An evoluonarily mobile antigen receptor variable region gene: Doubly rearranging NAR-TcR genes in sharks

Michael F. Criscitiello*, Mark Saltis*, and Martin F. Flajnik**†‡

*Department of Microbiology and Immunology, University of Maryland, Howard Hall, Suite 324, 660 West Redwood Street, Baltimore, MD 21200-1559; and ‡National Aquarium in Baltimore, 501 East Pratt Street, Baltimore, MD 21202

Edited by Max D. Cooper, University of Alabama at Birmingham, Birmingham, AL, and approved January 24, 2006 (received for review August 15, 2005)

Distinctive Ig and T cell receptor (TcR) chains define the two major lineages of vertebrate lymphocyte yet similarly recognize antigen with a single, membrane-distal variable (V) domain. Here we describe the first antigen receptor chain that employs two V domains, which are generated by separate VDJ gene rearrangement events. These molecules have specialized “supportive” TcR domains membrane-proximal to domains with most similarity to IgNAR. The ancestral NAR V gene encoding this domain is hypothesized to have recombined with the TRD locus in a cartilaginous fish ancestor >200 million years ago and encodes the first V domain shown to be used in both Ig’s and TcRs. Furthermore, these data support the view that γ/δ TcRs have for long used structural conformations recognizing free antigen.

cartilaginous fish | evolution | γ/δ T cells | T cell receptor | V(D)J rearrangement

The hallmark characteristics of the adaptive immune system are specificity and memory, both conferred by the V domains of Ig and T cell receptors (TcR). All genes encoding antigen receptor V domains are in the same superfamily and are generated by the same gene rearrangement mechanism, yet all jawed vertebrates studied have discrete Ig and TcR loci. Cartilaginous fish are the oldest animals having an adaptive immune system centered on rearranging antigen receptors. They have all four types of TcR (α, β, γ, δ) (1), three Ig isotypes (IgM (2, 3), IgW (4), and IgNAR (new antigen receptor) (5)), the recombination-activating gene recombinase (6), and polymorphic MHC genes (7, 8). Studies of modern sharks may shed light on the origins of adaptive immunity (9).

Although sharks have much of the basic molecular hardware required for adaptive immunity, some differences have been noted between their antigen receptors and those of other vertebrates. Shark Ig loci are found in many “clusters” as opposed to the single translocon organization common to mammals. Each of the hundreds of Ig loci in the shark genome contains V, D (diversity), J (joining), and C (constant) genes and may be prejoined in the germ line (10), begging questions of the regulation of rearrangement (11). It was originally suggested that horned shark TRB was multicluster (12), but this species seems to be an exception, because all four TcR genes are single translocon loci in skate (1). In addition to mono-meric and pentamer's IgM, cartilaginous fish have at least two other IgH isotypes. The poorly understood IgW isotype occurs in multiple forms (13) [as does IgM from elasmobranchs (14)]. IgNAR, which is apparently found only in cartilaginous fish, binds antigen by means of a single V domain (15), which is no more similar to IgV than to TcR V (16, 17). The IgNARV gene undergoes extensive hypermutation resulting in affinity maturation (18). Only the conventional α, β, γ, and δ TcR chains with single C and V domains have been described from shark or other vertebrates (19).

Here we describe a unique antigen receptor chain that blends characteristics of (previously incompatible) Ig and TcR into a TcR chain with two V domains, each encoded by separate rearranging VDJ segments, on a membrane-anchored TcRδ C domain. The membrane-distal domain, christened NAR-TcRδ, is most related to the IgNARV domain and was recombined into the TRD locus >200 million years ago. Considering the reported interaction with antigen by IgNARV domains, this TcR chain provides evidence for direct antigen binding by γ/δ TcRs.

