representations of type 1 and type 2 gene clusters are shown in figure 4.

To extend these analyses, large numbers (~700) of individual phage from a *Heterodontus* genomic DNA library that was selected with either C_H or J_H probes and

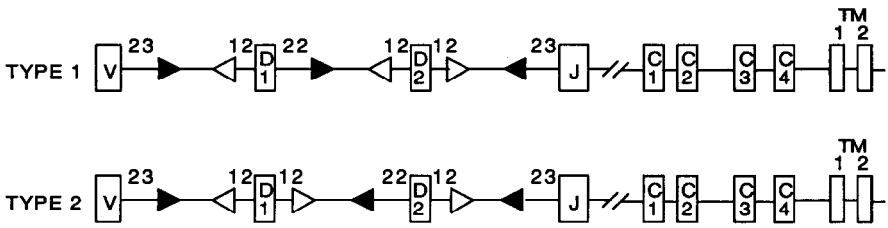

FIG. 4.—Comparison of type 1 and type 2 patterns of gene organization in *Heterodontus* immunoglobulin. Recombination signal sequences are denoted by blackened triangles (▶ and ◀) when they contain 22 or 23 nucleotide spacers or by unblackened triangles (▷ and ◁) when they contain 12 nucleotide spacers. The 23/12, 12/23 spacer rule states that recombination only occurs between an element possessing a 12-bp spacer and an opposing element possessing a 22- or 23-bp spacer (or the opposite relationship). From comparison of intervening sequences, it is apparent that the type 2 gene arose by inversion of the D_1-D_2 intervening sequence (Kokubu et al. 1988*b*). The type 1 gene can recombine as V-D_1-D_2-J or V-D_2-J, whereas the type 2 gene can recombine as V-D_1-D_2-J or V-D_1-J. Intersegmental distances are not to scale.

that should detect all immunoglobulin genes regardless of V_H family were screened with a V_H probe. Under the hybridization conditions employed, distant members of a unique V_H gene family (see below) would be expected to be V_H^-, J_H^+, or C_H. Wherever a lack of concordance between C_H and V_H hybridization was observed, C_H elements could be mapped near a λ arm, i.e., V_H was not integrated during the cloning. In a similar experiment, in which V_H, J_H, and C_H probes were used to detect possible V_H^+, J_H^-, or C_H^- variants, only two such clones were detected; both contained multiple V_H segments, of which some are clearly pseudogenes. Although one of the clones detected in this experiment is a variant of the "typical" cluster, i.e., the gene is V_H-J_H-C_H (D_H is deleted), this gene does not deviate from the basic cluster-type organization (K. R. Hinds-Frey and G. W. Litman, unpublished data). Thus the primary antibody repertoire in *Heterodontus* appears to derive from a sizable number of genes in the V_H-D_1-D_2-J_H configuration that are closely related members of a single gene family and from a single gene cluster that appears to be the monotypic representative of a second V_H gene family.

Despite the limitation in V_H gene diversity, N-region, junctional, and complementarity-determining region (CDR) diversity are extensive (Kokubu et al. 1988*b*; K. R. Hinds-Frey and G. W. Litman, unpublished data). Using an oligonucleotide-based approach that relates "parental" germ-line clusters to specific mRNA transcripts, we recently have been able to provide unequivocal evidence for somatic variation in both CDR and FR segments of rearranged, presumably functionally expressed *Heterodontus* genes. Similar observations have been made with the rearrangement products of the second (monotypic) V_H-type element (not illustrated). While it can be argued that all of the differences that we observe arise from mutated germ-line variants, which we are not able to detect or which are not integrated into the libraries, the number of such gene clusters that would be required is inconsistent with our current understanding of the size of the heavy-chain gene cluster in *Heterodontus* (Kokubu et al. 1987). *Heterodontus* exhibits two significant restrictions in the immune repertoire: (1) sequence diversity in V_H, D_H, and J_H elements is considerably lower than in higher vertebrates, and (2) combinatorial rearrangement between segmental elements found in different clusters has not been detected. In view of the marked restrictions in the D_H and J_H repertoires, this latter effect may be insignificant, i.e., no selective advantage would be realized in recombination between clusters. The near absence of variation

