Immunogenetics (2002) 54:501–512
DOI 10.1007/s00251-002-0479-z

ORIGINAL PAPER

Marilyn Diaz · Robyn L. Stanfield
Andrew S. Greenberg · Martin F. Flajnik

Structural analysis, selection, and ontogeny of the shark new antigen receptor (IgNAR): identification of a new locus preferentially expressed in early development

Received: 30 January 2002 / Revised: 3 June 2002 / Published online: 23 July 2002

Abstract The new antigen receptor (IgNAR) family has been detected in all elasmobranch species so far studied and has several intriguing structural and functional features. IgNAR protein, found in both transmembrane and secretory forms, is a dimer of heavy chains with no associated light chains, with each chain of the dimer having a single free and flexible V region. Four rearrangement events (among *1V*, *3D*, and *1J* germline genes) generate an expressed *NAR V* gene, resulting in long and diverse CDR3 regions that contain cysteine residues. *IgNAR* mutation frequency is very high and "selected" mutations are found only in genes encoding the secreted form, suggesting that the primary repertoire is entirely CDR3-based. Here we further analyzed the two IgNAR types, "type 1" having one cysteine in CDR3 and "type 2" with an even number (two or four) of CDR3 cysteines, and discovered that placement of the disulfide bridges in the IgNAR V domain differentially influences the selection of mutations in CDR1 and CDR2. Ontogenetic analyses showed that *IgNAR* sequences from young animals were infrequently mutated, consistent with the paradigm that the shark immune system must become mature before high levels of mutation accompanied with selection can occur. Nevertheless, also in agreement with the idea that the IgNAR repertoire is entirely CDR3-based, but unlike studies in most other vertebrates, N-region diversity is present in expressed IgNAR clones at birth. During the investigation of this early IgNAR repertoire we serendipitously detected a third type of *IgNAR* gene that is expressed in all neonatal tissues; later in life its expression is perpetuated only in the epigonal organ, a tissue recently shown to be a (the?) primary lymphoid tissue in elasmobranchs. This "type 3" *IgNAR* gene still undergoes three rearrangement events (two *D* regions are "germline-joined"), yet CDR3 sequences were exactly of the same length and very similar sequence, suggesting that "type 3" CDR3s are selected early in ontogeny, perhaps by a self-ligand.

Keywords Evolution · Gene rearrangement · Antigen receptors · Cartilaginous fish · Somatic mutation

M. Diaz
Department of Immunology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

R.L. Stanfield
Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

A.S. Greenberg
Box 111, Dept. of Medicine, VAMC, University of California at San Francisco, San Francisco, CA 94143, USA

M.F. Flajnik (✉)
Room 13-009 BRB, 655 West Baltimore Street, Department of Microbiology and Immunology, University of Maryland at Baltimore, Baltimore, MD 21201-1559, USA
e-mail: mflajnik@som.umaryland.edu
Tel.: +1-410-7065161, Fax: +1-410-7062129

Introduction

The jawed vertebrate adaptive immune system is defined by its diverse antigen receptors, the immunoglobulins (Ig) and T-cell receptors (TCR). Animals in different taxa have developed unique ways of generating diversity in their primary Ig repertoire, as well as evolving novel molecules specific for some groups (reviewed by Litman et al. 1999). The cartilaginous fish (sharks, skates, rays, and ratfish) are the oldest vertebrates with an adaptive immune system, and they possess Igs that seem unique to this taxon (reviewed by Flajnik and Rumfelt 2000a). In addition to IgM (present in two forms, 19S and 7s, in approximately equal amounts), two other Igs, IgW and Ig new antigen receptor (NAR) are present in all elasmobranchs so far examined (Nuttall et al. 2001, 2002; Rumfelt and Flajnik, unpublished).

The *IgNAR* gene family has been best studied in the nurse shark and is composed of only 4–5 loci, each with its own cluster of a single variable (*V*) region, three diversity (*D*) elements, and a single joining (*J*) gene (Greenberg et al. 1995). NAR lacks diversity found in mammalian Ig genes via Ig heavy (H) chain-light (L) chain associ-

ation, as NARs are covalently linked H chain dimers with free and flexible *V* regions (Roux et al. 1998). Furthermore, we showed previously that *NAR V* genes are extensively modified by somatic hypermutation, not as a repertoire-building strategy, but probably after antigenic stimulation (Greenberg et al. 1995; Diaz et al. 1998). Thus, given the small number of *NAR* loci, the lack of associated L chains, no combinatorial diversity, and the finding that somatic hypermutation does not generate the repertoire, NAR primary repertoire diversity might be severely limited. However, sequence analysis of the NAR third complementarity determining region (CDR3) has revealed extensive heterogeneity, principally through N-region addition at the four rearrangement breaks (*V-D1-D2-D3-J*). Thus, the NAR primary repertoire is quite heterogeneous but entirely CDR3-based, which also may be true of conventional antibodies (Xu and Davis 2000).

Although NAR repertoire diversity is dependent solely on CDR3, there are nevertheless restrictions on its CDR3 amino acid composition. *NAR* type 1 and type 2, the two loci specifying the adult repertoire, are distinguished by the number and positioning of various cysteines generally not found in vertebrate IgV regions (Roux et al. 1998). These non-canonical cysteines in both NAR types are under strong selection, as they are rarely replaced by somatic mutations (Du Pasquier et al. 1998; Diaz et al. 1999), and despite the unpredictable outcome of *VDJ* recombination, most NAR CDR3 include cysteine (Cys) residues. An example of evolutionary convergence, some forms of camel/llama IgG V_H, which also do not associate with L chains, have non-canonical cysteines in similar positions as the NAR proteins (including CDR3), suggesting that such cysteines are required to stabilize single-domain receptors (reviewed in Nguyen et al. 2001).

In mouse Ig and TCR genes, the neonatal repertoire is less diverse than the adult primary repertoire, in part because of the absence of N-region addition during rearrangement (Feeney 1990; Bogue et al. 1991). This is caused primarily by delay in expression of terminal deoxynucleotidyl transferase (TdT), the enzyme responsible for this activity. The mammalian neonatal B-cell repertoire is also characterized by predominant usage of particular V_H gene families (reviewed in Kearney et al. 1997). The protective mammalian uterine environment may not select for great diversity in antibodies and TCR, but rather for evolutionarily selected specificities that might protect against common pathogens or serve 'household' functions. In aplacental viviparous sharks, however, pups are bathed in seawater for at least some period of their intrauterine lives and may be exposed to pathogens (Kormanik 1988). Thus, given the dependence of the NAR repertoire on the generation of highly diverse CDR3s, we hypothesized that, unlike mouse Ig, the nurse shark neonatal NAR repertoire would display N-region additions. Similarly, we expected that the somatic hypermutation machinery would be fully operative in newborn sharks. To examine these issues, we PCR-amplified NAR *VDJ* sequences from various tissues of near-full-term pups and analyzed the entire V region for somatic hypermutation and the CDR3 for evidence of N-region addition and for amino acid composition. During these studies, a third NAR type, rarely expressed in adults, was discovered and found to dominate the neonatal repertoire.

Materials and methods

Animals

Three pregnant females were caught off the coast of Key West, Florida in 1996, 1997 and 1998, in late October/early November. Shark pups were delivered by Caesarian section as described (Ohta et al. 2000). Some of the shark pups were transported to the University of Miami Marine School Hatchery (Key Biscayne, Florida) and maintained in aerated seawater for the duration of the study. Three pups were immediately killed, while others were killed at 2-month intervals for harvesting of various organs.

