John P. Cannon
Robert N. Haire
Jonathan P. Rast
Gary W. Litman

The phylogenetic origins of the antigen-binding receptors and somatic diversification mechanisms

Authors' addresses
John P. Cannon[1,2], *Robert N. Haire*[3], *Jonathan P. Rast*[4]
Gary W. Litman[1,2,3]
[1]Department of Molecular Genetics, All Children's Hospital, St. Petersburg, FL, USA.
[2]H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL, USA.
[3]Department of Pediatrics, University of South Florida College of Medicine, Children's Research Institute, St. Petersburg, FL, USA.
[4]Sunnybrook and Women's College, Health Sciences Center, Toronto, Ontario, Canada.

Correspondence to:
Gary W Litman
Department of Pediatrics
University of South Florida College of Medicine
Children's Research Institute
830 First Street South
St. Petersburg, FL 33701
USA
Tel.: +1 727 553 3601
Fax: +1 727 553 3610
E-mail: litmang@allkids.org

Acknowledgements
We thank B. Pryor for editorial assistance and M. Sexton for providing graphic art. GWL is supported by grants from NIH. JPC is supported by a fellowship from the H. Lee Moffitt Cancer Center and Research Institute.

Immunological Reviews 2004
Vol. 200: 12–22

Immunological Reviews
0105-2896

Summary: The adaptive immune system arose in ancestors of the jawed vertebrates approximately 500 million years ago. Homologs of immunoglobulins (Igs), T-cell antigen receptors (TCRs), major histocompatibility complex I (MHC I) and MHC II, and the recombination-activating genes (RAGs) have been identified in all extant classes of jawed vertebrates; however, no definitive homolog of any of these genes has been identified in jawless vertebrates or invertebrates. RAG-mediated recombination and associated junctional diversification of both Ig and TCR genes occurs in all jawed vertebrates. In the case of Igs, somatic variation is expanded further through class switching, gene conversion, and somatic hypermutation. Although the identity of the 'primordial' receptor that was interrupted by the recombination mechanism in jawed vertebrates may never be established, many different families of genes that exhibit predicted characteristics of such a receptor have been described both within and outside the jawed vertebrates. Recent data from various model systems point toward a continuum of immune receptor diversity, encompassing many different families of recognition molecules whose functions are integrated in an organism's response to pathogenic invasion. Various approaches, including both genomic and protein-functional analyses, currently are being applied in jawless vertebrates, protochordates, and other invertebrate deuterostome systems and may yield definitive evidence regarding the presence or absence of adaptive immune homologs in species lacking adaptive immune systems. Such studies have the potential for uncovering previously unknown mechanisms of generating receptor diversity.

Introduction

The adaptive immune system represents a relatively late acquisition of the vertebrate lineage. The question of how such a complex integrated system of somatic cell recombination, selection, and regulation emerged in the ancestors of jawed vertebrates, apparently in a very short window of evolutionary time, is of significant interest to both immunology and evolutionary biology. A mechanism that most likely gave rise to the adaptive immune system has been identified (1, 2) and is discussed elsewhere in this text. However, to understand the evolution of vertebrate adaptive immunity in detail, the characteristics of the immune systems of pre-adaptive ancestors

must be reconstructed. Because we cannot access the molecular biology of the earliest vertebrate life forms, such reconstruction only can be accomplished by reference to extant jawless vertebrates, invertebrate chordates, and other invertebrate deuterostomes, each of which has acquired derived characteristics during the passage of phylogenetic time. Nevertheless, such reference models are potentially, highly instructive, and in the context of this review, a series of questions can be asked: Did the pre-adaptive immune system include lymphocyte-like cells? Did the antigen-specific recognition capacities of lymphocytes emerge before or after the appearance of clonal rearrangement? Did the pre-rearranging immunoglobulin (Ig)-domain ancestors serve as innate immune recognition proteins? If so, were they expressed on lymphocyte-like cells or a different class of immune cell (e.g. macrophages) whose molecular tool kit included some of the regulatory and functional networks that later came to be associated with lymphocytes in jawed vertebrates?

Recognition of non-adaptive immune cells that may help to answer these questions from animals outside of the jawed vertebrates requires the isolation of orthologous molecular markers. The most obvious and definitive targets for isolation are the theoretical orthologs of the molecules associated most intimately with adaptive immune function, namely the rearranging adaptive immune receptors [Ig and T-cell antigen receptor (TCR)], as well as the surface molecules of the major histocompatibility complex I (MHC I) and MHC II. The patterns of evolutionary diversification of all four of these groups of molecules have been established for representatives of all of the major radiations of jawed vertebrates (3, 4). The single overriding principle of adaptive immune receptor evolution is that all jawed vertebrates rearrange genetically separate segments to diversify immune receptors somatically. No homologs of Ig heavy chain (IgH), Ig light chain (IgL), TCR, recombination-activating genes (RAG), or terminal deoxynucleotidyl transferase (TdT) genes have been identified in either of the two extant orders of jawless vertebrates. Within the jawed vertebrates, three major systems of immune receptor rearrangement and diversification can be recognized in three different phylogenetic groups represented by the cartilaginous fish, bony fish, and avians (*Fig. 1*). Ig genes in cartilaginous fish are encoded by hundreds of independent chromosomally dispersed clusters, a large percentage of which have the unusual characteristic of being joined in the germline (see below). Segmental rearrangement of the non-germline-joined clusters is an intracluster phenomenon, i.e. does not involve combinatorial joining, and these species exhibit minimal variation in germline V_H and V_L sequences

