Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 10769–10770, November 1994

Commentary

Molecular evolution of the vertebrate immune system

Simona Bartl, David Baltimore†, and Irving L. Weissman‡*

**Department of Biological Sciences, University of North Carolina, Wilmington, NC 28403; †Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139; and ‡Departments of Pathology and Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305*

An understanding of the evolution of vertebrate immunity is slowly emerging from studies of chordates that share distant ancestors with mammals. In higher vertebrates, such as birds and mammals, we know that two receptor systems are operative. B cells use immunoglobulins to bind foreign agents (the functionally defined antigens). T cells use T-cell receptors (TCRs) to respond to antigen in the form of processed peptides bound to cell surface proteins encoded in the major histocompatibility complex (MHC). Thus, for T cells, two receptor molecules are required for recognition of antigen. First, the MHC molecule on the infected cell binds the processed antigenic peptide; second, the TCR binds the MHC molecule–antigenic peptide complex (for a model, see Fig. 1).

MHC molecules were first described as transplantation antigens. It was the non-self MHC molecule expressed on transplanted tissues that triggered a vigorous T-cell-mediated immune response. In retrospect, we know that most tissue transplants among individuals of the same species result in graft rejection due to the high degree of polymorphic MHC alleles within wild populations. The role of self MHC molecules in the presentation of antigen to T cells was a later discovery. Thus, MHC molecules function in two not necessarily separate processes: recognition of allogeneic tissues and presentation of antigen.

It seems plausible that TCRs and MHC molecules evolved together early in the vertebrate lineage. If so, one would expect to find allorecognition and antigen presentation in most vertebrates. On a molecular level, MHC molecules and TCRs should be present in the same organisms. In addition, if TCRs exist in lower vertebrates, then the genes and proteins required for gene rearrangement also must exist.

Both immunoglobulins and TCRs require the rearrangement of subgenic elements composed of two [V (variable), J (joining)] or three [V, D (diversity), J] large arrays, with 5–1000 members in each of the arrays, to form the V regions of single chains of the immunoglobulin or TCR antigen receptors (for a model, see Fig. 2). The combinatorial association of the subgenic elements [with a single constant (C) region subgenic element], and also the combinatorial assembly of two-chain receptor molecules [IgH (heavy) and IgL (light) for immunoglobulins, TCR $\alpha\beta$ or $\gamma\delta$ for TCRs] allows the generation of receptor repertoires that can specifically bind a large variety of potential antigens. Without rearrangement, it is difficult to envision the production of an effective, highly specific immune response.

For any organism to have an immune system akin to that seen in mammals, the minimally required molecules are the antigen receptors (immunoglobulin and TCR), the antigen presentation molecules (MHC), and the gene rearranging proteins. The vertebrate taxon that shares the ancestor most distantly related to mammals in which this minimal prerequisite has been found is the cartilaginous fish. In the 1980s, genes that encode immunoglobulins were first isolated from the horned shark (1). Like all other immunoglobulin genes, the shark homologues undergo rearrangement. Interestingly, the organization of gene segments within the shark genome differed markedly from that of mammals (for a model, see Fig. 3). While the subgenic elements that comprise the functional, rearranged immunoglobulin gene are called V, D, J, and C (though some loci lack D segments), shark immunoglobulin loci consist of tandem arrays of V, D, J, and C segment clusters, resulting in the number of copies of each segment in the genome being equal (2). In contrast, the mammalian genomes contain a pool of V regions found upstream of several D and then several J segments with usually a single 3′ C region (Fig. 2). The mechanism and driving force behind such a major reorganization of these gene loci during vertebrate evolution remains perplexing.

More recently, evidence for all three types of MHC loci that are found in higher vertebrates (class I, class IIA, and class IIB) has been documented in the shark (3–5). The shark class II genes appear to encode a heterodimeric protein that is structurally similar to its mammalian counterpart and includes a highly polymorphic cleft for the presentation of antigenic peptides. The presence of what appear to be functional MHC molecules would argue in favor of the existence of T cells with functional TCRs in the shark. The evidence of a thymus in several shark species also argues in favor of T cells in this organism (refs. 6 and 7; Carl A. Luer, personal communication). The thymus is the site of T-cell maturation in mammals, birds, and amphibians. Now, the isolation of TCR homologues in sharks provides the most compelling evidence to date for the presence of T cells in cartilaginous fish.