RNA isolation and RT-PCR

Spleen, kidneys, intestine, and epigonal organs were harvested from each pup and total RNA was isolated using the Trizol Reagent (Invitrogen, Carlsbad, Calif.) following the manufacturer's protocol. cDNA from each tissue was made with oligonucleotides complementing either the transmembrane or the secretory regions of NAR, as described (Diaz et al. 1998). Oligonucleotides used for cDNA synthesis and to PCR-amplify NAR V and J regions, as well as the PCR conditions, were also described previously. PCR products were run in 1.5% agarose gels, and bands of the expected size (~320 bp) were excised and purified using the Gene Clean II Kit (QBiogene, Carlsbad, Calif.). Ligation and transformation of competent cells were done using the Original TA cloning Kit (Invitrogen). Culture of transformed cells, and plasmid isolation was done as described (Diaz et al. 1998). Plasmid DNA (1 μg) from each clone was sequenced at the University of Maryland at Baltimore Biopolymer Facility.

Tertiary structure modeling of IgNAR

Computer modeling was carried out using the Swiss-PDB Viewer (Guex and Peitsch 1997). The template used for the final NAR models was the human $V\kappa_L$ domain of the chimeric human/mouse Fab B72.3 (PDB code 1bbj, Brady et al. 1992). CDR3 loops were modeled by hand, with bent loops in the NAR type-1 models based on an H3 loop of similar size in the camel V_H domain (PDB code 1mel, Desmyter et al. 1996). Additional energy minimization was done with CNS (Brunger et al. 1998) after modification of CDR3 loops.

Sequence alignment and analysis

Sequence analysis and preparation for alignments were done using the BCM Search Launcher sequence utilities Internet program (Smith et al. 1996) and sequences were aligned using the Clustal W program at EMB, also available through the internet (Thompson et al. 1994). Statistical analysis was done using the G-test for goodness of fit as described (Sokal and Rohlf 1981).

Results

Analysis of Cys residues in NAR CDR3

In our previous study of NAR V regions (Roux et al. 1998, Fig. 4a), we described two major forms that differ

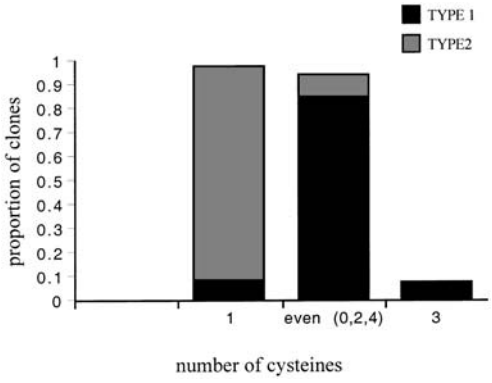


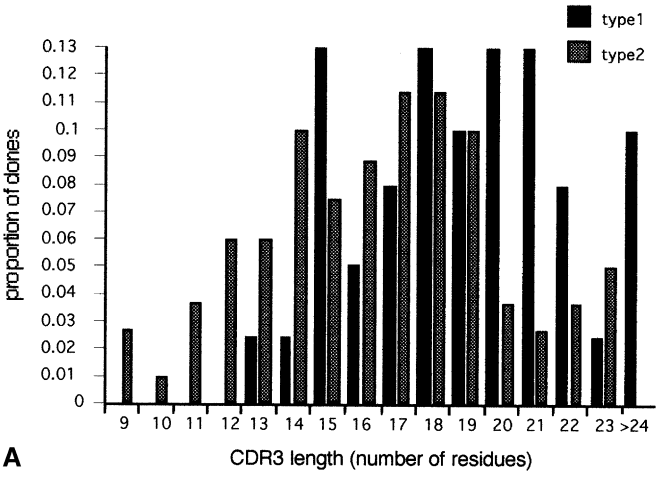
Fig. 1 The number of cysteines encoded within CDR3 in NAR type-1 and type-2 clones. The great majority of type-1 clones (*n*=43) contained two or four Cys residues, while most type-2 clones (*n*=77) encoded only one Cys in CDR3. The data were obtained from three adult nurse sharks

chiefly in the number of non-canonical Cys residues: type-2 NAR has a single Cys in CDR3 in addition to a germline-encoded Cys in CDR1, and type I has an even number of Cys residues in CDR3 (two or four) and two Cys residues in the core V domain (FR2 and FR4). Here we analyzed sequences from all nurse sharks so far studied and quantified the number of clones having the representative structure. Type-1 NAR CDR3 most often had an even number of Cys residues and the majority of type-2 clones contained one Cys in CDR3 in both adults (Fig. 1) and neonates (Table 1). Type-1 NAR CDR3 sequences are longer and somewhat more variable in length than type-2 NAR CDR3 (Fig. 2a). Figure 2b shows a representative sample of type-1 and type-2 CDR3 sequences from one adult shark (a display of CDR3 sizes and number of Cys residues).

Selection on type-1/type-2 NAR

To examine whether these features of NAR type 1/type 2 have an impact on antigen binding, we analyzed the replacement to synonymous (R/S) hypermutation ratios in CDR codons in a large library of clones obtained from adult PBL in a previous study (Diaz et al. 1998). A high R/S ratio in CDRs is indicative of positive selection (Wagner and Neuberger 1996). Surprisingly, a significant number of replacements in CDR1, but not CDR2, were found for type-2 NAR clones, while the opposite was true of type-1 NAR clones (Fig. 3, Fig. 4a displays the

Fig. 2a, b Type-1 NAR CDR3 tends to be longer and more variable in length than type-2 NAR. **a** Type-1 NAR CDR3s averaged 19 residues in length (SD 3.5), while type-2 NAR CDR3 averaged 16 residues (SD 3.4). Sample sizes/number of sharks are the same as in Fig. 1. **b** Type-1 and type-2 NAR CDR3 sequences from one adult shark. The CDR3 sequences encompass the canonical cysteines in the F strand of the V domain and the residue adjacent to the invariant Gly-X-Gly in Fr 4 (*J* region). CDR3 length and number of cysteines are shown on the right. Type 1: mean 20.9 (SD 3.04); type 2: mean 15.3 (SD 2.9)



A

NAR Type 1 CDR3

```
CGASGWCVAY SCDYVSHSPVLSHLNYAAC  27-2
CGVTIEAGVAGSCDYGPVCDLGT  NYAAC  26-2
CGVCRIGWCPGCETARSVCSRS     SAC  23-4
CAVGGIVGV  SCDYGL CAVEGA  YAAC  23-2
CGVYMLVWDY SCDYDL CSSPN    AAC  22-2
CGVRCDIRNSCEDGHTLVMFG      AAC  22-2
CAASWIAAR   CDYPELCSSRDF   AAC  22-2
CGVCMWDGGCCITKSS  CALDY    AAC  22-4
CGVPSRLLPGLGCDAL  CSWDYT    AC  22-2
CGAAVS   GLGCDPRV CSWSVGYH  AC  22-2
CGTAQGNSW  SCDYPL CALGGVL   AC  22-2
CGVWCGTRSVRDDKL   CSN    NYAAC  21-2
CGVCKRVCYP  CDLYLGCALL     AAC  21-4
CGRRAG      CDYG  CALGGHDDAIAC  20-2
CGVSRLLWCQTKL     CSFPST    AC  19-2
CGACFGVC   SCDYGV CSRG     AAC  19-4
CGVYRWRGAQLCSCSH           AAC  17-2
CGAPVSLLSA SCDCR          YAAC  17-2
CGVASC      CDYGLSLDED     AAC  16-2
CGYGH       CDYSI CSAR    YAAC  16-2
```