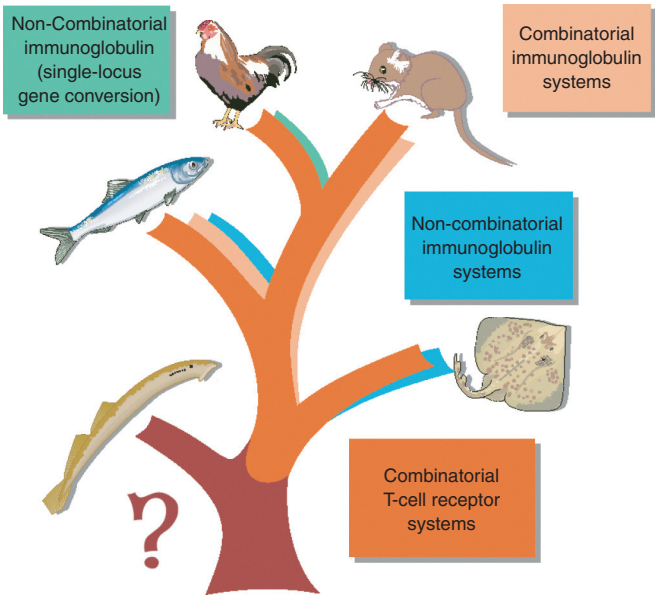

***Fig. 1.* General trends in the generation of somatic diversity of vertebrate adaptive immune receptors.** Four classes of T-cell antigen receptor (TCR), as well as immunoglobulin (Ig) light- and heavy-chain genes, most likely are present in all jawed vertebrates but have not been detected in either jawless vertebrates or invertebrates. The configuration of TCR genes in cartilaginous fish suggests that their combinatorial rearrangement has changed little throughout jawed vertebrate evolution; however, Ig genes in cartilaginous fish exist in a multiple cluster arrangement and do not undergo combinatorial rearrangement, producing in effect an innate form of an adaptive receptor. Somatic diversity is enhanced by various means, including somatic hypermutation and gene conversion. Mechanisms of primary receptor diversification can vary among species; certain classes, e.g. avians and certain mammals, utilize gene conversion rather than somatic hypermutation as the primary means of diversifying the rearranging receptor gene.

(3). Bony fish, which are related more closely to mammals, possess IgH loci with multiple rearranging V_H elements, analogous to those seen in mammals. Some of the IgL loci in bony fish also resemble mammalian forms; however, in several species of bony fish, light-chain genes are in a cluster organization at least superficially similar to that found in cartilaginous fish (5). In contrast to the cartilaginous fish, the sequence divergence within V_H and V_L gene families of bony fish approaches that seen in mammals and amphibians (3). Avians, which represent a more recent phylogenetic divergence from mammals, present another variation on the theme of B-cell receptor (BCR) diversity. In these species, IgH and IgL loci possess only single functional germline V_H–(D_H–J_H) or V_L–J_L, both of which undergo RAG-mediated recombination but (with the exception of D_H) do not undergo combinatorial rearrangement, a shared characteristic with the similarly compact Ig loci of cartilaginous fish. Diversity of the heavy- and light-chain gene loci in avians is achieved by gene conversion targeted to the V_H or V_L after V(D)J recombination.

This process also diversifies Ig genes in gut-associated lymphoid tissue found in a large number of mammals.

Many exceptions and variations of the above rules underscore the extensive plasticity of BCR gene organization and function in jawed vertebrates (reviewed in 3). As the thrust of this text is on the mechanisms of rearrangement, other somatic changes that influence receptor diversity and the evolutionary origins of receptor complexity, the details of the organizational differences in Ig loci between species occupying different phylogenetic positions will not be detailed further.

Germline joining of Ig loci

Before describing gene rearrangement in cartilaginous fish, it is instructive to consider the unusual case of germline joining of antigen-binding receptors. If indeed the rearrangement of antigen-binding receptors was the major defining innovation in immunity that accompanied the emergence of the jawed vertebrate form, why would otherwise 'rearranging gene clusters' be joined in the germline? To understand this phenomenon, it is first necessary to appreciate that the number of independent gene loci in cartilaginous fish exceeds the number found in mammals by almost two orders of magnitude (3). Although the complete genome sequence of a cartilaginous fish species is not yet available, several observations (6, G. Litman, unpublished observation) suggest that the Igs are not linked closely. Thus, unlike the single, combinatorial, rearranging Ig loci found in mammals or the highly efficient gene conversion capabilities of avians [and some mammals (3, 4)], the cartilaginous fish Ig system exhibits enormous basic redundancy. Germline-joined genes can be recovered from sperm (gonad tissue) DNA and thus represent a heritable specificity (3), and the same joined genes can be found in different animals (7). The expression of RAG in gonadal tissues explains how germline joining may have arisen (8). What role germline-joined genes play, if any, in B-cell immunity in cartilaginous fish is considerably more speculative. Although rearrangement and diversification of Ig genes occur early in development in cartilaginous fish (9), selection and expansion may not be sufficiently rapid to confer protection against pathogens. Germline joining could provide a selective advantage in early development, during which embryos are in direct contact with a potentially hostile microbial and viral environment. Although the multigene families encoding IgM, IgW (initially known as IgX), and novel antigen receptor (NAR) classes of Ig-heavy chains consist of both germline-joined and -unjoined members (3), some classes of light-chain genes are fully joined in the germline (10, 11), exhibit intrafamily variation, and apparently undergo somatic mutation (11). While seemingly less efficient in the generation of somatic diversity, germline joining of one (or both) member(s) of the IgH: IgL pair could provide a temporally favorable innate specificity.

Germline rearranging/non-rearranging Ig clusters and the origins of combinatorial joining

It now is quite clear that germline-joined genes are not derived directly from the types of structures that were interrupted at the time of the emergence of adaptive immunity in jawed vertebrates. Rather, they appear to be products of an exceptionally large multigene family in which individual members have secondarily acquired an innate function. Although an accurate description of the full extent of germline joining would require resolution of a very large number of genomic clones, it is estimated that approximately half of the heavy-chain genes are joined in the germline (3). Depending on the species of cartilaginous fish and the class of light-chain gene, the genes encoding light chains can be classified into three groups: completely germline joined, joined at a relatively high percentage, or completely 'unjoined' (3, 8, 12).

Initially, it was unclear whether V(D)J recombination in cartilaginous fish could occur within or among different clusters. However, based on the actual documentation of rearrangement within a cluster, consideration of the genomics of the three classes of Ig heavy-chain genes in cartilaginous fish and our current understanding of Ig rearrangement in mammals, the current consensus is that joining likely is restricted to an intracluster process (13). It would follow that (in the case of the Igs) the emergence of class switching was a later occurrence, after the time of the divergence of the cartilaginous fish. Whether or not somatic mutation preceded or succeeded somatic rearrangement remains open to question (see below).

The combinatorial joining of distinct TCR loci in cartilaginous fishes provides significant insight for reconstructing the state of adaptive immune receptors in the common ancestor of cartilaginous fishes and mammals. TCR genes from both of these modern groups demonstrate that combinatorial diversity existed in their ancestor (14). Furthermore, both IgM and NAR demonstrate the presence of somatic hypermutation mechanisms at this point in evolution. Ig heavy-chain class switching probably originated later, within a sarcopterygian (modern example is lungfish) or tetrapod (modern example is frog) ancestor. Going further back into the evolution of rearranging receptors becomes more speculative, as there are no living representatives of jawed vertebrates more primitive than

cartilaginous fishes. The prototypic rearranging gene most likely was relatively compact, containing an interrupted Ig/TCR-like V region linked tightly (approximately 200–400 base pairs) to a J-like region. This configuration is characteristic of the scattered, individual Ig clusters of modern cartilaginous fish as well as, in some aspects, TCR loci. From this point, diversity increased either by whole cluster duplication, resulting in a cartilaginous fish Ig-like configuration, or by segmental duplication, resulting in a TCR- or bony fish/tetrapod Ig-like configuration. It can be argued that because combinatorial diversity is a characteristic of both TCR genes and bony fish/tetrapod Igs, the prototypic locus diversified through combinatorial joining. Thus, the cartilaginous fish cluster organization would be derived. This argument is far from conclusive, given the extremely dynamic evolutionary history of Ig genes within the jawed vertebrates.

