Commentary

50 million years of chordate evolution: Seeking the origins of adaptive immunity

Diana J. Laird*†, Anthony W. De Tomaso*, Max D. Cooper‡, and Irving L. Weissman*

*Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305-5324; and ‡Howard Hughes Medical Institute, University of Alabama, 378 Wallace Tumor Institute, Birmingham, AL 35294

Agnathans, the most primitive chordates, are poised at a fascinating point in evolution. In the 50 million years between agnathan and chondrichthian divergence, something mysterious, even miraculous occurred: the adaptive immune system evolved. A complex, interdependent system of checks and balances in which fragments of both intracellular and extracellularly derived foreign molecules (antigens) are presented in the clutch of MHC cell surface molecules to 10^{16} possible different lymphocyte receptors arose seemingly overnight in evolutionary time. Moreover, the system that evolved before the divergence of jawed fish was so successful that the basic paradigm remained throughout the radiation of subsequent vertebrate lineages.

The immune mechanisms shared by all vertebrates are collectively termed "adaptive immunity," "combinatorial immunity," or the "anticipatory response" (1) for the unique receptors generated in response to foreign invaders, or antigens, such as microbes, parasites, and genetically altered cells. Unlike the innate immune components found in all multicellular organisms, this response is specific, selective, remembered, and regulated. T and B lymphocytes, which bear unique immunoreceptors, circulate throughout the body in search of antigens. B cell receptors, called immunoglobulins (Ig), recognize intact macromolecular complexes on the invaders; the cell progeny of B cells secrete massive amounts of the specific Igs, known as antibodies, that permeate the intercellular milieu. By contrast, T cell receptors (TCRs) recognize small fragments of antigen presented in the groove of the MHC (major histocompatibility complex) receptor found on nearly every cell in the body. TCR and Ig are encoded by interspersed V, D, and J subgeneic elements that are combinatorially rearranged during lymphocyte development; rearrangement is initiated by a unique DNA splicing enzyme complex, the recombinase activating genes (RAGs). Each lymphocyte expresses about 10^5 identical products of one successful V(D)J rearrangement pair, and collectively the antigen receptors of T and B lymphocytes anticipate the full repertoire of antigenic shapes the universe supplies (2). We would expect that living agnathans, hagfish and lamprey, might retain the genetic raw materials for these inventions, but no conclusive evidence of MHC, Ig, TCR, or RAG elements exists in metazoans that diverged before chondrichthyes.

Luminaries in the field of evolution and immunity describe the critical events that befell our last common ancestor with sharks between 500 and 450 million years ago in cosmic terms. Synthesizing many recent findings, Schluter *et al.* (3) describe the origin of the combinatorial immune system with the moniker "Immunological Big Bang." As in astrophysics, the study of immunology must infer the events of the past from the stuff—whether electromagnetic or genetic—that remains today. This endeavor is complicated by the possibility that genetic mechanisms and selective forces exerted by parasites differed in the era of early vertebrates—as did the physics of the nascent universe. Although molecular evolutionists rely on the genomes of extant organisms to peer back in time, this is commonly misunderstood to be like gazing at the light from stars millions of light years away. The genome of living sharks is neither ancestral to nor older than that of humans, but their comparison may reveal structures that existed in their last common ancestor 450 million years ago.

With the unavailability of either fossilized Cambrian nucleic acid or time travel, evolutionary immunologists primarily rely on comparative tools using extant species. Recent investigations into the phylogenetic origin of this complex system derive from four different approaches: (*i*) the study of the evolution of rearranging receptors, (*ii*) the study of the origin and assembly of MHC, (*iii*) the study of the origins of self-nonself recognition systems such as histocompatibility, and (*iv*) the study of lymphocyte phylogeny. A paper in this issue of PNAS by Shintani *et al.* (4) advances our understanding of the cellular mediators of adaptive immunity by tracing the origins of the *Spi* family of lymphocyte-specific transcription factors in the jawless fish. Here we attempt to situate this finding from the laboratory of Jan Klein within the field of evolutionary immunology.