NAR Type 2 CDR3

```
CKWSAGMDPEGYDCGVELK   DVY       21-1
CSGLTSDV   YDCQSWNSYEVD Y       18-1
CKGGL VWTD YDCSPDWPG   VY       18-1
CKVLPLVWIGVCTPSVG     DVY       18-1
CSDRRTYT  VCDSRTGS    DVY       17-1
CKTSPSVYCRDGDFGPS       Y       16-1
CKVPCGYHLDWRPRH        VY       15-1
CHAVTYG     DCAGQTD   DVY       15-1
CQVESWLGLEPCPYE         Y       14-1
CKRRSLSWCKSEP          DVY       14-1
CNSYWSS    YDCPLYEL     Y       14-1
CNPNTCIWGPYP           VY       12-1
CKGL      VCRGRNWD     VY       12-1
CKVAD       DCWRH      VY       10-1
```

B

Table 1 NAR types 1 and 2 in nurse shark neonates underwent extensive N-region addition

a. Type 1

Clone	F-strand	N/P	D1 ATAGCTG GTGTGGA	N/P	D2 TAGCTG TGACTAC	N/P	D3 CTGTG CTCTTG	N/P	J ACAACTAT GCTGCATGC	Translation
s-1649	ggtg	gcgtcccc	agctggt	ccgg	tgtgactac	ccctg	tgtgctcttg	tgggata	tatgctgcatgc	GGVPSWLGVTTPVCLCGIYA
s-1650	ggtg	at	ctggtgtgga	–	ctgtgactac	–	gct	–	ctatgctgcatgc	GDLVWTTLYA
s-1646	ggtgtc	ggc	gg	gc	gctgtgac	aag	ctgtgctctt	–	tatgctgcatgc	GVGGRCDKLCSYAA
s-1647	ggtgt	ggg	agctggtgtgg	t	tagctgtg	ggcag	ctgtgctctt	c	caactatgctgcctgc	GVGAGVVSCGQLCGSNYAA
s-1645	ggt	c	gtgtg	tatcggggtag	tagctgtga	aaacttt	ctgtgctct	caggactc	tgctgcctgc	GRVYRGSSCENFLCSQDSAA
s-181	ggtgtctg	–	ctggtgt	cc	agctgtgact		gctcttg	ggat	ctatgctgcatgc	GVCWCPAVTALGIYAA
s-183	ggtg	cctcat	ctggtgtg	agg	gtgac	cacaagt	ctgtgctcttg	ggggt	acaactatgctgcatgc	GASSGVRVTTSLCS WGYNYAA
s-185	ggtg	–	atactgtg	atgt	ctgtgctct	–		ta	tctgctgcctgc	GDTVMSVLLSAA
e-746	ggtgtctggt	cc	agctg	–	tgact	gg	gctcttg	atccgg	tatgctgcatgc	VSGPAVTGLLIRYAAC
e1732	ggtgt	a	ctggt	tcgcta	tgtgact	cgg	gctctt	cc	ctatgctgcatgc	GVLVRYVTRALPYAAC
e-1675	ggtg	gcgtcccc	agctggt	ccgg	tgtgactac	cc	ctgtgtgctcttg	tgggata	tatgctgcatgc	GGVPSWSGVTTPVC SCGIYAAC
e-1736	ggtgtctgctggt	ccc	gctg	–	tgac	c	gtgctc	cggt	ctatgctgcctgc	GVCWSRCDRAPVYAAC
e-1750	ggt	cgtgtgtatcgg	ggt	ag	tagctgtga	aaacttt	ctgtgctct	cagg	actatgctgcctgc	GRVYRGSSCENFL CSQDYAAC
k-tail-typ1–3	ggtgtctg	–	ctggtgtg	–	ctgtgactac	–	tgtgc	gcgg	ctatgctgcatgc	GVCWCAVTTVRGYAAC
k-tail-typ1–5	ggtgtc	aggg	tgg	–	ctgt	cggag	tgtgctc	gt	acaactatgctgcatgc	GVRVAVGVCSYNYAAC

b. Type 2

Clone	F-strand	N/P	D1 ATAGCTG GTATGGA	N/P	D2 TAGCTA TGACTGC	N/P	D3 CTGAA CTGGAGA	N/P	J ACTATGA TGTATAC	Translation
s-168	aacgtata	ct	tagctgg	cccc	tatgactgc	ga	ct	–	actatgatgtatac	NVYLAGPYDCDYYDVY
s-179	aacgt	c	ctggtatg	c	ctatg	ga	actgg	gg	tgatgtatac	NVLVCLWNWGDVY
s-169	aacgtat	cc	ct	acgc	ga	gtg	ctgg	gg	ctatgatgtatac	NVSLRECWGYDVY
e-680	aaggta	cc	gctggt	–	tat	–	tgaactggaga	tcccg	tatgatgtatac	KVPLVIELEIPYDVY
e- 1759	aaggt	gcg	g	–	–	–	gaac	cccaca	actatgatgtatac	VREPHNYDVY
k-tail-typ2–1	acg	ccttat	atagctggtat	tccc	gct	–	gaac	ga	gatgtatac	TPYIAGIPAERDVY

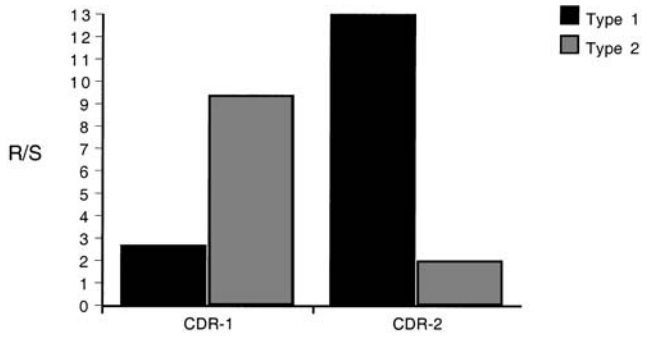


Fig. 3 Differential selection on CDR1 and CDR2 in NARs type 1 and type 2. CDR1 in type 2 and CDR2 in type 1 have high ratios of replacement/synonymous (R/S) substitutions. Sequences in this analysis were derived from secretory cDNA clones with less than 15 mutations from three adult sharks (n=72). See Fig. 4a for the definition of NAR CDR1 and CDR2

CDR). These data suggest that the positioning of disulfide bridges in NAR type 1 and type 2 profoundly influences the V-domain structure, differentially exposing CDR1 and CDR2. This result prompted us to re-examine models of NAR structures.