Results

RACE PCR Reveals a NAR Domain. We discovered an Ig superfamily V domain while studying the genetics and expression of TcRδV families in the nurse shark Ginglymostoma cirratum. Using a reverse primer to the TcRδC domain, 5' RACE PCR from RNA of several tissues amplified products of two sizes (Fig. 1A). Cloning and sequencing of the low-molecular-weight band yielded typical TcR leader-V-D-J-C-encoding transcripts, but the higher band unexpectedly contained transcripts encoding an additional V domain, N-terminal to TcRδV. This domain is most similar to V domains of IgNAR, the Ig class of cartilaginous fish that does not associate with Ig light (IgL) chains (Figs. 1B and 2). We named this V domain NAR-TcRδV, which reflects its relatedness to IgNARV and its expression as a TcR, and we call the entire chain NAR-TcRδV (genes encoding the NAR-TcRδV domain are denoted NT) (Fig. 3A). NT gene transcripts encode a typical leader peptide 5' to the NTV segment and have a NTV segment spliced directly to a TRDV segment. A parsimonious model of a γ/δ TcR complex including a NAR-TcRδ chain is depicted in Fig. 1B.

Several forms of NAR-TcRδ were identified, all most related to IgNARV (~30–40% identity; Fig. 1C) (20). Strands c and d deviate most from the Ig superfamily prototype V domain, and the canonical tryptophan of the WYRK motif is not present (21). NAR-TcRδ lacks the cysteines in the c and d strands and complementarity-determining region (CDR)3 that make the additional disulfide bonds in some IgNARV domains (17). However, a conserved cysteine in the a–b loop is free of intradomain cysteines partners in all NAR-TcRδV (see below).

Dedicated TcRδ Domains Support NAR-TcRδV Domains. Genes encoding the same subfamilies of NAR-TcRδ and TcRδV domains are always found in the same RNA transcripts (e.g., NTV1 is only found in transcripts 5' of TRDV1). Alignment of four TcRδV subfamilies that support NAR-TcRδV families 1–4 with typical δV subfamilies (families 5–8) shows the supporting δV domains lack leader peptides and share a cysteine residue in CDR1 (Fig. 1D). Threading these sequences onto crystal structures for IgNARV and human TcRδ predicts that the cysteine in the a–b loop of the NAR-TcRδV domains and in CDR1 of supporting TcRδV domains are in close proximity.

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: TcR, T cell receptor; RSS, recombination signal sequence; CDR, complementarity-determining region; PBL, peripheral blood leukocyte.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ022688–DQ022718).

‡To whom correspondence should be addressed. E-mail: mflajnik@som.umaryland.edu.

© 2006 by The National Academy of Sciences of the USA
proximity and could form a disulfide bond (R. L. Stanfield, personal communication) (22, 23). Thus, the NAR-TcRV domain likely has two covalent anchors to the /H9254 domain.

Genomic Organization of a Doubly Rearranging Receptor. High variability in the sequence and length of CDR3 in NAR-TcRV sequences strongly suggested that these genes undergo V(D)J rearrangement events (Fig. 4 A), prompting an investigation of NTV and supporting TRDV genomic organization. RT-PCR resulted in some incompletely rearranged transcripts that revealed the intergenic sequence between V and D, and D and J segments encoding the NAR-TcRV domain (Fig. 4 B). In two transcripts, the associated TRDV gene is completely rearranged to the typical TcR DD and DJ gene segments. Genomic PCR using NTV forward primers and reverse primers to the corresponding TRDV revealed that a cluster comprising one NTV, one D, and one J segment is 5'-3' of each TRDV exon that supports NAR-TcRV. These TRDV's lost the exon encoding the leader peptide, and thus they can be expressed as protein only when attached to the N-terminal NAR-TcRV domain (Figs. 1 A and 4 B). The NTJ exon splices to a canonical splice site in the TRDV exon. The NTV, D, and J segments are bordered by recombination signal sequences (RSS) known to regulate recombination-activating gene-mediated rearrangement events (24, 25). Like in IgH loci heptamers and nonamers separated by 23-nucleotide spacers lie 3'-5' of the NTV and J segments, respectively, and NTD segments are flanked by RSSs with 12-nucleotide spacers. Like in all other vertebrates examined, TRDD segments, which are flanked by RSS with 5'-12-nucleotide and 3'-23-nucleotide spacers, rearrange to a J segment RSS having a 12-nucleotide spacer. There is neither rearrangement between different NTV clusters (as is true of the IgH and IgL clusters of the cartilaginous fish) (10, 26) nor joining of NTV segments with DD or DJ segments. The incompletely rearranged transcripts provide preliminary evidence that the TRDV, D, and J segments rearrange...
before NTV, D, J joining. As mentioned, the CDR3 sequences at both junctions are diverse in length and sequence (Fig. 4A). Notably, the TcRδVs that support NAR-TcRV are no less diverse in CDR3 than the typical TcRδVs (data not shown) despite the fact that the antigen-binding site of this domain might be occluded by NAR-TcRV. Both identified TRDD segments are used in the supporting TRDV rearrangements, and no preference for particular TRDJ segments has been observed.