In a recent issue of this journal, Rast and Litman (8) report the isolation of a cDNA clone that appears to encode a functional, rearranged TCR. They also isolated several V regions that display significant sequence diversity. By analyzing genomic clones, they found that the germ-line genes are unrearranged and appear to be organized in multiple clusters in the same manner as shark immunoglobulin genes. In mammals, both TCR and immunoglobulin genes are arranged similarly, as described above.

The isolation of shark TCR homologues emphasizes the importance of several questions. (*i*) When and how did the two receptor systems (immunoglobulin and TCR) undergo the massive chromosomal reorganization that resulted in the genomic structure observed in mammals today? It is intriguing that the TCRγ, TCRβ, and Igλ loci in mammals display a gene organization that can be considered

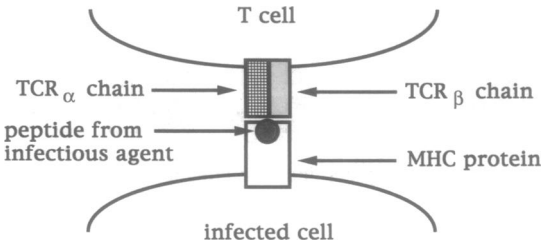

FIG. 1. Model of an infected cell with a MHC protein presenting a foreign peptide to a T cell.

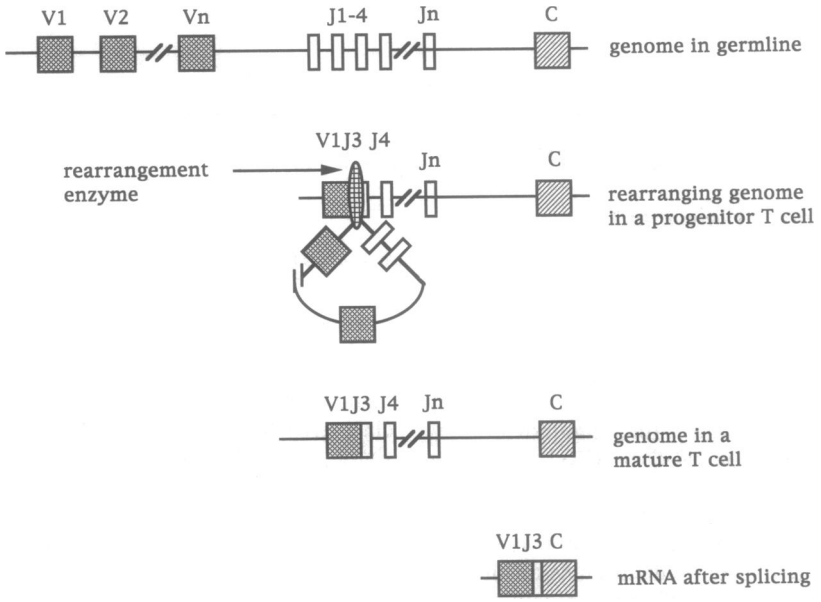

FIG. 2. Model of a mammalian TCR gene locus that rearranges to form a functional gene.

somewhat intermediate between the other mammalian immunoglobulin and TCR loci and those found in the shark. Isolation of TCRγ, TCRβ, and the Igλ loci from other vertebrates may provide insight into possible chromosomal reorganization schemes. (*ii*) Which shark cells express these TCRs and are they a distinct population from immunoglobulin-positive, presumptive B cells? This question should be easily addressed after the development of shark TCR-specific antibodies. (*iii*) Do these TCRs function in the recognition of MHC molecule–antigenic peptide complexes? This may be the most difficult question to answer, since it will probably require *in vitro* stimulation and maintenance of shark blood cells, but it is key to the understanding of shark T-cell function.