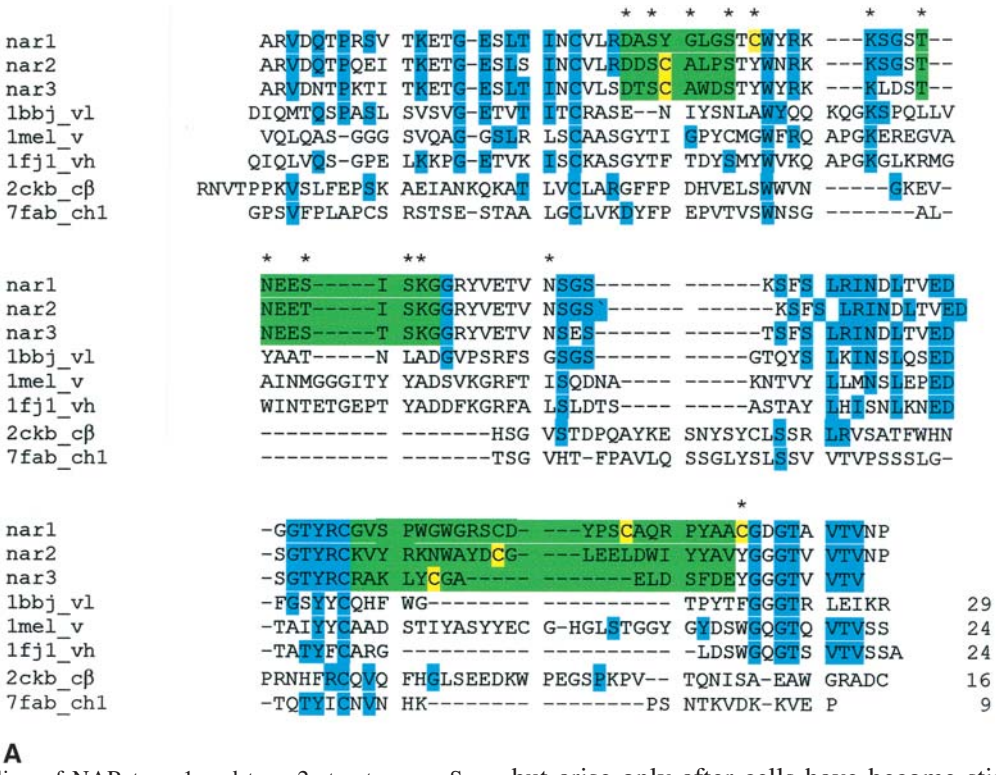
Cysteines in NAR CDR3 are predicted to facilitate the formation of a large antigen-binding surface encompassing CDR3/CDR1 in type 2 and CDR3/CDR2 in type 1

The importance of the germline (CDR1/2) and the somatically encoded (CDR3) cysteines in NAR proteins was examined by modeling NAR types 1 and 2. Ig-like domains are classified into topological subtypes depending on the number of β-strands in each sheet, and the position of the edge strands in each sheet (reviewed by Bork et al. 1994). While NAR type-1 and -2 sequences are members of the Ig fold family, a deletion in FR2/CDR2 makes it difficult to assign the sequence to a topology subtype. Proteins within an Ig family subtype usually have the same number of residues separating the first invariant Cys and the invariant Trp, and the separation between that Trp and the second invariant Cys (see Fig. 4a). In NAR type-1 and -2 sequences these distances are 13 residues (Cys1-Trp) and 46 residues (Trp-Cys2). While the Cys1-Trp value is common among other domain types such as IgV_H, the second Trp-Cys2 value is unusual, found mainly in TCR, IgL and MHC C1-type domains (Williams and Barclay 1988). However, NAR/C1 domain sequence identity is lower than that between NAR and IgV domains. Unsure of the correct topology subtype, we modeled the NAR type-1 and -2 structures using IgV_L, IgV_H, TCR Cβ, IgC_H, and camel V_H domains as templates. Models based on the IgV_H, TCR C_β, IgC_H, and camel V_H templates all resulted in CDR2 residues at the "bottom" of the molecule, not contiguous with CDR1 or 3. However, a model based on the IgV_L domain, which like NAR has a very short CDR2 region (Fig. 4a), resulted in a structure where the three CDR clustered at the "top" of the V domain. From these results, we feel that the NAR models based on IgV_L domains are most reasonable. NAR type-1 and -2 CDR3 loops are both long (particularly for type 1), and, with no canonical structure to guide us, this loop cannot be modeled with certainty. However, models that we have generated show that it is possible for several arrangements of potential disulfide bonds to be made. In addition to the two characteristic Cys residues in type-1 CDR3, there are two other non-canonical cysteines: one in FR2 (C-35), adjacent to the invariant Trp, and one in FR4 (C-107), flanking the invariant Gly-X-Gly (Fig. 4a). In our previous model we suggested that, like other CDR3 regions encoding an even number of cysteines, NAR CDR3 cysteines covalently bonded with each other (Roux et al. 1998). We also suggested that the FR2 and FR4 cysteines, which point outward in the area normally occupied by L-chain V domains, would form a disulfide bond. However, we could never persuade this FR2-FR4 disulfide bridge to form when modeling onto existing Ig superfamily structures, and thus now we propose an alternative model in which the two somatically generated CDR3 cysteines form disulfide bonds with the two FR cysteines (Fig. 4b). In this model the CDR3 loop is bent over and "rests upon" CDR2, closely resembling the camel V_H domain structure that also has a long CDR3 loop disulfide-bonded to a Cys encoded in the main body of the protein (Desmyter et al. 1996, but note there is only one disulfide bridge formed in camel Ig). Type-2 NAR is predicted to form a disulfide bond between the germline-encoded cysteine in CDR1 and the somatically generated cysteine in CDR3, as suggested previously (Roux et al. 1998, Fig. 4b).

This model proposes that the germline-encoded CDR loop adjoining CDR3 (CDR1 in type-2 NAR and CDR2 for type-1 NAR) contributes most significantly to antigen binding; the finding of a high R/S ratio in the CDR adjacent to CDR3 is also consistent with the idea that CDR3 is responsible for the primary NAR repertoire (Diaz et al. 1998). In the stereoimages (Fig. 4b), note that two residues in CDR2 (*purple*) under strong selection in type-1 but not type-2 structures, N45 and N48, are in close proximity to CDR3 only in NAR type 1. Conversely, CDR1 (*red*) residues D26 and S28 are under strong selection in NAR type 2 and are near CDR3 only in that structure. Obviously, the next mission is to generate true structures of NAR type 1/2 and test our models.

CDR3 are diversified by N-region addition in neonates

Because the primary NAR repertoire is entirely CDR3-based, and shark pups are exposed to seawater for at least some phase of intrauterine life, we predicted that CDR3 would be well diversified in neonates. Indeed both NAR type 1 and type 2 had P-nucleotide and N-region addition (examples in Table 1 for spleen-derived clones). The bias for G and C nucleotides in the identified N-regions suggests that TdT, recently identified in sharks (Rumfelt et al. 2001; Bartl et al. unpublished) and



A

Fig. 4a, b Modeling of NAR type-1 and type-2 structures. **a** Sequence alignments of NAR type 1 and 2 (and type 3) with other Ig superfamily domains (based upon structural alignments). Non-canonical cysteines in NAR are highlighted in *yellow*. CDRs are highlighted in *green*, and residues with sequence identity to NAR type 1 are highlighted in *cyan*. The number of residues identical to the NAR type-1 sequence is shown at the end of each sequence. *Asterisks* denote residues displayed on the structures in **b**. Sequences and coordinates used are the chimeric antibody B72.3 (1bbj), camel VH domain (1mel), mouse αβ TCR 2C (2ckb), and human λ chain NEW (7fab). **b** Space-filling (*top*) and steroimages of NAR type 1 and type 2. CDRs are colored *blue* (CDR1), *magenta* (CDR2), and *red* (CDR3). Residues marked with an *asterisk* in **a** are displayed on the structures. The two somatically generated cysteines in type-1 CDR3 were modeled to make disulfide bonds with C-35 (FR2) and C-99 (FR4); the one CDR3 cysteine in type 2 was modeled to make a bond with C-25 in CDR1. Figures were generated with MolScript (Kraulis 1991) and rendered with Raster3D (Merritt and Bacon 1997)