in D_H is particularly intriguing, as these segments, which are located <350 nucleotides from V_H, maintain sequence stability in different chromosomal locations, while the V_H elements exhibit appreciable sequence variation in CDR2. On the basis of analyses of rearrangement products, it is evident that the role of D_H elements in this species is in mediating/facilitating joining rather than in contributing unique coding specificities.

The Origins of the Single-Extended-Locus Form of Immunoglobulin Gene Organization

Having established the presence of a rearranging immunoglobulin gene system in the cartilaginous fish, we would expect that the system is present in all higher vertebrate species, including the bony fishes, dipnoi, amphibians, reptiles, avians, and mammals. This essentially is the case; however, there are major variations in the organization of the rearranging gene system, which is remarkable for a system that is associated with only a single primary function, i.e., antibody specificity. In the bony fishes, we and others have shown that V_H elements are tandemly linked upstream of J_H (and supposedly D_H elements) and C_H elements (mammalian-type organization; fig. 1). Preliminary evidence from other bony fishes, including a representative "chondrostean" and a representative "holostean," indicates variation in the numbers of V_H and C_H elements (Amemiya and Litman 1991). Diversification of V_H elements into specific families has been established in teleost fish (Amemiya and Litman 1990; Ghaffari and Lobb 1991), in marked contrast to *Heterodontus,* which possesses one large V_H element family and a single monotypic second V_H element family. In the coelacanth, *Latimeria chalumnae,* two distinct forms of gene organization are encountered. Variable regions either are tandemly linked or are associated with a single D segment in the same close linkage arrangement as seen in *Heterodontus* but lacking the closely linked D_H2 and J_H elements, as well as C_H elements (fig. 2). The tandemly linked variable regions are pseudogenes, which may have lost recognizable, flanking D_H elements. This has led us to propose that *Latimeria* may occupy a phylogenetic point intermediate between the cartilaginous and bony fishes, a conclusion that is consistent with many anatomical, physiological, and endocrinological considerations (C. T. Amemiya and G. W. Litman, unpublished data). Extensive diversification of V_H families is evident in *Xenopus laevis,* with 11 distinct, highly interspersed families now identified (Haire et al. 1990, 1991). *Xenopus* also possesses three constant-region isotypes (Schwager et al. 1988; Amemiya et al. 1989; Haire et al. 1989). Extensive V_H element families also are found in reptiles, with evidence for V_H elements in four different chromosomal linkage groups (C. T. Amemiya and G. W. Litman, unpublished data).

A unique variation on the general theme of segmental rearrangement is found in avians. For both the heavy- and light-chain genes in *Gallus domesticus,* a single, functional gene is the target for extensive gene conversion by flanking pseudogenes. The evolutionary mechanism(s) that gave rise to this unique variation may relate to extensive chromosomal dispersal of the immunoglobulin gene loci, as noted above for an ancestral reptilian species, and to loss of chromosomes during speciation. Similar findings have been made with another bird (McCormack et al. 1989). With the exception of the avians, two primary patterns of gene organization are found throughout the vertebrates; the cluster type and the extended, single-tandem-locus type. Furthermore, in each of the three major types of gene organization—cluster, extended locus, and single gene (avian-type)—the organization and mechanisms for generating diversity are identical for the heavy- and light-chain gene families, within the represen-

tative taxa. Such observations suggest that these independent loci are coevolving, which we have attributed to an obligatory need for utilization of the same recombination mechanisms (Shamblott and Litman 1989).