Rearranging Antigen Receptors. Much attention has focused on the epicenter of rearrangement, RAG1/2, the lymphocyte-specific proteins that create nicks between germline chromosomal V, D, and J components of Igs and TCR. The curious genomic proximity of RAG1 and -2, their absence of introns (5), and their detection in sharks but not protochordates first inspired the hypothesis of their origin by horizontal transfer to a vertebrate ancestor (6). Similarity between the mechanisms of DNA cleavage for RAG recombination, retroviral integration, and transposition corroborated the hypothesis that RAG1 and RAG2 entered together on a retrotransposon (7–11). It is thought that a retrotransposon containing RAG1 and -2 integrated into a large cis promoter element. Once translated, RAG proteins mediated transposition of a gene segment flanked by short recombination signal sequences (RSS) into a primordial receptor gene, first splitting the gene into fragments that became V, D, and J through duplication (12). Whether RAG fortuitously planted itself into a uniquely lymphoid promoter or acquired specificity later is unclear.

Non-Rearranging Ig-Like Receptors. Ig and TCR molecules feature a structural motif comprised of β-pleated sheets, known as the "Ig fold," whose archtypical presence in receptors and adhesion molecules through phylum Chordata and even in invertebrates (13, 14) dates it before the entrance of RAG. Many investigations have approached the hypothesized primordial receptor by searching for relics of Ig-type molecules in classes of organisms that diverged before agnathans. Sequences isolated from hagfish (15, 16), tunicate (17), and sponge (14, 18) aligned with canonical Ig domains produce a similarity score that falls into the "twilight zone" of questionable

See companion article on page 7417.

†To whom reprint requests should be addressed. E-mail: dlaird@leland.stanford.edu.

homology (19). Another line of investigation begins with the analysis of vertebrate non-rearranging Ig-like receptors in hopes of gleaning information about the structure of the ancestral pre-RAG receptor. One such group of related receptors includes the paired Ig-like receptors (PIR) in mice (20, 21), the Ig-like transcripts (ILT) in humans [also called leukocyte inhibitory receptors (LIR) or monocyte inhibitory receptors (MIR) (22)], and the killer inhibitory receptors (KIR) of humans (23). The PIR, ILT, and KIR multigene families are located in syntenic regions of the mouse and human genomes. Members of these receptor families vary slightly in their extracellular Ig domain, and they come in two functionally distinct forms. They either possess intracellular tyrosine-based inhibitory motifs (ITIM) to serve as inhibitory receptors, or they have transmembrane region polarity allowing them to associate with another transmembrane chain containing a tyrosine-based activation motif (ITAM) to form a cell activating receptor complex. The ligands so far identified for these pairs of activating and inhibitory receptors include MHC class I and class I-like molecules. The sequence diversity, polymorphic nature, and counterfunctional capability of these receptors make them reasonable candidates for a primitive allorecognition system. Although the KIR family of receptors in humans probably arose as recently as 40 million years ago, the identification of PIR/ILT relatives in birds, the chicken paired Ig-like receptors (CHIR),[§] suggests a much earlier origin for this family of Ig-like activating and inhibitory receptors.

Another large multigene family of novel immune-type receptors (NITR) has been identified in the pufferfish (24). These nonrearranging Ig-like genes encode patterned sequence variability in their V-like domains. Many members of this family possess ITIM sequences in their cytoplasmic region whereas others may have polar transmembrane sequences that would allow them to associate with ITAM-bearing partners. Apart from shared general features that support their candidacy as primitive allorecognition receptors, the NITR and PIR-types of receptor genes are not obviously related, and data acquired to date suggest it is unlikely they are located in syntenic chromosomal regions (G. W. Litman, personal communication). In the ongoing search for non-rearranging Ig-like allorecognition receptors that may predate the TCR and Ig receptors, it is interesting to note the upregulation of two Ig-like receptor molecules at the interface of sponge autografts (14).

[§]Dennis, G., Kubagawa, H. & Cooper, M. D. (2000) *FASEB J.* **14,** 1020 (abstr.).

MHC. It is difficult to imagine a need for peptide-presenting MHC receptors before the existence of TCRs that simultaneously bind MHC and peptide. To date, MHC receptors have not been isolated in species that lack RAG or TCR. However, each type of MHC receptor resides in its own genetic locus spanning 2,000–3,000 kb and containing many linked immune and some nonimmune genes, including complement and inflammatory genes. Because many of the genes associated with vertebrate MHC loci are phylogenetically conserved and perform functions common to all metazoans, it is feasible to test linkage between homologs of MHC-related genes in nonvertebrate species (2). In the earliest known chordates, the protochordates, close linkage is not observed between two HSP70 genes, which are contained in the MHC region of vertebrates, or to the self-nonself histocompatibility system called Fu/HC (refs. 25 and 26; see below). It remains to be tested whether homologs of other MHC-associated genes such as TNF, LMP, TAP, complement, and collagen lie in genetic proximity in agnathans or protochordates. The present data suggest that the MHC assembled very quickly during the 50 million years of "big bang" in question.

