Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 11504–11505, September 1998

Commentary

In an immunological twilight zone

*Jan Klein**

Max-Planck-Institut für Biologie, Abteilung Immungenetik, Corrensstrasse 42, Tübingen, D-72076 Germany

Early in their evolution, perhaps during their transition from Agnatha (jawless fish) to Gnathostomata (jawed animals), vertebrates committed as much as 1% of their protein-encoding genome to a new system of defense against parasites (1). Central to this anticipatory (adaptive) immune system (2) are three types of antigen receptor—immunoglobulin (Ig), T cell receptor (Tcr), and major histocompatibility complex (Mhc) molecules (3)—each consisting of an antigen-binding part and a part concerned with other functions. In the Tcr and Ig molecules, both parts, the variable (V) and constant (C) domains, are drawn from the multifarious Ig superfamily of proteins; in the Mhc molecules, only the part that does not bind antigen comes from the Ig superfamily, whereas the antigen-binding portion has been derived from a different, as yet unidentified source (4). Although jawed vertebrates represent only a tiny fraction of the living world diversity, the parochialism of our species has led immunologists to devote a disproportionate amount of effort to the study of the anticipatory immune system and to neglect the nonanticipatory system on which all other living forms depend. The seemingly sudden appearance of the anticipatory system and specifically of the three types of antigen receptor is acknowledged by both immunologists and evolutionary biologists as a puzzle worthy of resolution. Among the many questions this puzzle raises, two in particular are fundamental; both questions concern the manner in which the antigen receptors function. The individual Tcr and Ig molecules are quite fastidious in their interactions with antigens, each receptor binding a narrow range of antigens and different receptors binding different antigens. The diversity underlaying this receptor selectivity is generated to a large degree somatically by pasting together various combinations of gene segments and then, often but not always, mutating the resulting pastiche. The first fundamental question is therefore: How did the diversity-generating mechanisms come into being? The Mhc molecules, by contrast, are quite promiscuous in their propensity for antigens and correspondingly lack a somatic diversification mechanism; they possess, however, a different, equally bizarre characteristic: they bind antigens only to be seen in their company by the Tcr (i.e., they function as receptors that "present" antigens to other receptors). Hence the second fundamental question is: When, how, and why has this Mhc restriction of antigen recognition arisen? To answer these two questions, it would be of great help to know how the three receptor types originated. Specifically, evolutionarily minded immunologists have been after "primitive" forms of antigen receptors, forms resembling the common ancestor of Ig and Tcr, assuming there was one. A few years ago Greenberg and his coworkers (5) announced that they may have possibly found one such form. Doubts have persisted, however, whether their "new or nurse shark antigen receptor" (NAR) is really a transitional form between Tcr and Ig or simply an Ig variant. To me, the most recent contribution by this group (6) indicates that the doubts were justified.

The nurse shark in which Greenberg and his coworkers (5–7) found the receptor is a representative of cartilaginous fish, the oldest extant branch of jawed vertebrates. The secreted form of the NAR molecule is a homodimer with each chain comprised of six Ig-like domains—one N terminal V-domain and five C-domains. The genes coding for the V-domain undergo somatic diversification in the same way as Tcr and Ig receptors, and their products presumably bind antigen, although this point still needs to be demonstrated formally. So is the NAR an Ig or a Tcr? The authors' original phylogenetic analysis of the five C-domain protein sequences (5) failed to affiliate this part of the NAR unambiguously with either Ig or Tcr, but subsequent analyses (7–9) indicated a clear relationship to Ig heavy chains. The domains are distinctly related to the C-domains of the sandbar shark IgW (8) and skate IgX or IgR (10); all three receptors (NAR, IgW, and IgX) are, in turn, related to shark Ig H-chains of the μ isotype (9). Hence, as far as the constant part of the molecule is concerned, the NAR is clearly a variant of an Ig molecule—an Ig H-chain isotype. This conclusion is further supported by the observation that like Igs, but unlike Tcrs, the NAR—judging from the presence of corresponding signal sequences (5)—occurs in both soluble and membrane-bound forms.

On phylogenetic trees, the V-domain of the NAR seems to be affiliated with the V-domains of some Tcrs (5), and this observation is the sole reason for holding the NAR for a possible intermediate between Ig and Tcr. If, however, the NAR V-domain were really Tcr-like, it would mean either that it has been grafted on the Ig-part of the molecule, for example by an exon shuffling mechanism, or that it acquired its Tcr-likeness by convergent evolution. In the former case, the donor of the grafted domain would presumably be a fully evolved Tcr molecule (gene), whereas in the latter case the origin of the domain would have nothing to do with Tcr. In either case, there would be no reason to consider the NAR a precursor or an intermediate form between Tcr and Ig receptors. The third, and in my mind the most likely possibility is that the Tcr-likeness of the NAR V-domain is an artifact of the phylogenetic analysis.