**Tissue Expression.** Highest expression of NTV was detected in the thymus followed by spleen, spiral valve (shark intestine), and peripheral blood (Fig. 3A). The NTV1 probe cross-hybridized with the faster migrating IgNAR transcript under low-stringency conditions in the spleen. This band is not surprising considering the sequence homology between the two gene families and the high levels of IgNAR expression in spleen (27). Increasing the wash stringency resulted in loss of the cross-hybridizing IgNAR band. The NTV1 signal was also decreased because of the range of different NAR-TcRV family members (Fig. 1C). Hybridization with a typical TRDV probe that does not support NAR-TcRV revealed the same expression pattern (data not shown) as does a TRDC probe that recognizes all TRD transcripts regardless of V usage, with the exception of higher relative liver and peripheral blood leukocyte (PBL) expression for the TRDC probe and a higher relative expression of NTV1 in the spiral valve. Signal strength from the membrane-expressed, individual NTV family probes is, as expected, much less than that of secreted Ig or housekeeping probes and only a fraction of that of pan-V recognizing TRDC. Unlike what has been hypothesized for some unusual γδ TcR in mammals (28), the NAR-TcR tissue distribution is consistent with the great junctional diversity in CDR3 in suggesting that the chains have not been selected to recognize tissue-specific ligand. A second blot was prepared with immune tissue RNA, electrophoresed longer, and probed with only TRDC and NDPK to confirm that the higher 2.7-kb TRDC-positive transcript is indeed NAR-TcR and not old activity from previous NTV1 probings. This second blot shows significant enrichment of NAR-TcR in the spiral valve.

**Phylogenetic Distribution of NAR-TcR and Number of Families.** The existence of NAR-TcR in other organisms was examined by genomic Southern blotting (Fig. 3B). All sharks tested (separated in phylogeny by up to 200 million years) were positive and had multiple copies of NTV, even under higher-stringency conditions. Weakly hybridizing bands were seen with the batoid species (ray and skate), fugu, and even human, which were washed away under high-stringency conditions; thus, it is not clear whether NAR-TcR existed in the ancestor of all cartilaginous fish and even perhaps all extant vertebrates. However, in silico investigations of fugu, *Xenopus*, and mammalian databases have failed in finding NTV homologues or telltale leaderless TRDV genes. An IgNARV probe hybridized to different bands on the same blot, dismissing cross-
hybridization to the IgNAR loci as an explanation for NAR-TcR’s existence in sharks (Fig. 3B).