Conclusions

On the basis of these observations and additional data from our laboratory, hypotheses about the principal events that gave rise to the immunoglobulin gene systems of contemporary vertebrates can be made. The first event in the evolution of this rearranging gene system presumably involved the segmentation of a single exon encoding a protein that possessed the basic conformation of an immunoglobulin domain. This may have been mediated by a transposon containing an RSS-like sequence or by intragenic recombination. We have described such a transposon-like structure within an immunoglobulin intervening sequence (Litman et al. 1985*b*). In functional terms, rearrangement of the segmented gene could introduce, at the joining junction, sequence variation analogous to junctional and N-type diversity seen in higher vertebrates. This segmented exon may have evolved into a V_H-J_H-C_H- or V_H-D_H-J_H-C_H-type structure, through duplication and/or recombination; however, this discussion should not be taken to suggest that a heavy-chain gene preceded a light-chain gene in the evolution of this rearranging gene system.

There are reasons to believe that a light-chainlike structure may have preceded the heavy chain in the evolution of the rearranging gene system: it is noteworthy that light chains can form a homodimer that simulates the folding patterns of immunoglobulin heterodimers. Without readdressing whether TCRs preceded immunoglobulins in phylogeny (discussed above), we can say that the most conservative interpretation is that a primordial, segmented V-D-J-type structure duplicated, giving rise to V_1-J_1-C_1 and V_2-J_2-C_2. One of these underwent an internal rearrangement, recombination, or secondary inversion by a transposon, yielding V_2-D_2-J_2-C_2. Both the immunoglobulin and TCR heterodimers are encoded by pairs of D-containing and non-D-containing segmental rearranging genes, e.g., heavy (D_H) and light chain, α and β (D_β) TCR, and γ and δ (D_δ) TCR.

In their most fundamental role, the D_H elements, which in at least one phylogenetically early vertebrate (*Heterodontus*) lack appreciable sequence variation, serve as mediators of genetic change, through the recombination process, rather than as contributors of intrinsic genetic diversity in the antibody repertoire, through sequence variability. Further development of this system presumably involved unit duplication of these clusters. Sequence diversification in the CDRs of the V_H elements would expand the range of combining site specificities. An alternative pathway of gene evolution resulted in duplication (and diversification) of the individual V_H, D_H, and J_H elements in a single, extended cluster. Significant chromosomal distances between the individual elements would serve to promote nonrestrictive rearrangement of segmental elements, facilitating combinatorial diversity. Further diversification of V_H FRs resulted in the separate V_H element families found in teleosts and tetrapods, and duplication and diversification of C_H elements expanded constant-region diversity. Throughout this process, somatic mutation (and hypermutation) targeted to the immunoglobulin locus most likely became an increasingly important factor in the generation of antibody diversity. Alternatively, somatic mutation may have been an early component in the system, and this process coevolved with cellular selection mechanisms that could improve the quality (affinity and specificity) of antibody reactivity to a defined antigen.

It may never be possible to project the time frame in which these various changes

took place, other than to state that the first point in vertebrate evolution where heterodimeric (mammalian-like) immunoglobulin genes can be recognized is in the phylogenetically primitive cartilaginous fishes. The recent findings that the heterodimeric structure found in an Agnathan, the hagfish (*Eptatretus stoutii*), has only limited amino acid sequence identity with immunoglobulin (and many other peptides from the hagfish heavy and light chains have no identity with immunoglobulin) are not unexpected, since extensive efforts using molecular genetic approaches to detect immunoglobulin genes in this species [and lamprey (*Petromyzon marinus*), another Agnathan] have been unsuccessful, as have been efforts to identify plasma cells, the terminally differentiated B lymphocyte form. A simplified phylogeny of inducible immunity is used to emphasize this point (fig. 5), as well as to illustrate the major change in gene organization that occurred at the time of the divergence of the cartilaginous fish lineage versus the bony fish lineage. The search for the origins of the B cell repertoire may need to be restricted to the earliest cartilaginous fishes, as the jawless vertebrates either may lack recognizable immunoglobulin genes or rely on other mechanisms and molecules, which may or may not be related to contemporary members of the immunoglobulin gene superfamily, to mount effective challenges to