The third set of molecules that are minimally required for an immune system similar to the mammalian system are the gene rearranging proteins. At least two of these proteins are encoded by two linked genes called *RAG-1* and *RAG-2* in mammals, birds, and amphibians (9–13). In all the organisms for which detailed genomic mapping has been reported, the *RAG* genes are almost intronless, having only a leader sequence separated by an intron. We (S.B. and I.L.W.) and others (14) (J. J. Marcholonis, personal communication) have now found evidence for the presence of *RAG-1* homologues in sharks. Thus, sharks appear to possess at least some representatives of the three classes of minimally required immunological molecules.

The question remains which of these molecules evolved first. It is possible to envision a function for MHC molecules in the absence of TCRs. MHC molecules may have served as peptide transporters or as allogeneic markers for recognition by non-TCR receptors; for example, the somewhat mysterious MHC class I molecule recognizing receptors on natural killer (NK) cells (15). It is conceivable that TCRs at some point in evolution recognized antigen in the absence of MHC molecules. The most difficult scenario to consider would be the presence of receptor systems requiring rearrangement without the presence of RAG homologues. Immunoglobulin and TCR genes both require RAG proteins for rearrangement. On the other hand, RAG proteins require specific recombination signals to rearrange immunoglobulin and TCR genes. Although this appears to be a "chicken or egg" dilemma, an intriguing possibility is that ancestral *RAG* homologues were the trigger for cooption and/or evolution of the other immunologically important molecules in the chordate lineage.

RAG genes in mammals, birds, and amphibians have intronless coding regions and are very close to each other, both rare situations in vertebrate genomes. They have the aspect of a unit brought into the genome from a single-celled organism like a yeast or bacterium, where intronless, closely apposed genes are common (10). They could have been part of retrotransposons and had a DNA rearrangement function in their previous life. It is possible that the ancestors of *RAG* genes may have been horizontally transferred into a metazoan lineage at some relatively recent point in evolution. The newly introduced *RAG* genes may have acted, most likely with other proteins, on preexisting recombination signals (which consists of conserved heptamer and nonamer sequences) that may have been present for some other function or by random chance. In that view, the signal sequences captured the ancestors of present day TCR and immunoglobulin gene segments. Such a scenario is highly speculative but if true would imply a startling role for horizontal information transfer as the pivotal event in the evolution of vertebrate immunity.

FIG. 3. Model of a germ-line immunoglobulin gene locus in the shark.

1. Litman, G. W., Berger, L., Murphy, K., Litman, R., Hinds, K. & Erickson, B. W. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2082–2086.
2. Hinds, K. R. & Litman, G. W. (1986) *Nature (London)* **320**, 546–549.
3. Hashimoto, K., Nakanishi, T. & Kurosawa, Y. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2209–2212.
4. Kasahara, M., Vazquez, M., Sato, K., McKinney, E. C. & Flajnik, M. F. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 6688–6692.
5. Bartl, S. & Weissman, I. L. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 262–266.
6. Pulsford, A., Morrow, W. J. W. & Fänge, R. (1984) *J. Fish Biol.* **25**, 353–360.
7. Lloyd-Evans, P. (1993) *Dev. Comp. Immunol.* **17**, 501–524.
8. Rast, J. P. & Litman, G. W. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 9248–9252.
9. Schatz, D. G., Oettinger, M. A. & Baltimore, D. (1989) *Cell* **59**, 1035–1048.
10. Oettinger, M. A., Schatz, D. G., Gorka, C. & Baltimore, D. (1990) *Science* **248**, 1517–1522.
11. Ichihara, Y., Hirai, M. & Kurosawa, Y. (1992) *Immunol. Lett.* **33**, 277–284.
12. Carlson, L. M., Oettinger, M. A., Schatz, D. G., Masteller, E. L., Hurley, E. A., McCormack, W. T., Baltimore, D. & Thompson, C. B. (1991) *Cell* **64**, 201–208.
13. Greenhalgh, P., Olesen, C. E. M. & Steiner, L. A. (1993) *J. Immunol.* **151**, 3100–3110.
14. Greenhalgh, P. & Steiner, L. A. (1994) *Immunogenetics*, in press.
15. Yokoyama, W. M. & Seaman, W. E. (1993) *Annu. Rev. Immunol.* **11**, 613–635.