Other Histocompatibility Systems. Although the MHC and adaptive immunity apparently first coexist in the chondrichthians, histocompatibility is a broader phenomenon in multicellular organisms. Natural tissue compatibility systems are well documented in organisms ranging from slime molds, which spend the majority of their lives as single cells and only aggregate to sexually reproduce (27), to sponges, the simplest metazoans (28), and more complex invertebrate phyla (cnidarians) to the prevertebrates, or protochordates (e.g., ascidians). However, with little exception, nothing is known about the molecular mechanisms that underlie allorecognition in phyla that predated vertebrates.

Protochordates, which occupy a strategic phylogenetic position at the cusp of chordate evolution, provide an excellent model to study the possible origins of vertebrate adaptive immunity. Colonial protochordates are endowed with a genetically regulated histocompatibility system, termed Fu/HC, in which a single Mendelian locus with hundreds of codominantly expressed alleles controls fusion or rejection of allogeneic individuals (25). Although early attempts failed to identify the protochordate histocompatibility genes by homology to vertebrate MHC receptors, modern genomics places us on the brink of identifying the key players via positional cloning. The Fu/HC locus in the protochordate *Botryllus schlosseri* has been mapped within a 1-cM region of the genome (29, 2), and the analysis of expressed sequences in this region should lead to identification of these histocompatibility genes.

Lymphocytes. Another approach to the problem of immune evolution is to inquire into the phylogenetic origins of the cellular bearers of recombined receptors. One would like to know the characteristics of lymphocytes in that ancestral recipient of the first RAG transposon; whether separate T and B cell lineages existed before TCR and Ig; and what mechanisms of lymphocyte fate determination were in place. The paper by Shintani *et al.* (4) in this issue makes an important foray into this question with the identification and characterization of the ancestor to a family of lymphocyte transcription factors called Spi. In mammals, three known family members, Spi-1 (also called PU.1), Spi-B, and Spi-C, are present at various points of lymphocyte lineage and interact via an Ets DNA-binding domain to turn on many genes important in lymphocyte development. Jan Klein's group has isolated a single homolog of the Spi genes in the jawless fish, as verified by both amino acid similarity and exon structure. Data not presented in the paper confirm that no other family members with similar Ets domains exist in the hagfish, which lends credence to the hypothesis that this gene is derived from a single common ancestor to all three vertebrate forms.

The present study raises many new questions about the ancestral lymphocytes that became T and B cells through transformation with the RAG genes. Very little is known about the functional properties of prevertebrate lymphocytes. Data from teleost and shark suggest that ancient lymphocytes had a spontaneous or NK-like cytotoxicity against parasites (30), but neither the mechanism of recognition nor the mode of killing are known. Parsimony argues that the ontogeny of lymphocytes is conserved across evolution, but how did the present system of lineage determination evolve? In mouse, one can isolate a pure population of common lymphoid progenitors that use specific cytokine receptors (e.g., IL-7, Flt-3, SLF) to interact with bone marrow stroma (31) and that express distinct signal transduction molecules and transcription factors (TFs) (32). Do similar microenvironments in the agnathan tissues support similar lymphoid progenitors? Furthermore, is the subsequent differentiation of those cells directed by homologous TFs? Detective work using TFs may be complicated. Ikaros, for example, is a family of TFs with many known mammalian isoforms that interact with one another as well as many promoter elements. Although early investigations identified Ikaros as a lymphocyte-specific TF (33), the Ikaros gene has a number of zinc finger encoding domains, and virtually all bloodforming stem and progenitor cells, as well as lymphocytes, express different