We have identified four nurse shark NAR-TcRV families through cDNA and genomic cloning, but the Southern blot suggests that more members await discovery. Consistent with this proposal, screening of a 10× coverage nurse shark bacterial artificial chromosome (BAC) library with a NTV1 probe resulted in 208 clones, 168 of which also hybridized to TRDV probes. In contrast, no NAR-TcRV clone contained the single-copy TRDC gene, suggesting that the NTV genes are 150 kb away from TRDC. Five of the 11 TRDC-positive BAC clones were also positive for TRDV genes that do not support NAR-TcRV, which advocates a large, complex TRD locus with NTV-supportive TRDV minicircles upstream of the traditional V, D, and J arrays (Fig. 5). No BAC clones positive for any TcR probes were also positive for IgNAR.

Initial characterization of nurse shark TcRδ and γ revealed no NAR-like mRNA transcripts by 5′ RACE or cDNA screenings (unpublished observations), nor did differential screening of a shark cDNA library with TRDC and NTV1 probes. Published studies of horned shark TcRβ also did not reveal NAR-TcR (12, 29).

**Relatedness of NAR-TcR to IgNARV and Other Antigen Receptors.**

Amino acid alignment and phylogenetic analysis of many vertebrate V domains confirms the close relationship between NAR-TcR, TcRV, and TcRδ (unpublished data) families used in clones named at left. Conserved cysteine of the f strand and GXG motif of the g strand are highlighted flanking each CDR3. Known and predicted D-segment-encoded amino acids are underlined, ‘‘-’’ are gaps introduced to align Vs and Js, and ‘‘/’’ marks framshifts resulting from nonfunctional rearrangements. (B) Diagrams of clones showing NTV assembly mechanism. Recombination and splicing signals are marked by filled triangles for RSSs and open squares for GT/AG intronic splice sites. An incompletely rearranged 5′ RACE clone from PBL shows RSS 3′ of TRDD and 5′ of TRDJ. These data suggest that the mammalian RSS system for δ is also used in shark (Inset) (52, 53).
It is also notable that the supporting TcRγδV domains form a separate cluster within the major TcRγδV group. The genetic distances between the supporting δV are similar to those of the NAR-TcRV, consistent with en bloc duplications of NTV/NTV-DV genes over evolutionary time (Fig. 5).

**Discussion**

The maintenance of many functional NAR-TcR and their supporting TcRγδV with conserved cysteines (presumably for interdomain stability), many NTV/NTV-DV genes in shark genomes divorced hundreds of millions of years ago, and high expression in lymphoid tissues are all consistent with NAR-TcR playing an important role in the shark immune system. NAR-TcR transcripts comprise 21 of 99 of the TcRγδ repertoire by sequenced full-length 5’ RACE clone count, and the higher band accounts for 51%, 42%, and 53% of the amplification products from PBL, spleen, and thymus by semiquantitative densitometry in Fig. 1A. Because the IgNARV domain recognizes antigen as a single chain (15, 22), and additional domains have not been found on shark TcRγ, it is likely that the NAR-TcR domain will recognize antigen in a similar fashion. Crystalllography of human γδ has shown a smaller angle between the C and V domains than is seen in α/β heterodimers (23); such an orientation in shark could accommodate steric constraints of an additional V domain in an immunological synapse. This angle could lend accessibility to ligand binding at either V domain, so we cannot rule out use of the typical γ/δ binding site. However, we propose that the NAR-TcR uses the typical γ/δ TcR as a scaffold (Fig. 1B), and thus after antigen recognition by the NAR-TcR V domain, the T cell signaling machinery is directed to induce cytokine secretion or cell killing. Quaternary structure modeling predicts that the TcRγδ CDR3 might be occluded by the NAR-TcR domain, which would replace the TcR γδ Vδ Vs in antigen recognition. Like other Igs, IgNAR has both transmembrane and secreted forms, and its gene undergoes extensive somatic mutation after antigenic stimulation. NAR-TcR neither has a secreted form (single band by Northern blotting and absence of alternative splicing from sequence data) nor is somatically mutated (extensive sequence analysis of particular families); thus, although the NAR-TcRV sequence is most related to IgNARV, its other basic characteristics are the same as those of TcRs.