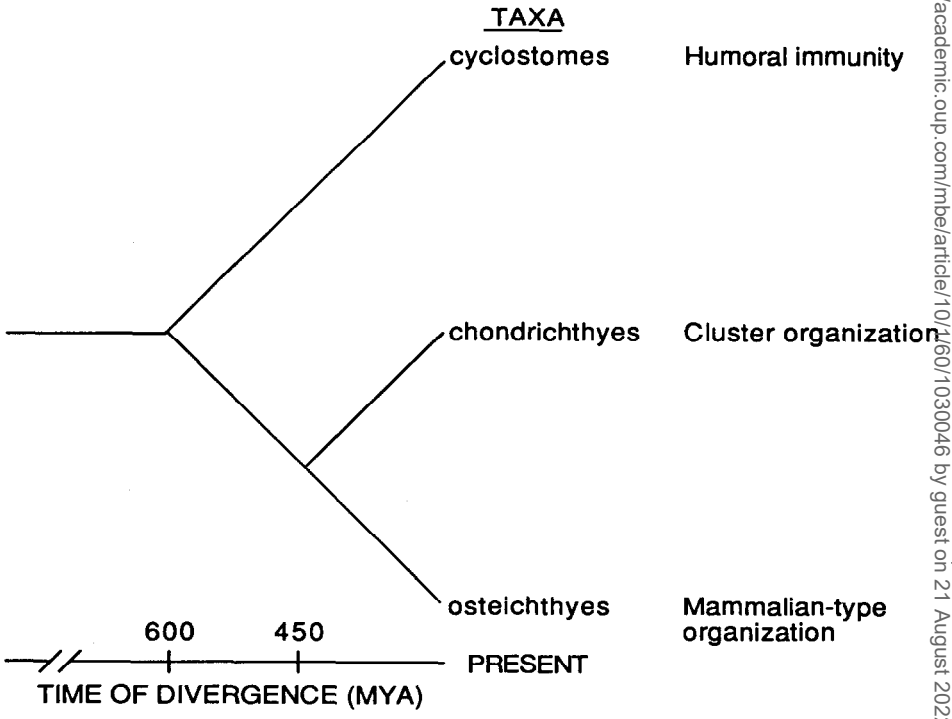

FIG. 5.—Simplified view of the vertebrate radiation as this relates to the occurrence of two different patterns of vertebrate immunoglobulin heavy-chain gene organization and inducible humoral immunity of an as yet unknown type in jawless vertebrates (cyclostomes). This macroscheme divides the phylogeny of inducible immunity into three major branches; however, no a priori genetic relationship is implied between the inducible system found in the contemporary cartilaginous (chondrichthyes) and bony (osteichthyes) fishes (and subsequent vertebrate radiations) vs. that found in the cyclostomes. Major organizational differences, associated with extensive sequence identities between segmental elements, are the basis for the branching shown for the radiations of the chondrichthyes and osteichthyes. The branch points are indicated relative to approximate geologic time [in Mya (MYA)].

invading foreign pathogens. The Agnathans may possess highly derived mechanisms for antigen reactivity, but the more phylogenetically primitive protochordates may retain characteristics that are related more closely to immunoglobulin found in contemporary species. Phylogenetic investigations into the evolution and phylogenetic origins of this complex system have provided and will continue to provide essential information as to the mechanisms of evolution of this and other multigenic systems that comprise multiple cell types.

Acknowledgment

The editorial assistance of Barbara Pryor is appreciated. This work was supported by National Institutes of Health grants AI-23338 and GM-38656.

LITERATURE CITED

AMEMIYA, C. T., R. N. HAIRE, and G. W. LITMAN. 1989. Nucleotide sequence of a cDNA encoding a third distinct *Xenopus* immunoglobulin heavy chain isotype. Nucleic Acids Res. **17**:5388.