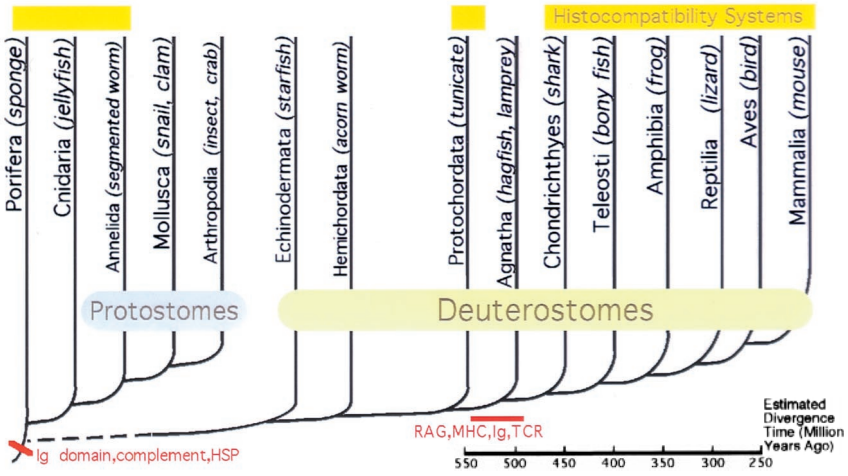

Fig. 1. Dendrogram showing evolutionary relationships between selected animal phyla and classes according to the current model (http://phylogeny.arizona.edu/tree/phylogeny.html). Speculative origin times of adaptive immune structures are indicated, as are documented histocompatibility systems. Divergence times are not to scale unless indicated.

collections of splice-variant isoforms (34). Perhaps an easier route is found by examining the downstream genes regulated by the transcription factor in question. In mammals, Spi activates transcription of several B cell genes CD72, LSP1, CD20, BTK, mb-1, μ HC, κ and λ LCs, and the J chain (35), the last of which may be identifiable in lamprey by hybridization to known homologs in chicken and segmented worm (36, 37). The study of phylogenetically ancient lymphocytes might exploit the conservation of vertebrate cytokines to ask whether candidate lymphocytes respond to proliferative factors from other species.

Perhaps another pathway to discovery of the phylogenetic origins of adaptive immunity will come from a reconsideration of immune system functions critical to the survival and the genomic integrity of each examined species. Although the prevailing paradigm of vertebrate immune and MHC system function is the protection from parasites with distinct genomes (microbes, metazoan parasites, etc.), many complex invertebrate metazoans have an equally pressing danger to their genomic integrity—stem cells from other members of their species (38). Many protochordate and other invertebrate metazoans can fuse their vascular systems with other individuals, opening their bodies to foreign cells, including germline stem cells (39, 40). As might be expected, selection for more predatory germline stem cells occurs, and the only barriers the species have from this spreading, species-homogenizing protoneoplasm is a highly genetically polymorphic histocompatibility barrier, such as Fu/HC (25), that limits genome sharing to siblings (40). Thus, the kind of histocompatibility immune system that allows sibling cell lineage transfer (by a shared Fu/HC allele) but prevents invasion by cells with no common allele might have been the primordial alloimmune state in the protochordate ancestor that bordered the emerging chordate phyla (Fig. 1).

Finally, it should be pointed out that there are serious limits to morphological descriptions of lymphoid cells in both the contexts of ontogeny and phylogeny. Neither microscopic (histological or immunohistochemical) nor molecular (collections of transcription factors, etc) morphologies can substitute for the isolation of the cells in question, the identification of their ontogenetic precursors and progeny, and the direct demonstration between species of their morphological and functional similarity (41).