Hallmark transmembrane charged residues of TcRγδC, as well as its short cytoplasmic tail void of obvious signaling potential, predict that the NAR-TcR δ chain is part of a γδ/CD3 complex for cell surface expression and signal transduction. This notion is consistent with the short cytoplasmic tails found on IgNAR and other antigen receptor chains from cartilaginous fish to man. CD3 and CD79 genes have been found in diverse vertebrate groups, including the amphian Xenopus and fish Takifugu (36, 37), and we predict that these signaling chains will be found in elasmobranchs. Thus, it is likely that the evolutionarily mobile NAR V domain signals via CD79 orthologues on B cells (as IgNAR) and CD3 orthologues on T cells (as NAR-TcRδ).

An IgNAR VDJ cluster may have recombined with TRDV genes sometime early in the evolution of modern sharks, 200 million years ago (Fig. 5). The original NTV-DV gene cluster then duplicated many times and is apparently used to a different extent in different shark taxa (as exemplified by fewer bands in the horn shark digests in Fig. 3B). We previously suggested that an IgNAR VDJ cluster recombined with an IgW cluster (an extant elasmobranch Ig isotype), which then gave rise to the IgNAR isotype gene, a possibility that is supported by these data (16). The origins of the NAR V domain is a mystery (38), because it does not show high similarity to any of the known antigen receptor V domains; the simplest interpretation is that the NAR V arose as a gene encoding a single domain, which was distributed to different antigen receptor families (Fig. 5).

Lateral gene transfer and endosymbiosis has forced major reconsideration of prokaryotic evolution (39). Horizontal gene transfer mechanisms have occurred in eukaryotes (40) but are often excluded from evolutionary hypotheses in favor of simpler Darwinian vertical-descent models. The concept of lateral gene flow is also a common theme for evolutionary immunologists, in which the demonstrable transposase activity of the recombination-activating gene products suggested an invading transposon in antigen receptor evolution (41, 42). The NAR domain now requires consideration as a nomadic gene shuttled intragenomically by means of conventional recombination. The exon-shuffling theory asserts that exon-encoded domains facilitated evolutionary swapping and experimentation of functional protein domains (43), and the best empirical evidence for this theory has been obtained from analysis of exon–intron structure of complex eukaryotic genes in relation to the proteins they encode (44). The use of the NAR domain in two very different antigen receptors on two very different lymphocytes is the most explicit example of the exon-shuffling theory of which we are aware.

Comparative antigen receptor studies throughout vertebrates have thus far shown contrasting natural histories for Ig and TcR. Although Ig have used various genomic arrangements, isotypes, and somatic diversification mechanisms in different vertebrates to achieve repertoire diversification (reviewed in ref. 45), TcR studies have yielded relatively few surprises (19). This dichotomy has been linked to the different stringencies required for function as secreted binders of free antigen and recognition of peptide presented by MHC molecules (46). However, because most γδ T cells are not MHC-restricted (47), they may have been afforded evolutionary freedom more similar to that of the direct binding B cell receptors (Igs). Therefore, perhaps it is not surprising that a mobile genetic element encoding an antigen-binding domain would thrive on TcRγδ, giving it an additional tool for antigen recognition by means of a protruding single domain in addition to the planar recognition landscape possible with the TcRγδ and TcRγγ V domains. The NAR-TcR on a clonally expandable T cell could conceivably recognize cell-bound antigen of fungi, parasites, or virally infected cells and direct cellular cytotoxicity in a manner akin to antibody-dependent cellular cytotoxicity by natural killer cells through the TcR instead of Ab and FcR.
Materials and Methods

**RACE PCR.** 5’ RACE PCR products were amplified by using TaqRc primer (5’-GCTGGCCAGAAGACTCTGCCAGTTGAGCACC-3’) and nested TaqRc primer (5’-TTTGTTGATAAAAGCGAG-3’) from adult shark PBLs, spleen, and thymus total RNA and then separated in 1% agarose and visualized with ethidium bromide. The SMART RACE system (BD Biosciences) was used according to the manufacturer’s protocol with 2 × 20 cycles annealing at 58°C. All products were cloned into pCR2.1 with T4 Cloning kit (Invitrogen) and sequenced by the University of Maryland’s Biopolymer Core facility.