Cluster-type Igs produce extensive junctional diversity

Two observations regarding cluster-type Ig loci in cartilaginous fish are most significant when considering gene rearrangement and the generation of junctional diversity. First, the prototypic IgM (or IgW/IgX)-type gene clusters possess two D elements. By virtue of classical 12/23 pairing of recombination signal sequences, joining can occur at the VD_1, D_1D_2, D_1J, and/or D_2J interfaces (3). The extent of somatic variation at the junctional boundaries is reminiscent of that seen in mammals (12, 13). The second observation relates to Ig (NAR) (15). Unlike the large numbers of related IgM-type genes, the few NAR genes are easily distinguished from one another and can be analyzed unambiguously. NAR possesses a third D region, such as seen in TCRα in mammals and cartilaginous fish, thus presenting an additional joining interface.

RAG1/RAG2 and TdT have been identified in representatives of all major classes of jawed vertebrates. The overall level of predicted amino acid sequence identity between RAG1 and RAG2 identified in jawed vertebrates as phylogenetically divergent as sharks and mammals is approximately 65 and 50%, respectively (16). Furthermore, exon structure, tight physical linkage, and opposite transcriptional orientation of RAG1 and RAG2 are well conserved. Although RAG1 and RAG2 exhibit predicted functional similarity to bacterial transposases, their phylogenetic origin is not entirely clear (17, 18). TdT also has been identified in jawed vertebrate species as phylogenetically divergent as man and cartilaginous fish. TdTs in cartilaginous fish are approximately 70% identical to mammalian TdT, which itself is related to polymerases found in fungi (19). Co-expression of RAG1 and RAG2 proteins is essential for segmental recombination of antigen-receptor genes and is restricted to lymphoid and olfactory sensory neuron cells (20). However, selective expression of RAG1 or RAG2 occurs in a variety of tissues. In the clearnose skate (*Raja eglanteria*), our current primary cartilaginous fish model system, expression of RAG1 is observed in lymphopoietic tissues in the embryo and is primarily thymic in newly hatched animals (9).

The origins of somatic mutation

Although antibody affinity does not change appreciably during immunization of horned sharks (*Heterodontus franciscii*), replacement substitutions in mRNAs (cDNAs) that derive from a single-IgM cluster localize preferentially to the portions of V_H that correspond to Ig V_H complementarity determining regions (CDRs) (3, 13). Far more extensive somatic hypermutation was noted in later studies of the NAR genes in nurse shark (*Ginglymostoma cirratum*) (15, 21). Mutations in nurse shark NAR lack the G–C bias seen in the studies with other shark Ig loci, suggesting a second biochemical mechanism of somatic hypermutation. A more recent comprehensive study of substitution patterns in shark Ig light-chain V regions confirmed the existence of somatic mutation and also identified significant numbers of tandem mutations, an effect that is seen relatively infrequently in mammals (12). Thus, point mutation, tandem mutation, and mutation with G–C bias are associated with the Ig rearrangement process at an early stage in vertebrate phylogeny. However, correlating this phenomenon with actual antibody specificity is not possible at this point. Thus, although Ig genes in cartilaginous fish do not undergo true combinatorial rearrangement, extensive junctional diversity and somatic mutation/hypermutation can be inferred to have evolved relatively 'early' in jawed vertebrate phylogeny, if not earlier at other levels of receptor evolution.

Class switching

Ig heavy-chain class switching, as defined by the association of one rearranged (VDJ-joined) gene with a different heavy-chain isotype in which the intervening genetic sequence is excised from the genome of the B cell undergoing switching, is not a factor in the regulation of Ig expression in cartilaginous fish. This conclusion is supported not only by our current understanding of the IgH loci in these species at the genomic level but also by the patterns of association of specific V_H regions. For example, V_μ only associates with the μ-type constant region and not with IgW/IgX or NAR. An isotype switch (differential RNA processing) similar to that associated

with IgM-IgD has been observed in bony fish, where both μ and another class of Ig heavy chain gene (a δ chain equivalent) are contiguous in the genome. Despite a general similarity in genetic linkage, the IgD-like gene is chimeric and contains the Cμ1 exon, which is required for incorporation of a critical internal disulfide (22). Thus, there is no evidence for true class switching in the bony fishes. The African lungfish (*Protopterus aethiopicus*), a representative of the ancient fleshy finned fish that diverged prior to the amphibians, transcribes both an IgM and an IgW/IgX-like heavy-chain gene; however, these genes appear to be encoded at independent chromosomal sites (23). Amphibians possess two distinct Ig isotypes in addition to IgM and possibly exhibit class switching. Sequences found upstream of the IgX and IgY genes of the African clawed frog (*Xenopus laevis*) possess short repeat sequences, resembling in a general sense the switch recombination sequences found upstream of the various constant region isotypes in mammals; circular DNA containing the repeats has been characterized (24). Although the Ig locus in the amphibians is organized in a mammalian-like manner and may undergo some form of class switching, definitive evidence has not yet been presented that classical mammalian-type switch recombination, as defined above, occurs.

Activation-induced deaminase and somatic diversification

Having considered somatic rearrangement, mutation/hypermutation, and class switching, it would appear that cartilaginous fish somatically rearrange and hypermutate Ig but do not switch classes. The discoveries of activation-induced deaminase (AID) and subsequently uracil-DNA-glycosylase (UDG) have been instrumental in our understanding of the somatic diversification of the Ig repertoire in mammals, as AID is involved in conjunction with UDG in somatic hypermutation, gene conversion, and class switching of Ig genes (25). AID certainly plays a role in gene conversion in avians, although its role in class switching in these species is not known. The recent description of uncoupling of somatic hypermutation/gene conversion and class-switch recombination by deletion of the 10 C-terminal amino acid residues of mouse AID raises the possibility that AID may have acquired its functions separately during evolution (26). Thus, while AID or an AID-like molecule may be common to all jawed vertebrates (25), putative AID orthologs in more phylogenetically primitive species may not necessarily have evolved interactions with cofactors that are obligatory for class switching (27). Genes with significant sequence identity to AID have been identified in several different species of bony fish in which somatic hypermutation occurs; however, it remains to be seen if they exhibit homologous function or represent related APOBEC RNA editing (28, S. Fugmann and N. Trede, personal communications). The multicluster organization of the IgH loci in cartilaginous fish would seem to preclude a role for AID in class switching, although it is possible that AID is an important factor in somatic hypermutation in these species. Until AID and its associated biochemical pathways are identified in species outside of birds and mammals, it remains an open question whether or not AID is required universally for modification of Ig genes in vertebrates.