1. Klein, J. (1990) *Immunology* (Blackwell Scientific, Boston.).
2. Magor, B. G., De Tomaso, A., Rinkevich, B. & Weissman, I. L. (1999) *Immunol. Rev.* **167,** 69–79.
3. Schluter, S. F., Bernstein, R. M., Bernstein, H. & Marchalonis, J. J. (1999) *Dev. Comp. Immunol.* **23,** 107–111.
4. Shintani, S., Terzic, J., Sato, A., Saraga-Babic, M., O'hUigin, C., Tichy, H. & Klein, J. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7417–7422.
5. Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. (1990) *Science* **248,** 1517–1522.
6. Bartl, S., Baltimore, D. & Weissman, I. L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 10769–10070.
7. Schatz, D. G. (1999) *Immunol. Res.* **19,** 169–182.
8. Lewis, S. M. & Wu, G. E. (1997) *Cell* **88,** 159–162.
9. Thompson, C. B. (1995) *Immunity* **3,** 531–539.
10. Hiom, I., Melek, M. & Gellert, M. (1998) *Cell* **94,** 463–470.
11. Agrawal, A., Eastman, Q. M. & Schatz, D. G. (1998) *Nature (London)* **394,** 744–751.
12. Yu, W., Nagaoka, H., Jankovic, M., Misulovin, Z., Suh, H., Rolink, A., Melchers, F., Meffre, E. & Nussenzweig, M. C. (1999) *Nature (London)* **400,** 682–687.
13. Williams, A. F. & Barclay, A. N. (1988) *Annu. Rev. Immunol.* **6,** 381–405.
14. Blumbach, B., Diehl-Seifert, B., Seack, J., Steffen, R., Muller, I. M. & Muller, W. E. G. (1999) *Immunogenetics* **49,** 751–776.
15. Varner, J., Neame, P., Litman, G. W. (1991) *Proc. Natl. Acad. Sci. USA.* **88,** 1746–1750.
16. Raison, R. L. & Hildemann, W. H. (1984) *Dev. Comp. Immunol.* **8,** 99–108.
17. Pancer, Z., Diehl-Seifert, B., Rinkevich, B. & Muller, W. E. (1997) *DNA Cell Biol.* **16,** 801–806.
18. Pancer, Z., Skorokhod, A., Blumbach, B. & Muller W. E. (1998) *Gene* **207,** 227–233.
19. Doolittle, R. F. (1987) *URFs and ORFs: A Primer on How to Analyze Derived Amino Acid Sequences* (University Science Books, Mill Valley, CA).
20. Kubagawa, H., Burrows, P. D. & Cooper, M. D. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 5261–5266.
21. Hayami, K., Fukuta, D., Nishikawa, Y., Yamashita, Y., Inui, M., Ohyama, Y., Hikida, M., Ohmori, H. & Takai, T. (1997) *J. Biol. Chem.* **272,** 7320–7327.
22. Colonna, M., Nakajima, H. & Cella, M. (2000) *Semin. Immunol.* **12,** 121–127.
23. Long, E. O. (1999) *Annu. Rev. Immunol.* **17,** 875–904.
24. Strong, S. J., Mueller, M. G., Litman, R. T., Hawke, N. A., Haire, R. N., Miracle, A. L., Rast, J. P., Amemiya, C. T. & Litman, G. W. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 15080–15085.
25. Scofield, V., Schlumpberger, J. M., West, L. A. & Weissman, I. L. (1982) *Nature (London)* **295,** 499–502.
26. Fagan, M. B. & Weissman, I. L. (1998) *Immunogenetics* **47,** 468–476.
27. Buss, L. W. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 5337–5341.
28. Humphreys, T. & Reinherz, E. L. (1994) *Immunol. Today* **15,** 316–320.
29. De Tomaso, A. W., Saito, Y., Ishizuka, K. J., Palmeri, K. J. & Weissman, I. L. (1998) *Genetics* **149,** 277–287.
30. Haynes, L. & McKinney, E. C. (1991) *Dev. Comp Immunol.* **15,** 123–134.
31. Kondo, M., Weissman, I. L. & Akashi, K. (1997) *Cell* **91,** 661–672.
32. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. (2000) *Nature (London)* **404,** 193–197.
33. Georgopoulos, K., Bigby, M., Wang, J. H., Molnar, A., Wu, P., Winandy, S. & Sharpe, A. (1994) *Cell* **79,** 143–156.
34. Klug, C. A., Morrison, S. J., Masek, M., Hahm, K., Smale, S. T. & Weissman, I. L. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 657–662.
35. Rao, S., Garrett-Sinha, L. A., Yoon, J. & Simon, M. C. (1999) *J. Biol. Chem.* **274,** 34245–34252.
36. Takahashi, T., Iwase, T., Takenouchi, N., Saito, M., Kobayashi, K., Moldoveanu, Z., Mestecky, J. & Moro, I. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 1886–1891.
37. Takahashi, T., Takashi, I., Tachibana, T., Komiyama, K., Kobayashi, K., Chen, C., Mestecky, J. & Moro, I. (2000) *Immunogenetics* **51,** 85–91.
38. Buss, L. W. (1987) *The Evolution of Individuality* (Princeton Univ. Press, Princeton).
39. Stoner, D. S. & Weissman, I. L. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 15254–15259.
40. Stoner, D. S., Rinkevich, B. & Weissman, I. L. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 9148–9153.
41. Weissman, I. L. (1994) *Immunity* **1,** 529–531.