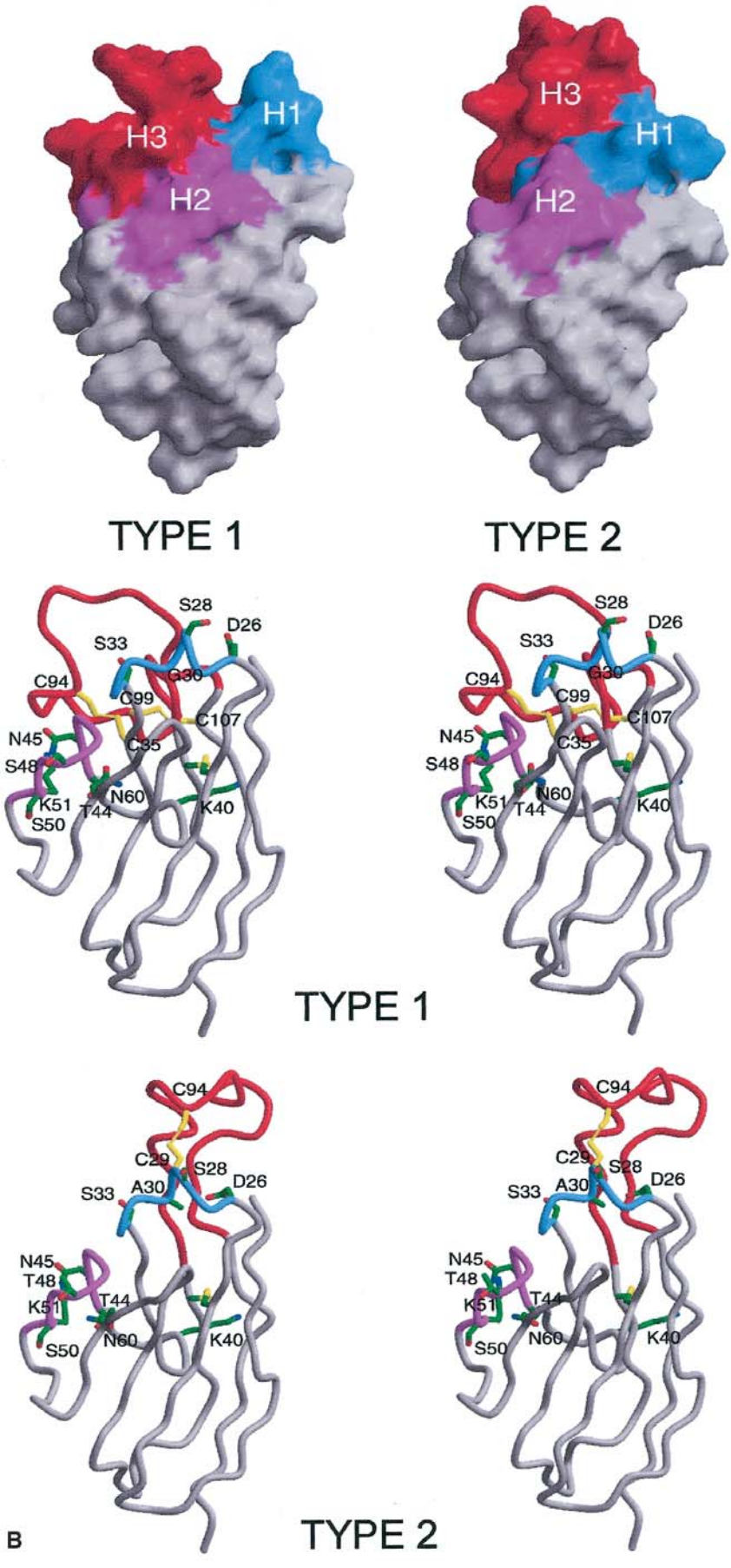
skates (Miracle et al. 2001) to be expressed early in development, is responsible for the diversity. Thus, in contrast to studies of the neonatal Ig/TCR repertoire in most other vertebrates (reviewed by Flajnik and Rumfelt 2001b), shark type-1 and type-2 *NAR* genes undergo extensive N-region addition in neonates. Indeed, the skate study mentioned above showed that IgM H-chain *V* genes also have N-region addition from the earliest stages analyzed (Miracle et al. 2001); also true of nurse shark IgM genes, Rumfelt and Flajnik, unpublished

Shark pup NAR hypermutation: low frequency and no CDR targeting

As mentioned above, our previous studies reported that mutations do not generate the primary NAR repertoire, but arise only after cells have become stimulated to become secretory cells (Diaz et al. 1998, 1999). The NAR loci expressed in adults, types 1 and 2, and a third type rarely seen in adults (type 3, see below) were found in cDNA derived from neonatal spleen, epigonal, kidney and intestine. A total of 180 mutations were observed in 228 clones from neonates and two-,four-, and six-month-old pups (84/138 secretory). Neonatal clones had fewest mutations (39 mutations/99 clones, 27/56 secretory), while the overall frequency for all four age groups was 0.85 (SD 0.33) mutations per clone. This low level of mutation in secretory clones from young animals contrasts with that detected in adult secretory clones, where the overall frequency was 5.14 (SD 0.49) mutations/clone, but rather resembles adult transmembrane (Tm) clones that had an overall frequency of 1.65 (SD 1.13) mutations/clone (Diaz et al. 1998).

Several features of these mutations indicate that they result from the previously reported NAR hypermutation mechanism (Diaz et al. 1998, 1999): (1) the mutation frequency, albeit low, is significantly higher than the empirically determined PCR error rate (0.0009/base); (2) previously identified hotspots of NAR hypermutation, such as AGC/T, AAC, and AAA, were highly targeted in pup clones (33%); (3) there was a transition bias (56%), a characteristic of somatic hypermutation in all species. However, in contrast to adult secretory clones (59%), but like adult Tm clones (35%), only 28% of mutations in shark pup secretory clones fell within CDR1/2. Particularly unlike adult NAR cDNA clones, there was a paucity of tandem mutations, a unique feature of NAR hypermutation (and perhaps of shark hypermutation in general, Lee et al. 2002), among neonatal- and young shark-derived clones (Diaz et al. 1999). Like adult Tm clones,

Fig. 4b


H3
H1
H2
TYPE 1
H3
H1
H2
TYPE 2
S28
D26
S33
C94
G30
C99
C107
N45
C35
S48
K51
T44
N60
K40
S50
S28
D26
S33
C94
G30
C99
C107
N45
C35
S48
K51
T44
N60
K40
S50
TYPE 1
C94
C29
S28
D26
S33
A30
N45
T48
T44
K51
S50
N60
K40
C94
C29
S28
D26
S33
A30
N45
T48
T44
K51
S50
N60
K40
B
TYPE 2