Given the compressed nature of the Ig loci in cartilaginous fish and the similarity of their recombination signal sequences to those of mammals, it would be interesting to examine somatic rearrangement and germline modifications of transfected cartilaginous fish 'clusters' in mammalian or avian tester lines. However, such studies may be difficult to interpret, owing to differences in transcriptional regulation, e.g. we have reported that the upstream promoter region of V_H genes in one species of cartilaginous fish exhibits similarities to those of mammalian TCRβ (3).

The primordial antigen-binding receptor gene

In spite of our understanding of the primary mechanisms that likely gave rise to somatic recombination (i.e. RAG transposition), neither the identity of the target of the initial event that gave rise to a rearranging receptor nor its basic characteristics are known. While it is unlikely that the actual target molecule or even a related receptor exists in a recognizable form in modern species, it is reasonable to assume that it could have been an innate receptor or molecule involved in the binding of ligands. The function of this receptor would not necessarily have been directly immune-related at the time of integration but could have been related to cell-cell recognition. We can speculate as to what this receptor may have resembled and, perhaps more importantly, discuss manners by which innate receptors could acquire adaptive functions. To do this, it is first necessary to consider the jawless vertebrates, protochordates, and invertebrates, which diverged from the jawed vertebrate lineage hundreds of millions of years ago (*Fig.* 2). Immune function in these organisms presently is being studied by both direct molecular approaches, which rely on a certain degree of relatedness to modern immune receptors, and genomic approaches, which do not necessarily depend on a high degree of sequence similarity to known immune receptors but rather recognize shared motifs and only modest sequence relatedness.

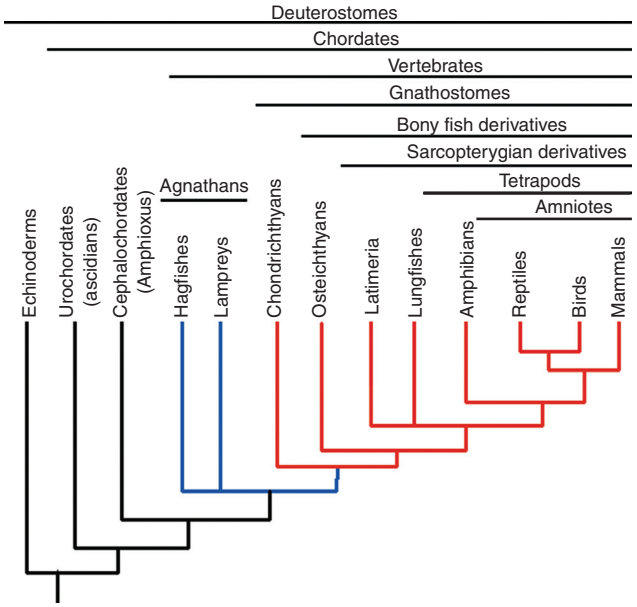

***Fig.* 2. Deuterostome phylogeny.** Major lineages of echinoderms, protochordates, and vertebrates are shown with particular emphasis on species that are amenable to experimentation.

The jawless vertebrates

The jawless vertebrates cannot be viewed as a minor progressive step 'down' from jawed vertebrates. Rather, these species represent ancient lineages that last shared a common ancestor with jawed vertebrates some 500 million years ago. Similarly, they should not be expected to possess a simplified form of Ig/TCR or other Ig-type receptor that can be aligned readily with those identified in modern forms of jawed vertebrates. Jawless vertebrates may possess antigen-recognition receptors that bear no relationship to any member of the Ig superfamily, as has been suggested both by earlier work at the protein-functional level (3) and more recently through molecular genetic approaches (Z. Pancer, personal communication) in one of the surviving orders of jawless vertebrates, the sea lamprey (*Petromyzon marinus*).

Four sets of observations are particularly critical to understand immunity in the sea lamprey. First, this species generates a specific humoral response to immunization with bacteriophage, human erythrocytes, and Gram-negative bacteria. The reactivity towards bacteriophage is resident in what was interpreted to be a heterodimeric molecule that is devoid of interchain disulfide bonds. In contrast, immune reactivity to erythrocytes or Gram-negative bacteria is associated with a multimeric molecule with an estimated mass of approximately 310 Kd. Unlike conventional Igs, the protein was shown to possess a high content of α-helical secondary structure. Both observations emphasize that the inducible, antigen-specific reactivity is highly labile, an uncommon feature of Igs. The second group of data derives from an extensive series of investigations that sought to identify homologs of Ig, TCR, RAG, TdT, MHC I, and MHC II in sea lamprey. Using both conventional cross-hybridization and expressed sequence tag (EST) approaches, none of these canonical adaptive immunity-related genes have been identified. However, both approaches have significant limitations; in the former, it is necessary to have a region of significant contiguous sequence identity or two regions of known sequence for polymerase chain reaction (PCR) amplification. Even though we ultimately were able to demonstrate that these regions could be as short as three amino acid codons (14), the method does require accurate predictions for both forward and reverse primers and minimization of spurious priming of other genetic regions during amplification. The EST approach is particularly complicated by difficulties in normalizing libraries as well as in addressing cell lineage and developmental stage-specific effects on putative immune gene expression. The third set of observations comes from two major lamprey EST screens, which failed to reveal any V region-containing products but concluded that some of the ESTs identified are 'consistent' with genes expressed by lymphoid lineages of higher vertebrates (29, 30). V-like regions are found in many classes of proteins other than antigen-binding receptors; however, they were not encountered at the levels of transcript representation examined, even though the RNA was derived from cells possessing convincing lymphoid-like characteristics. A question that follows is whether or not there was bias in the construction of the libraries, as EST approaches are confounded by a number of factors that influence representation of rare transcripts. Negative findings in the EST screens cannot be considered definitive evidence that V-type genes related to adaptive immunity are absent in the sea lamprey.