AMEMIYA, C. T., and G. W. LITMAN. 1990. The complete nucleotide sequence of an immunoglobulin heavy chain gene and analysis of immunoglobulin gene organization in primitive teleost species. Proc. Natl. Acad. Sci. USA **87**:811–815.

———. 1991. Early evolution of immunoglobulin genes. Am. Zool. **31**:558–569.

BLACKWELL, T. K., and F. W. ALT. 1988. Immunoglobulin genes. Pp. 1–60 *in* B. A. HAMES and D. M. GLOVER, eds. Molecular immunology. IRL, Oxford.

GHAFFARI, S. H., and C. J. LOBB. 1991. Heavy chain variable region gene families evolved early in phylogeny: Ig complexity in fish. J. Immunol. **146**:1037–1046.

HAIRE, R. N., C. T. AMEMIYA, D. SUZUKI, and G. W. LITMAN. 1990. A high degree of immunoglobulin V_H gene complexity in a lower vertebrate: *Xenopus laevis*. J. Exp. Med. **171**: 1721–1737.

HAIRE, R. N., Y. OHTA, R. T. LITMAN, C. T. AMEMIYA, and G. W. LITMAN. 1991. The genomic organization of immunoglobulin V_H genes in *Xenopus laevis* shows evidence for interspersion of families. Nucleic Acids Res. **19**:3061–3066.

HAIRE, R. N., M. J. SHAMBLOTT, and G. W. LITMAN. 1989. A second *Xenopus* immunoglobulin heavy chain constant region isotype gene. Nucleic Acids Res. **17**:1776.

HANLEY, P. J., I. M. SEPPELT, A. A. GOOLEY, J. W. HOOK, and R. L. RAISON. 1990. Distinct Ig H chains in a primitive vertebrate, *Eptatretus stouti*. J. Immunol. **145**:3823–3828.

HARDING, F. A., C. T. AMEMIYA, R. T. LITMAN, N. COHEN, and G. W. LITMAN. 1990*a*. Two distinct immunoglobulin heavy chain isotypes in a primitive, cartilaginous fish, *Raja erinacea*. Nucleic Acids Res. **18**:6369–6376.

HARDING, F. A., N. COHEN, and G. W. LITMAN. 1990*b*. Immunoglobulin heavy chain gene organization and complexity in the skate, *Raja erinacea*. Nucleic Acids Res. **18**:1015–1020.

HINDS, K. R., and G. W. LITMAN. 1986. Major reorganization of immunoglobulin V_H segmental elements during vertebrate evolution. Nature **320**:546–549.

ISHIGURO, H., K. KOBAYASHI, M. SUZUKI, K. TITANI, S. TOMONAGA, and Y. KUROSAWA. 1992. Isolation of a hagfish gene that encodes a complement component. EMBO J. **11**:829–837.

KOBAYASHI, K., S. TOMONAGA, and K. HAGIWARA. 1985. Isolation and characterization of immunoglobulin of hagfish, *Eptatretus burgeri*, a primitive vertebrate. Mol. Immunol. **22**: 1091–1097.

KOBAYASHI, K., S. TOMONAGA, and T. KAJII. 1984. A second class of immunoglobulin other than IgM present in the serum of a cartilaginous fish, the skate, *Raja kenojei*: isolation and characterization. Mol. Immunol. **21**:397–404.

KOKUBU, F., K. HINDS, R. LITMAN, M. J. SHAMBLOTT, and G. W. LITMAN. 1987. Extensive

families of constant region genes in a phylogenetically primitive vertebrate indicate an additional level of immunoglobulin complexity. Proc. Natl. Acad. Sci. USA **84**:5868–5872.

———. 1988*a*. Complete structure and organization of immunoglobulin heavy chain constant region genes in a phylogenetically primitive vertebrate. EMBO J. 7:1979–1988.