Evolutionary biologists refer to sequence identities below 25% as the "twilight zone" (11), an area of sequence comparisons in which phylogenetic relationships among taxa are extremely difficult, if not impossible to decipher. The similarity of the NAR V-domain sequence to Tcr lies well within the twilight zone. The effect of the twilight zone is illustrated best by comparing the phylogenetic trees in the two successive publications of Greenberg and his coworkers (5–7): The additions of new sequences have changed the clustering of some of the sequences considerably. The available sequence data are simply not good enough to resolve the branching order and the affinities between individual branches in this case. A

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: Tcr, T cell receptor; Mhc, major histocompatibility complex; NAR, new or nurse shark antigen receptor.

The companion to this Commentary begins on page 11804.

*To whom reprint requests should be addressed. e-mail: jan.klein@tuebingen.mpg.de.

region of the genome evolving under selection pressure exerted by parasites is probably least suited for making assertions about deep-branch divergences. Even at the protein level, most sites have turned over repeatedly (have been saturated) and any phylogenetic signals that may have existed some 400 million years ago have become obliterated. Disregarding the sequence information, two other observations can be used to argue that, in fact, the entire NAR molecule, including its V-domain, is an Ig. One is that modeling (6), for whatever it is worth, matches the NAR V-domain extremely well with the V-domain of the camel Ig molecule which, like the NAR, is an H-H chain dimer rather than H-L chain tetramer of a typical Ig (12). Hence, the tertiary structure of the NAR V-domain is that of an Ig and not of a Tcr molecule. The second observation further supports this conclusion: All known Tcr V-domains contain a hypervariable region 4 (HV4) and certain other characteristics, which are lacking in Ig V-domains and also in the NAR V-domain. There is thus strong support for the thesis that the entire NAR molecule, with all its six domains, is an Ig, one of several Ig isotypes that cartilaginous fish have evolved (9). It differs from other isotypes in its absence of association with light (L) chains, but on this score it resembles the camelid Ig molecules (12). The large divergence of the NAR V-domain from other V-domains is probably the result of the adaptations in the primary and tertiary structure necessary to keep the molecule functional after its loss of L-chains. The divergence is not a sign of an old age but of structural adjustments made by natural selection. A more reliable estimate of NAR divergence from other antigen receptor molecules is provided by its five constant regions and these indicate that the NAR separated well after the Ig-Tcr split.

I suggest, therefore, that the precursor of Ig and Tcr has not been found; I doubt, in fact, that it will ever be found. Assuming that it once existed (the alternative being that the two antigen receptor families evolved independently from similarly structured, but otherwise distinct members of the Ig superfamily), can it be expected to have been retained along with Tcr and Ig in the same group of animals? Tcr and Ig presumably evolved because they satisfied the needs of the organism better than their common precursor. If so, then retaining the precursor form would be tantamount to having, in 1998, an airline flying the transatlantic route with a Spirit of St. Louis-type of airplane, along with other airlines transporting passengers in jumbo jets and airbuses. The precursor may have fulfilled its function 400 million years ago under the then existing conditions, but it became obsolete under conditions in which the Tcr and Ig, the immunological equivalents of jumbo jets and airbuses, have taken over. A search for "primitive" immune systems engaging "ancestral" antigen receptors therefore may be all but futile. Studies like those of Greenberg and coworkers reveal the immune system of the descendants of the most ancient jawed vertebrates to be as sophisticated as that of the upstarts, the mammals.

I thank Dr. Colm O'hUigin, Max Planck Institute for Biology, Tübingen, Germany and Dr. Peter Walden, Humboldt University, Berlin, Germany for discussion and advice and Ms. Jane Kraushaar for editorial assistance.

1. Klein, J. (1997) *Scand. J. Immunol.* **46**, 558–564.
2. Klein, J. (1996) *Scand. J. Immunol.* **29,** 449–505.
3. Klein, J. & Hořejší , V. (1997) *Immunology* (Blackwell Scientific, Oxford), 2nd Ed.
4. Klein, J. & O'hUigin, C. (1993) *Curr. Opin. Genet. Dev.* **3,** 923–930.
5. Greenberg, A. S., Avila, D., Hughes, M., Hughes, A., McKinney, E. C. & Flajnik, M. F. (1995) *Nature (London)* **374,** 168–173.
6. Roux, K. H., Greenberg, A. S., Greene, L., Strelets, L., Avila, D., McKinney, E. C. & Flajnik, M. F. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 11804–11809.
7. Greenberg, A. S., Hughes A., Guo, J., Avila, D., McKinney, E. C. & Flajnik, M. F. (1996) *Eur. J. Immunol.* **26**, 1123–1129.
8. Bernstein, R., Schluter, S. F., Shen, S. & Marchalonis, J. J. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 3289–3293.
9. Schluter, S. F., Bernstein, R. M. & Marchalonis, J. J. (1997) *Immunol. Today* **18,** 543–549.
10. Harding, F. A., Cohen, N. & Litman, G. W. (1990) *Nucleic Acids Res.* **18**, 1015–1020.
11. Doolittle, R. F. (1987) *URFs and ORFs: A Primer on How to Analyze Derived Amino Acid Sequences* (University Science Books, Mill Valley, CA).
12. Spinelli, S., Frenken, L., Bourgois, D., de Ron, L., Bos, W., Verrips, T., Anguille, C., Cambillau, C. & Tegoni, M. (1996) *Nat. Struct. Biol.* **3**, 752–757.