**RT-PCR.** Clones in Fig. 4A ending in 810 and incompletely rearranged cDNA clones in Fig. 4B were gathered with Nar-TcR to TaqRc RT-PCR by using the nonnested 6C primer (Fig. 4A) and primers specific for Nar-TcRV1 (5’-ACCCGTGTCCAGGAAAGGTGTG-3’), Nar-TcRV2 (5’-CATCCCTCTATTATTTAGGATTG-3’), Nar-TcRV3 (5’-CTTCAAACAAAGCCAGGATCCAGG-3’), and Nar-TcRV4 (5’-CTTACGGCACAACCAGCACAGACG-3’). Oligo-dT-primed cDNA was made from 5 μg of RNA as described and used as template for PCR amplification (14).

**Genomic PCR.** PCR from genomic DNA isolated from shark erythrocytes using described Nar-TcRV primers (Fig. 2A) and 6V1 (5’-TTCCTGCTTTCTGAAATGTTGCTG-3’) and 6V2 (5’-TCTCTGCTTCTGAAATTTCGCTG-3’) reverse primers shows the arrangement of two entire NAR-TcRV clusters. One microgram of genomic DNA was used as template in a PCR of five cycles annealing at 50°C followed by 25 cycles annealing at 54°C.

**Northern and Southern Blotting.** Total RNA was prepared for Northern blotting as described (48), and 15 μg was loaded for each lane. All probes were labeled with 32P dCTP PCR as described (49), and free nucleotides were removed with Quick-Spin Sephadex G-50 columns (Roche). Probes were amplified by using primers to include most of the V segment (Nar-TcRV1, 5’-GACACACGCAGGACACATTGCTGGTG-3’ and 5’-GTTGTGTTTGATGTGGTCAGAC-3’; IgNAR7A, 5’-GCTTGAGTGGACACACACG-3’ and 5’-ACCGCAGATACGTGCCAC-3’; TaqRc, 5’-TTAAGCCGAATTCTGCGCT-3’ and 5’-AGAATAACCAGATCCTGGCAG-3’) or the loading control nucleotide diphosphate kinase (5’-AACAGGAGGGACACCTTCTC-3’ and 5’-TCACTCATAATCCAGT-3’). Probes routinely labeled to 3 × 10^6 cpm. Southern blot was performed on HindIII-digested and PstI-digested (Roche) genomic DNA as described (50). Organisms included are Pacific hagfish, spotted ratfish, shark, horned shark, sand tiger shark, lemon shark, cownose ray, little skate, Japanese pufferfish, African clawed frog, and human. Final low-stringency wash conditions were 20-min agitation in 2× SSC/0.1% SDS at 55°C, and final high-stringency wash conditions were 20-min agitation in 0.2× SSC/0.1% SDS at 65°C.

**Phylogenetic Analysis.** Multiple alignment of 78 chordate V domains, with emphasis on IgH, IgL, and the four TcR from diverse groups of vertebrates was performed in CLUSTALX (51) and then rearranged cDNA clones in Fig. 4

We thank Yuko Ota for RNA preparation; Helen Dooley and Robin Stanfield for helpful discussions of NAR-TcR structure; Becky Lohr for the identification of the nurse shark TcR domain gene; and J. Cerny, D. Miaz, D. L. Pasquier, and D. Nemazee for critique of the manuscript. The work was supported by grants from the National Institutes of Health (to M.F.F. and M.F.C.).