KOKUBU, F., R. LITMAN, M. J. SHAMBLOTT, K. HINDS, and G. W. LITMAN. 1988*b*. Diverse organization of immunoglobulin V_H gene loci in a primitive vertebrate. EMBO J. **7**:3413–3422.

LEE, N. E., and M. M. DAVIS. 1988. T cell receptor β-chain genes in BW5147 and other AKR tumors deletion order of murine Vβ gene segments and possible 5′ regulatory regions. J. Immunol. **140**:1665–1695.

LITMAN, G. W., L. BERGER, K. MURPHY, R. LITMAN, K. R. HINDS, and B. W. ERICKSON. 1985*a*. Immunoglobulin V_H gene structure and diversity in *Heterodontus,* a phylogenetically primitive shark. Proc. Natl. Acad. Sci. USA **82**:2082–2086.

LITMAN, G. W., L. BERGER, K. MURPHY, R. T. LITMAN, K. R. HINDS, C. L. JAHN, and B. W. ERICKSON. 1983. Complete nucleotide sequence of an immunoglobulin V_H gene homologue from *Caiman,* a phylogenetically ancient reptile. Nature **303**:349–352.

LITMAN, G. W., K. MURPHY, L. BERGER, R. LITMAN, K. HINDS, and B. W. ERICKSON. 1985*b*. Complete nucleotide sequences of three V_H genes in *Caiman,* a phylogenetically ancient reptile: evolutionary diversification in coding segments and variation in the structure and organization of recombination elements. Proc. Natl. Acad. Sci. USA **82**:844–848.

LITMAN, G. W., M. J. SHAMBLOTT, R. HAIRE, C. AMEMIYA, H. NISHIKATA, K. HINDS, E. HARDING, R. LITMAN, and J. VARNER. 1989. Evolution of immunoglobulin gene complexity. Pp. 361–368 *in* F. MELCHERS, ed. Progress in immunology, vol. 7. Springer, Berlin.

MCCORMACK, W. T., L. W. TJOELKER, L. M. CARLSON, B. PETRYNIAK, C. F. BARTH, E. H. HUMPHRIES, and C. B. THOMPSON. 1989. Chicken IgL gene rearrangement involves deletion of a circular episome and addition of single nonrandom nucleotides to both coding segments. Cell **56**:785–791.

REYNAUD, C.-A., V. ANQUEZ, H. GRIMAL, and J.-C. WEILL. 1987. A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell **48**:379–388.

REYNAUD, C.-A., A. DAHAN, V. ANQUEZ, and J.-C. WEILL. 1989. Somatic hyperconversion diversifies the single V_H gene of the chicken with a high incidence in the D region. Cell **59**: 171–183.

SCHWAGER, J., M. BÜRCKERT, M. COURTET, and L. DU PASQUIER. 1989. Genetic basis of the antibody repertoire in *Xenopus:* analysis of the V_H diversity. EMBO J. **8**:2989–3001.

SCHWAGER, J., C. A. MIKORYAK, and L. A. STEINER. 1988. Amino acid sequence of heavy chain from *Xenopus laevis* IgM deduced from cDNA sequence: implications for evolution of immunoglobulin domains. Proc. Natl. Acad. Sci. USA **85**:2245–2249.

SHAMBLOTT, M. J., and G. W. LITMAN. 1989. Genomic organization and sequences of immunoglobulin light chain genes in a primitive vertebrate suggest coevolution of immunoglobulin gene organization. EMBO J. **8**:3733–3739.

VARNER, J., P. NEAME, and G. W. LITMAN. 1991. A serum heterodimer from hagfish (*Eptatretus stoutii*) exhibits structural similarity and partial sequence identity with immunoglobulin. Proc. Natl. Acad. Sci. USA **88**:1746–1750.

MASATOSHI NEI, reviewing editor

Received July 13, 1992; revision received August 10, 1992

Accepted August 11, 1992