We have explored a fourth line of investigation of V genes in lamprey utilizing a recently described cDNA cloning method termed Amptrap, which incorporates five different levels of selection to maximize identification of transcripts encoding secreted or transmembrane proteins. This method requires that only a single sequence region of three amino acid identities be postulated, of which the middle position in certain cases can be an unknown (31). Employing this approach, we thus far have been able to identify four different V-type molecules, including a putative homolog of the Lutheran antigen and a nectin-like molecule in the sea lamprey. A third gene, encoding a putative transmembrane protein with a single large V-type ectodomain, cannot be assigned to any known protein family. The fourth gene encodes a secreted molecule

containing an unequivocal V-type domain, which at this point is the most interesting in terms of a potential link to the adaptive immune system. This molecule contains several charged C-terminal amino acids (both positive and negative), a particular cysteine distribution in the leader, and a trp-glu (WE) amino acid sequence motif near the C-terminus (J. Cannon, R. Haire, and G. Litman, unpublished observations), all of which are shared structural characteristics with mammalian VpreB, which functions in early stages of B-cell development by facilitating BCR presentation on the lymphocyte membrane (32). The overall sequence identity of the lamprey molecule to VpreB would not score it as being a significant homolog; however, in that the charged C-terminus may be the critical aspect in pre-receptor display, it is not unexpected that the sequences in the V domain are diverged, given the phylogenetic distances at issue. To our knowledge, single, secreted non-glycosylphosphatidylinositol-linked V domains are exceedingly rare in mammals in molecules other than VpreB family members. Perhaps a molecule that has a significant role in B-cell ontogeny in mammals has a counterpart in species that we presently do not consider as possessing adaptive immunity.

The characteristics of 'pre-adaptive' immune cells in jawless vertebrates remain an enigma. The somewhat ambiguous molecular findings to date regarding the presence or absence of such cells are not completely unexpected, setting aside the overly simplistic concept that a lamprey (or hagfish, representative of the other extant order of jawless vertebrates) reflects a primitive state. More appropriately, it is necessary to consider not only that the jawless–jawed vertebrate split is ancient but also that the remaining extant forms of jawless vertebrates are highly derived. Interestingly, the current studies suggest that the specialization of the lymphocyte may have preceded the development of the rearranging Ig and TCR genes, and that some of the factors that govern the commitment to the 'modern' lymphocyte lineage are present in jawless fish. The subsequent specialization of additional transcription factors may have been essential for the high degree of lymphocyte complexity (multiple organs, organ-specific development, etc.) seen in representatives of early jawed vertebrates (9). In the case of jawless vertebrates, the structural definition of adaptive immunity as a lymphocyte-Ig/TCR-associated process likely will need to be modified to encompass broader molecular classes of lymphocyte receptors. Furthermore, resolution of functional antigen-binding receptors in jawless vertebrates does not automatically mean that primordial Ig/TCR molecules will be found. Two central questions remain: (i) What was the nature of the primordial lymphocyte antigen receptor and (ii) Does any modern species outside of the phylogenetic level of the jawed vertebrates possess an ortholog? Given the enormous anatomic and physiological specialization of lamprey and hagfish, these species may not be the best place to look for answers.

Immunity in protochordates and echinoderms

The protochordates represent another significant group for exploring the phylogenetic progression of immunity (*Fig.* 2). Our interest in these species is based on the possibility that they manifest a type of recognition that is more representative of an ancestral form than would be found in the highly derived jawless vertebrates. The central issue in examining candidate protochordate immune receptors is the nature of the hypothetical primordial Ig/TCR-like receptor. We believe that such a molecule would: (i) possess both a V and a J (joining) region, (ii) be a membrane receptor (such as TCR or the transmembrane stage of the BCR), (iii) bind a ligand, and (iv) have the capacity to transmit a signal across the membrane, possibly by associating with an adaptor molecule or by being able to signal directly. If the molecule were to have the capacity to recognize multiple ligands, it would need to be diverse at the genomic level and/or to have the capacity for somatic diversification, which could have been a relatively early event in the evolution of immunity, developing independently from RAG-mediated rearrangement.

As was the case with jawless vertebrates, discussed above, complex cross-hybridization screens to identify homologs of Ig/TCR, RAG, TdT, MHC I, and MHC II in protochordates have failed, although these studies likely have not been as exhaustive as those carried out in jawless vertebrates. Furthermore, comprehensive analysis of the complete genome of the urochordate *Ciona intestinalis* (a type of sea squirt) has failed to identify such 'adaptive immune' gene homologs, although molecules with Ig domains and various signaling molecules have been found (33). Notably, the well-defined fusibility phenomenon, which is somewhat akin to histocompatibility, in the colonial tunicate *Botryllus schlosseri* does not appear to be MHC-related (34). The purple sea urchin (*Strongylocentrotus purpuratus*), an echinoderm, represents one of the most comprehensively characterized developmental systems and already has been shown to possess molecules associated with innate immunity (35). These animals also possess enormously diversified immune molecules of the scavenger receptor family, which are upregulated after immunization (36, 37). The resolution of the sea urchin genome is nearing completion and already is yielding information that may provide data to reconstruct the primitive deuterostome immune system. The emerging sea urchin genome sequence (Baylor College of

Medicine, Human Genome Sequencing Center) further reveals the enormous diversity of the innate immune gene repertoire in this animal and raises the possibility that some of the selection and expansion capacities of lymphocytes may be of use even in such a divergent receptor system. That is, the complexity of innate immune receptors may reach a point where cellular clonality is necessary to effect specific function. In this case, mechanisms of selection and expansion may have arisen before RAG-mediated recombination evolved. Most strikingly, a family of RAG1-like sequences is present in the sea urchin genome in multiple copies that appear to be the remnants of transposable elements. These repetitive sequences display about 30% amino acid identity to the human RAG1 core region. The elements are diversified and some are pseudogenes with internal stop codons. These loci may provide much needed new data to explore the mechanistic origins of immune gene rearrangement (J. Rast, unpublished observations).

The special case of the variable region-containing chitin-binding proteins

Among the protochordates, our efforts have focused on the amphioxus (*Branchiostoma floridae*), a cephalochordate that represents the most phylogenetically proximal invertebrate form on a direct line with the vertebrates. Early efforts to identify adaptive immune genes in this species by cross-hybridization were not successful; however, the Amptrap approach (see above) has identified a multigene family that encodes the variable region-containing chitin-binding proteins (VCBPs). These molecules possess two tandem V region domains as well as a chitin-binding domain (CBD) and can be classified into five major families with varying numbers of members (31). Comparisons of pooled mRNAs (cDNAs) as well as genomic sequences derived from individual animals have revealed several regions of considerable sequence substitution, i.e. VCBPs are diversified at both the inter- and intrafamily levels. One particular N-terminal region seems to represent the most variable portion of the molecule, and recent studies in our laboratory have shown that this variation occurs at the genomic level in individual animals. Comparisons of genomic DNA and cDNA derived from single animals have not indicated differences beyond an occasional single-nucleotide difference that could be attributable to PCR wobble during cloning. There is no evidence that the VCBPs differentiate somatically; rather, they appear to sustain an extraordinarily high degree of individual polymorphism, leading us to speculate that these genes are under enormous selection, a characteristic of immune receptors. In this case, selection would be at a population level and not the result of an individual somatic process.