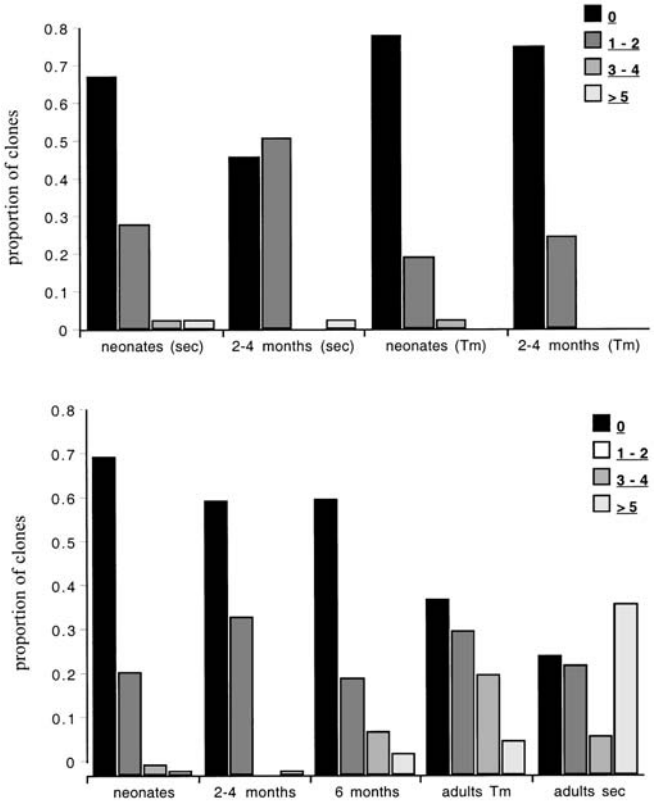


Fig. 5 Low frequency of mutations in both NAR secretory (*sec*) and transmembrane (*Tm*) (*top*) and total (*bottom*) clones derived from shark pups. Clones are displayed having 0, 1–2, 3–4 or >5 mutations. Note that the mutation frequency for pup secretory clones is similar to adult Tm clones

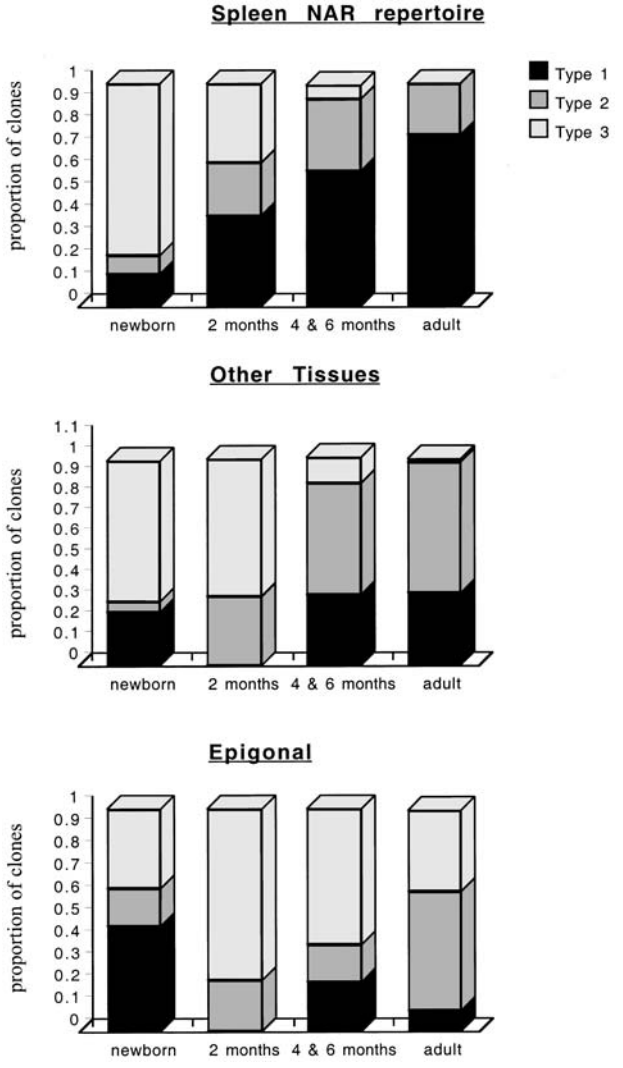


Fig. 6 NAR repertoire shift during ontogeny. Two hundred and fifty-seven NAR RT-PCR-derived clones from shark pups ranging from newborn to six months of age, and 184 adult-derived clones were examined for NAR type-1, -2, and -3 frequencies. Note that expression of NAR type 3 is perpetuated only in the adult epigonal organ. 'Other tissues' included kidney, intestine, and in one case (six months) thymus

the low levels of mutation in shark pups were due to a prevalence of clones with 0–2 mutations in both secretory and Tm NAR (Fig. 5). In the previous analysis of adult Tm vs secretory clones, we attributed this low mutation frequency and poor CDR clustering to "leaky hypermutation machinery," possibly operating during development. Interestingly, while 27% of neonatal-derived clones were out-of-frame, only 8.6% of the mutations were found in these clones, suggesting that 'leaky' mutation occurs only after selection of NAR$^+$ cells with productive V(D)J rearrangements (see Discussion).

The low frequency and poor CDR targeting of mutations, as well as the relative lack of tandem mutations in shark pup NAR clones, is consistent with the lack of developed lymphoid structures in neonates (Rumfelt, McKinney, Taylor, and Flajnik, submitted), and suggests that somatic hypermutation does not contribute to NAR diversity or specificity early in shark development.

The majority of neonatal-derived NAR clones originated from a third locus

All adult cDNA clones obtained from peripheral blood lymphocytes (PBL) (Diaz et al. 1998) and spleen (Greenberg et al. 1995) were either type-1 or type-2 NAR. Surprisingly, a third form, uninspiringly named type 3, constituted 61% of the neonatal-derived clones, particularly in the spleen, but it was prevalent in all tissues tested. To examine whether this finding was reflective of a repertoire shift during development, we obtained clones from progressively older shark pups, and found a gradual decline in the frequency of type-3 clones in all tissues except the epigonal organ, where the diminution was much less dramatic (Fig. 6).

A type-3 genomic clone was isolated by PCR amplification of genomic DNA and sequenced (Fig. 7). The *D1* and *D2* segments of a type-2-like clone were fused in what appears to be a germline-joining event (Kokubu et al. 1988); otherwise, this gene is much like the previously reported *NAR* type-1 and type-2 genes (Greenberg et al. 1995). Note that it is possible that the two-D type-3 gene is

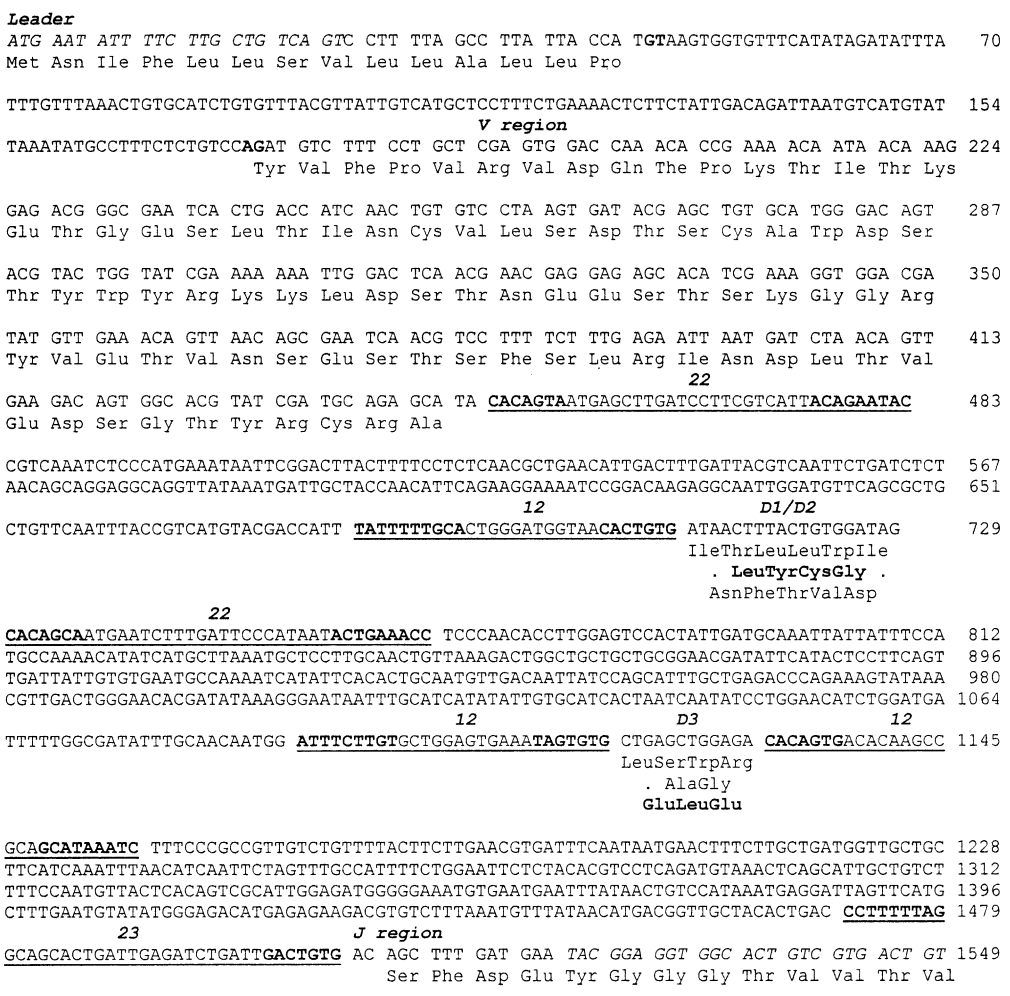


Fig. 7 NAR type-3 gene cluster. The gene was isolated by PCR from genomic DNA as described (Greenberg et al. 1995). Recombination signal sequences are *underlined*, with *bolded* nonamers and heptamers. Splice junctions are *bolded and italicized*. Translations for the three D segment reading frames are shown and the only reading frame found in cDNA clones is bolded for each segment. Primers in the leader and *J* region used for PCR amplification are *italicized*

ancestral to the three-D conventional *NAR* genes, but the high frequency of germline-joined genes in sharks and the early expression of the type-3 gene (see Discussion) suggest that the type-3 configuration is derived from type 1/2.

All type-3 cDNA clones employed the same reading frame in *D1/D2* (Leu-Tyr-Cys-Gly) and *D3* [Glu-Leu-Glu(Asp)] (Fig. 7 and Table 2); thus, like type-2 clones there was always a single Cys in CDR3 (and one Cys in CDR1). However, unlike type-1 and -2 CDR3s, type-3 CDR3s had three unique characteristics: (1) less N-region addition; (2) high amino acid similarity among clones; (3) remarkable conservation of CDR3 length, all clones but one being 15-amino-acids long (Table 2). A model of the type-3 gene suggests a rather rigid CDR3, with a CDR1 Trp specific to type-3 NAR (residue 31, adjacent to the C29-C89 disulfide bridge connecting CDR1 and CDR3) apparently packing with CDR3 (Fig. 8). The two amino acids encoded by the rearrangement breaks (A91 between *D1/D2* and *D3*, and *D94* between *D3* and *J*) are shown to complement the alignment in Table 2.

There are several features of type-3 NAR junctions and D elements that can partially account for the conservation of CDR3 length: (1) type-3 NAR undergoes only three rearrangement events (*V-D1/D2-D3-J*), because of the *D1/D2* fused element [types 1 and 2 generally undergo four rearrangement events (*V-D1-D2-D3-J*)]; (2) there is a paucity of N-region addition at the rearrangement breaks; (3) the *D3-J* junction may sometimes undergo homology-based joining (Table 2). However, these characteristics do not fully account for the remarkable conservation of type-3 CDR3 length. Even when few (or no) nucleotides were added at one junction, there was compensation either by increasing the number of nucleotides used in the other D element or in the N-regions. Furthermore, the few clones with significant N-region addition were always accompanied by fewer nucleotides (probably by nibbling into the D) used in one of the D elements or in the end of the *V* gene (see clones 171, 286, 283, 294, and 291; but note also that there was not much nibbling into the Ds). The type-3 NAR CDR3 most closely resembles rearrangement junctions in neonatal mice, where homology-based joining and lack of N-regions yields Ig *V* gene CDR3s with similar amino acid composition (Asarnow et al. 1988; reviewed in Kearney et al. 1997). Nevertheless, type-3 NAR neonatal clones provide the first example of perfect conservation of CDR3 length and high similarity in amino acid composition following (three!) *VDJ* rearrangements that incorporate N-region additions. This result suggests that a very stringent selection mechanism for length and perhaps amino

Table 2 Type-3 NAR is the predominant form in neonates: little-N region addition and high similarity in amino acid composition of CDR3