Are the VCBPs representative of the 'primordial' Ig-like receptor? Based on the predicted features of the primordial receptor described above, it is unlikely, as VCBPs are not membrane-bound and do not possess a contiguous VJ sequence. However, regarding this latter point, we have resolved the genomic organization of one VCBP, and it appears as if a J could arise through a genomic rearrangement event involving the V1 and V2 regions. We presently have not identified the exact cell types that express VCBPs, but we recognize that VCBP mRNAs are expressed abundantly in the gut, a site of potential pathogen infiltration in this colloidal filter feeder.

Many theories can explain the role of chitin binding by VCBPs. One possibility is that the CBD of a VCBP interacts with chitin degradation products in ingested seawater; such an interaction in effect could multimerize VCBPs as V region-based immune receptors. Other equally attractive possibilities include interactions between the CBD and chitin on the surfaces of pathogens, with the V region itself recognizing a phagocytic or other cell type involved in immune recognition. Alternatively, VCBPs could be mediators of self-nonself recognition. The association of two V regions with a C-terminal region possessing a different biochemical function is broadly reminiscent of the relationship of Ig V regions (Fab) to an Fc, which effects functions that are only indirectly associated with V recognition (*Fig.* 3). Progress is now being made towards the

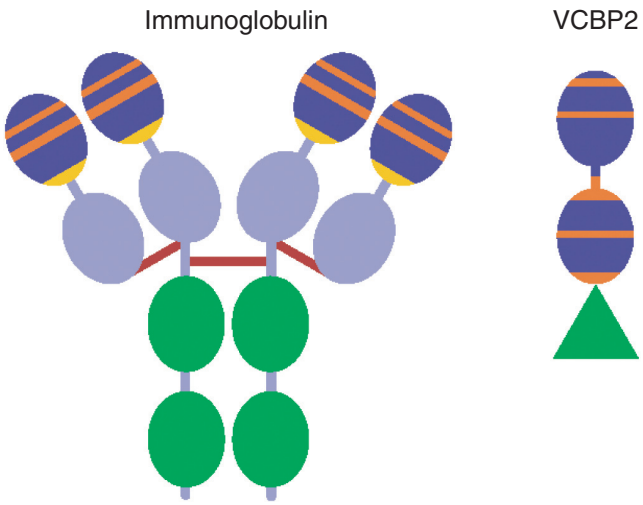

Fig. 3. Regionalization of biological function in immune molecules. The two variable (V)-type domains of an amphioxus variable region-containing chitin-binding protein are joined to a chitin-binding domain (right), whereas the two V regions that form the antigen-combining sites of an immunoglobulin are contiguous with the Fc, which serves as a ligand for other endogenous receptors and does not mediate conventional antigen binding (left). The horizontal lines in the V domain represent regions of localized sequence variation in both receptor families.

expression of functional recombinant VCBPs, which will help resolve these possibilities. At this point, we are inclined to believe the following: (i) the VCBPs play a role in immunity, (ii) the extensive regionalized sequence diversification is integral to that role, and (iii) VCBPs reflect parallel evolution of a V-type recognition molecule rather than a primordial receptor. V-domain bearing protein (VDB), another amphioxus molecule that might fit the role of a progenitor, was described recently (38). The VDB gene encodes a V-type region (with the J-like sequence FGXG) and a multipass transmembrane domain. The gene appears to be single copy, and no evidence has been presented to suggest that it is diversified at either the genomic or cDNA levels.

Candidate primordial receptors in jawed vertebrates

In regards to meeting the definition of a primordial receptor, it is useful to consider the novel immune-type receptor (NITR) genes in bony fish, which were described first in pufferfish, a tactically advantageous model system for genome studies (39, 40). The predicted NITR proteins possess V ectodomains that exhibit extensive germline variation and often contain J-like sequences. With only a few exceptions, the membrane-bound NITRs contain either transmembrane regions with charged residues, which presumably interact with oppositely charged transmembrane adapter molecules, or cytoplasmic tails that contain immunoreceptor tyrosine-based inhibitory motifs. Recent studies have shown that variation in NITRs is distributed in a manner that resembles CDR diversity in Igs and TCRs and that their predicted folding is remarkably similar to that of Igs and TCRs (D. Ostrov, J. Hernandez, and G. Litman, unpublished observation). NITRs do not undergo somatic rearrangement, a property of BCRs and TCRs. Their distribution seems limited to bony fish, although this conclusion cannot be considered definitive until the complete genomic sequences of other critical vertebrate groups are resolved. We have suggested that NITRs are functionally analogous but structurally unrelated to killer cell Ig-like receptor-type molecules in mammals and potentially could be involved in natural killer function. We also have identified another large family of immune-type receptors (ITRs) in the clearnose skate. ITRs have many structural properties in common with NITRs but do not encode V ectodomains. Preliminary data indicate that the ITR gene family may exceed the NITRs in overall complexity.

Recently, we have noted that the signal regulatory proteins (SIRPs), which are expressed on several different leukocyte lineages, contain not only VJ domains but also C1-type domains that otherwise are unique to BCR/TCR and MHCI/II (41). This latter feature distinguishes them from other non-rearranging, VJ-containing candidates that include the NITRs as well as CD8β, CD79b, CD7, NKp30, and ChT1 (CTX in *Xenopus*) (42, 43), a group of putative receptors/adapters containing single V domains. SIRPs may be involved in recognizing self ligands; SIRPα1 has been shown to bind CD47, an Ig superfamily, multipass transmembrane protein expressed widely by leukocytes (44). In addition to the molecules mentioned above and various other Ig-related receptors, such as cell adhesion molecules, broader molecular classes of innate receptors also are useful to consider in the emergence of adaptive immunity. Specifically, an increasingly extensive continuum of molecules that effect both innate and adaptive function is being defined (*Fig.* 4). As our information base increases for these types of molecules, rational strategies for addressing homologous forms in jawless vertebrates and protochordates should become more apparent.

Conclusions

Tracing the evolution of adaptive immunity, as defined by the somatic rearrangement and modification of germline genes encoding immune receptors, remains a most challenging task. In reviewing all of the studies carried out to date on non-mammalian models, there is no question that mechanisms of diversification of the BCR have changed dramatically throughout the course of evolution, whereas the mechanisms of variation within TCRs likely have remained stable. There is little reason to believe that any of the processes that modify the germline information encoded in V regions of BCRs or TCRs in mammals are appreciably different from those in the most phylogenetically distant of the cartilaginous fish (phylogenetically, the most distant extant jawed vertebrates relative to mammals). Convincing evidence has been presented to explain how the germline was modified to create junctional rearrangement as a means for expanding inherited genetic complexity, an obligatory requirement for a system dedicated to recognition of highly variable determinants. Speculation abounds as to the identity of the initial target of the proposed RAG 'transposition' event that ushered in V(D)J recombination, and many attractive candidates that reflect some of the central features of the target can be identified, even though these molecules appear to function in very different capacities today.

The innate immune response has a far more ancient origin than adaptive immunity, to which it is linked functionally in modern vertebrates (45–47). The primary advantage of innate function is to provide immediate recognition of pathogenic

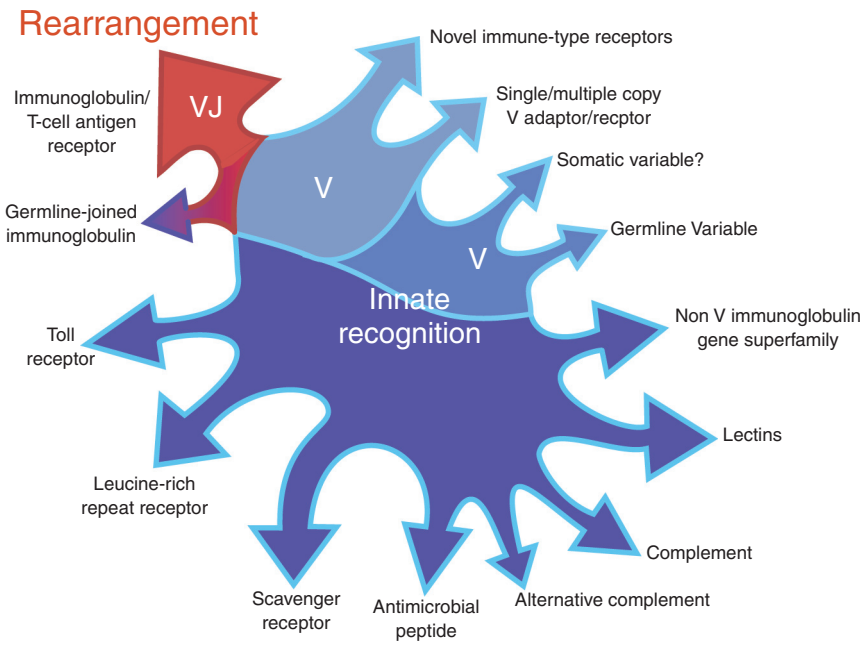

***Fig. 4.* Diversification and specialization by many different families of immune molecules in animals can be seen as a continuum.** Simultaneous evolution of many different receptor families, each becoming specialized for particular immune functions in an organism and synergizing functionally with other families, results in a complex network of defense against pathogens. The area of each region is not meant to be proportional to its complexity or relative impact on host defense. Blue regions of the chart indicate innate immune molecules; differential blue shading in the top portion indicates different phases of evolution of the V-type innate receptor family. The red region indicates molecules whose diversity is generated by somatic recombination. It is plausible that adaptive immunity arose from an innate receptor precursor. In this sense, rearranging antigen receptors can be seen as extensions of the continuum after the addition of RAG-mediated recombination. It is likely that joined genes of cartilaginous fish represent a reverse transition from adaptive to innate function. Notably, extensive diversification is seen in at least one family of innate receptor. The possibility exists that other families of receptors, both immunoglobulin-related and unrelated, can be somatically modified, further blurring the traditional distinction between innate and adaptive recognition.

invasion and an important temporal buffer for the development of an adaptive immune response, which depends on essentially random receptor diversification and subsequent expansion of cells with relevant receptors from extremely small selected pools. Innate receptors could have been the target molecules, or at least share features with those targets, to which RAG-mediated variation was introduced during evolution of adaptive immunity. Some putative innate receptors indeed exhibit considerable germline variation. Although the source of this variation is not known, it is consistent with binding of diverse sets of ligands and introduces the concept of 'localized hypervariability' in innate function. If that is the case, perhaps such receptors also could have acquired a capacity to vary somatically and even to affect the variation so that innate recognition functions at a single cell level. The adaptive immune system as we know it may have succeeded these parallel developments, which may still exist today in modern jawless vertebrates, protochordates, and invertebrates.

References

1. Agrawal A, Eastman QM, Schatz DG. Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. Nature 1998;**394**: 744–751.
2. Hiom K, Melek M, Gellert M. DNA transposition by the RAG1 and RAG2 proteins: a possible source of oncogenic translocations. Cell 1998;**94**:463–470.
3. Litman GW, Anderson MK, Rast JP. Evolution of antigen binding receptors. Annu Rev Immunol 1999;**17**:109–147.
4. Flajnik MF. Comparative analyses of immunoglobulin genes: surprises and portents. Nat Rev Immunol 2002;**2**: 688–698.
5. Ghaffari SH, Lobb CJ. Structure and genomic organization of a second class of immunoglobulin light chain genes in the channel catfish. J Immunol 1997;**159**: 250–258.
6. Anderson M, et al. Complete genomic sequence and patterns of transcription of a member of an unusual family of closely related, chromosomally dispersed immunoglobulin gene clusters in Raja. Int Immunol 1994;**6**:1661–1670.
7. Rast JP, Amemiya CT, Litman RT, Strong SJ, Litman GW. Distinct patterns of IgH structure and organization in a divergent lineage of chrondrichthyan fishes. Immunogenetics 1998;**47**:234–245.

8. Lee SS, Fitch D, Flajnik MF, Hsu E. Rearrangement of immunoglobulin genes in shark germ cells. J Exp Med 1999;**191**: 1637–1648.
9. Miracle AL, et al. Complex expression patterns of lymphocyte-specific genes during the development of cartilaginous fish implicate unique lymphoid tissues in generating an immune repertoire. Int Immunol 2001;**13**:567–580.
10. Hohman VS, Schluter SF, Marchalonis JJ. Diversity of Ig light chain clusters in the sandbar shark (*Carcharhinus plumbeus*). J Immunol 1995;**155**:3922–3928.
11. Anderson MK, Shamblott MJ, Litman RT, Litman GW. Generation of immunoglobulin light chain gene diversity in *Raja erinacea* is not associated with somatic rearrangement, an exception to a central paradigm of B cell immunity. J Exp Med 1995;**182**:109–119.
12. Lee SS, Tranchina D, Ohta Y, Flajnik MF, Hsu E. Hypermutation in shark immunoglobulin light chain genes results in contiguous substitutions. Immunity 2002;**16**:571–582.
13. Hinds-Frey KR, Nishikata H, Litman RT, Litman GW. Somatic variation precedes extensive diversification of germline sequences and combinatorial joining in the evolution of immunoglobulin heavy chain diversity. J Exp Med 1993;**178**:825–834.
14. Rast JP, Anderson MK, Strong SJ, Luer C, Litman RT, Litman GW. α, β, γ, and δ T cell antigen receptor genes arose early in vertebrate phylogeny. Immunity 1997;**6**: 1–11.
15. Greenberg AS, Avila D, Hughes M, Hughes A, McKinney EC, Flajnik MF. A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 1995;**374**:168–173.
16. Schluter SF, Marchalonis JJ. Cloning of shark RAG2 and characterization of the RAG1/ RAG2 gene locus. FASEB J 2003;**17**:470–472.
17. Hughes AL. Genomic catastrophism and the origin of vertebrate immunity. Arch Immunol Ther Exp 1999;**47**:347–353.
18. Aravind L, Koonin EV. Gleaning non-trivial structural, functional and evolutionary information about proteins by iterative database searches. J Mol Biol 1999;**287**:1023–1040.
19. Bartl S, et al. Terminal deoxynucleotidyl transferases from elasmobranchs reveal structural conservation within vertebrates. Immunogenetics 2003;**55**:594–604.
20. Jessen JR, Jessen TN, Vogel SS, Lin S. Concurrent expression of recombination activating genes 1 and 2 in zebrafish olfactory sensory neurons. Genesis 2001;**29**:156–162.
21. Diaz M, Velez J, Singh M, Cerny J, Flajnik MF. Mutational pattern of the nurse shark antigen receptor gene (NAR) is similar to that of mammalian Ig genes and to spontaneous mutations in evolution: the translesion synthesis model of somatic hypermutation. Int Immunol 1999;**11**:825–833.
22. Wilson M, Bengten E, Miller NW, Clem LW, Du Pasquier L, Warr GW. A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD. Proc Natl Acad Sci USA 1997;**94**:4593–4597.
23. Ota T, Rast JP, Litman GW, Amemiya CT. Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox. Proc Natl Acad Sci USA 2003;**100**: 2501–2506.
24. Mussmann R, Courtet M, Schwager J, Du Pasquier L. Microsites for immunoglobulin switch recombination breakpoints from Xenopus to mammals. Eur J Immunol 1997;**27**:2610–2619.
25. Fugmann SD, Schatz DG. RNA AIDs DNA. Nat Immunol 2003;**4**:429–430.
26. Barreto V, Reina-San-Martin B, Ramiro AR, McBride KM, Nussenzweig MC. C-terminal deletion of AID uncouples class switch recombination from somatic hypermutation and gene conversion. Mol Cell 2003;**12**: 501–508.
27. Ta V-T, et al. AID mutant analyses indicate requirement for class-switch-specific cofactors. Nat Immunol 2003;**4**:843–848.
28. Saunders HL, Magor BG. Cloning and expression of the AID gene in the channel catfish. Dev Comp Immunol 2004;**28**: 657–663.
29. Mayer WE, Uinuk-Ool T, Tichy H, Gartland LA, Klein J, Cooper MD. Isolation and characterization of lymphocyte-like cells from a lamprey. Proc Natl Acad Sci USA 2002;**99**:14350–14355.
30. Uinuk-Ool T, Mayer WE, Sato A, Dongak R, Cooper MD, Klein J. Lamprey lymphocyte-like cells express homologs of genes involved in immunologically relevant activities of mammalian lymphocytes. Proc Natl Acad Sci USA 2002;**99**:14356–14361.
31. Cannon JP, Haire RN, Litman GW. Identification of diversified immunoglobulin-like variable region-containing genes in a protochordate. Nat Immunol 2002;**3**:1200–1207.
32. Burrows PD, Stephan RP, Wang YH, Lassoued K, Zhang Z, Cooper MD. The transient expression of pre-B cell receptors governs B cell development. Semin Immunol 2002;**14**:343–349.
33. Azumi K, et al. Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: ''waiting for Godot''. Immunogenetics 2003;**55**:570–581.
34. De Tomaso A, Weissman IL. Initial characterization of a protochordate histocompatibility locus. Immunogenetics 2003;**55**:480–490.
35. Smith LC, Chang L, Britten RJ, Davidson EH. Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags. J Immunol 1996;**156**: 593–602.
36. Pancer Z, Rast JP, Davidson EH. Origins of immunity: transcription factors and homologues of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes. Immunogenetics 1999;**49**:773–786.
37. Pancer Z. Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. Proc Natl Acad Sci USA 2000;**97**:13156–13161.
38. Sato A, Mayer WE, Klein J. A molecule bearing an immunoglobulin-like V region of the CTX subfamily in amphioxus. Immunogenetics 2003;**55**:423–427.
39. Strong SJ, et al. A novel multigene family encodes diversified variable regions. Proc Natl Acad Sci USA 1999;**96**:15080–15085.
40. Litman GW, Hawke NA, Yoder JA. Novel immune-type receptor genes. Immunol Rev 2001;**181**:250–259.
41. van den Berg TK, Yoder JA, Litman GW. On the origins of adaptive immunity: innate immune receptors joining the tale. Trends Immunol 2004;**25**:11–16.
42. Katevuo K, et al. ChT1, an Ig superfamily molecule required for T cell differentiation. J Immunol 1999;**162**:5685–5694.
43. DuPasquier L, Chretien I. CTX, a new lymphocyte receptor in Xenopus, and the early evolution of Ig domains. Res Immunol 1996;**147**:218–226.
44. Seiffert M, et al. Human signal-regulatory protein is expressed on normal, but not on subsets of leukemic myeloid cells and mediates cellular adhesion involving its counterreceptor CD47. Blood 1999;**94**:3633–3643.
45. Medzhitov R, Janeway CA Jr. Decoding the patterns of self and nonself by the innate immune system. Science 2002;**296**:298–300.
46. Medzhitov R. Toll-like receptors and innate immunity. Nat Rev 2001;**1**:135–145.
47. Hoffmann JA, Reichhart J-M. Drosophila innate immunity: an evolutionary perspective. Nat Immunol 2002;**3**:121–126.