	F-strand AGAGCATA	N	*D1/D2* ATAACTTTA CTGTGGATAG	N	*D3* CTGAG CTGGAGA	N	*J* ACAGCTTT GATGAATAC	Translation
a. Tm								
182	agagc	ca	aactttactgtgga	g	ctgagctgg		acagctttgatgaatat	RAKLYCGAELDSFDEY
176	agagc	ttt	actttactgtgga	g	ctgagctgg		acagctttgatgaatat	RALLYCGAELDSFDEY
174	agagcata	t	ctttactgtgga	g	ctgagctgg		acagctttgatgaatat	RAYLYCGAELDSFDEY
180	agagcat	c	actttactgtgga	g	ctgagctgg		acagctttgatgaatat	RASLYCGAELDSFDEY
171	agagca	gcc	ctttactgtgga	g	ctgagctgg		acagctttgatgaatat	RAALYCGAELDSFDEY
175	agagca	cg	actttactgtgg	cg	ctgagctgg		acagctttgatgaatat	RARLYCGAELDSFDEY
177	agagc	tc	aactttactgtgga	g	ctgagctgg		acagctttgatgaatat	RAQLYCGAELDSFDEY
170	agagc	gg	aactttactgtggat		ctgagct	ct	acagctttgatgaatat	RAELYCGSELYSFDEY
1648	agagc	gg	aactttactgtggat		ctgagct	ct	acagctttgatgaatat	RAELYCGSELYSFDEY
1652	agagcata	t	ctttactgtgga	g	ctgagctgga	g	agctttgatgaatat	RAYLYCGAELESFDEY
b. Secretory								
290	agagca	c	aactttactgtggat		ctgagctgga		cagctttgatgaatat	RAQLYCGSELDSFDEY
282	agagcata	t	ctttactgtggat		ctgagctgga		cagctttgatgaatat	RAYLYCGSELDSFDEY
280	agagca	c	aactttactgtgga	g	ctgagctgga		cagctttgatgaatat	RAQLYCGAELDSFDEY
281	agagc	cc	aactttactgtgga	g	ctgagctgga		cagctttgatgaatat	RAQLYCGAELDSFDEY
294	agagc	cctc	ctttactgtgga	g	ctgagctgga		cagctttgatgaatat	RALLYCGAELDSFDEY
296	agag	tcc	aactttactgtggat		ctgagctgga		cagctttgatgaatat	RVQLYCGSELDSFDEY
286	agagc	cc	aactttactgtgga	cggat	gctgga		cagctttgatgaatat	RAQLYCGRMLDSFDEY
279	agagc	ttt	actttactgtgg	gt	ctgagctgga		cagctttgatgaatat	RALLYCGSELDSFDEY
291	agagc	tttt	ctttactgtgga	g	ctgagctgga		cagctttgatgaatat	RAFLYCGAELDSFDEY
289	agagcata	t	ctttactgtggat	c	gacgctgga		cagctttgatgaatat	RAYLYCGSTLDSFDEY
293	agagc	ctcc	ctttactgtggat		ctgagctgga		cagctttgatgaatat	RASLYCGSELDSFDEY
283	agagc	cg	aactttactgtgg	ccccc	agct	cca	cagctttgatgaatat	RAELYCGPQLHSFDEY
287	agagca	cg	actttactgtgg	cg	ctgagctgga		cagctttgatgaatat	RARLYCGAELDSFDEY
292	agagcata	t	ctttactgtgga	g	ctgagctgga		cagctttgatgaatat	RAYLYCGAELDSFDEY
1639	agagc	cc	aactttactgtgga	g	ctgagctgg		acagctttgatgaatat	RAQLYCGAELDSFDEY
1640	agagcat	gg	ctttactgtgga	g	ctgagctgga	g	agctttgatgaatat	RAWLYCGAELESFDEY
1642	agagc	gtt	actttactgtgga	g	ctgagctgg		acagctttgatgaatat	RALLYCGAELESFDEY
1726	agagcatat		ctttactgtgga	g	ctgagctgg		acagctttgatgaatat	RAYLYCGAELDSFDEY
1728	agagcatat		ctttactgtgga	g	ctgagct	tg	acagctttgatgaatat	RAYLYCGAELDSFDEY
1644	agagc	cg	aactttactgtgga	g	ctgagctgga	t	agctttgatgaatat	RAELYCGAELDSFDEY
1643	agagc	cc	aactttactgtgga	g	ctgagctgga	g	agctttgatgaatat	RAQLYCGAELESFDEY
182	agagc	ca	aactttactgtgga	g	ctgagctgg		acagctttgatgaatat	RAKLYCGAELDSFDEY

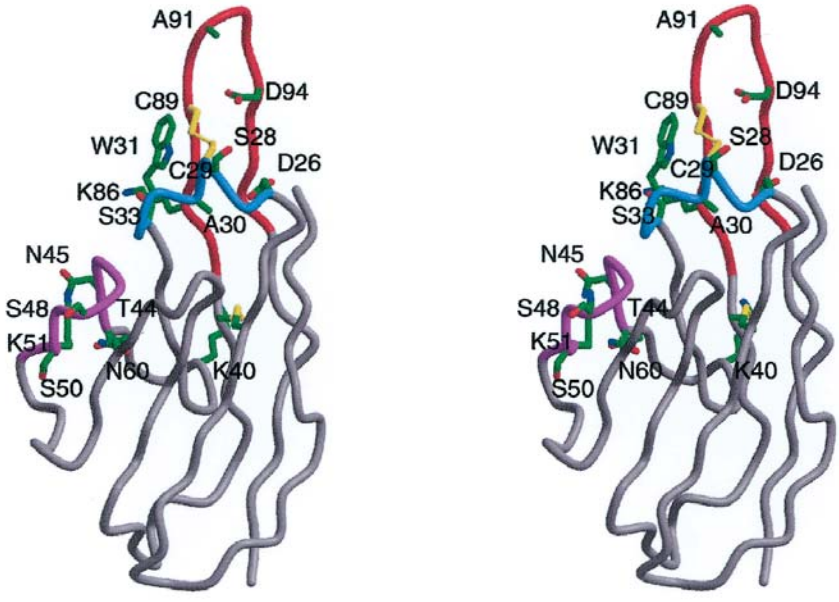


Fig. 8 Model of type-3 NAR. The sequence used was no. 182 from Table 2; alignment of the deduced amino acid sequence with types 1 and 2 and other V domains is shown in Fig. 4a

acid composition is operative in neonatal shark type-3 NAR. Its restricted expression in development and its limited diversity suggest that type-3 NAR may yield an evolutionarily selected specificity, and/or may play a role in the development of the adult NAR repertoire.

Discussion

To examine the source of diversity of the primary NAR repertoire in neonatal and young sharks, we analyzed over 200 NAR cDNA clones for evidence of somatic hypermutation of the entire V region and N-region additions in rearrangement junctions. We detected very low levels of mutation in shark pups that were characterized by insignificant CDR clustering, but that had most characteristics of NAR hypermutation, such as a preponderance of base substitutions, identical hotspots and a transition bias (Diaz et al. 1999). We previously detected a low mutation frequency in adult Tm clones (Diaz et al. 1988), and this result, combined with analysis of the neonatal clones, is in sharp contrast with the highly-mutated and CDR-targeted adult secretory NAR clones. A low mutation frequency, with no hint of positive selection, suggests that NAR hypermutation is "leaky," likely occurring during lymphocyte development. Most mutations in neonates were nonetheless obtained from productively rearranged clones, and since such clones probably originated from a pool of cells having functional surface receptors, the "leaky mutation" likely occurs after selection for cells with productive NAR rearrangements. We propose that this leaky mutation occurs during a period of division following the emergence of NAR^+ cells, perhaps akin to the period of 5–6 cell divisions that precedes L chain rearrangement in mouse B-cell development (Hardy and Hayakawa 2001). Perhaps transcription factors, important both early in lymphocyte development and following antigen activation, target mutational machinery that is fully functional in mature antigen-activated cells, but only partially active during early development of NAR^+ cells.

The low mutation frequency in shark pup NAR secretory clones may result from a poorly developed lymphoid microstructure in the spleen of young sharks (Rumfelt et al., submitted). T-cell zones are not found in the only shark secondary lymphoid tissue, the spleen, until five months after birth, concomitant with the appearance of many NAR transmembrane-positive cells in B-cell zones. The low number of NAR secretory cells in young sharks may have been stimulated in a T-independent manner, which failed to engage the full-blown hypermutation mechanism operative in adults. The low frequency of mutations in young animals is not limited to the *NAR* loci; few mutations also were detected recently in a study of expressed nurse shark germline-joined Ig light chains (Lee et al. 2002).

We detected significant N-region additions in neonatal type-1 and -2 NAR clones, the predominantly expressed NAR forms in adult sharks. This is in contrast to murine and *Xenopus* neonatal (larval) B cells, where the lack of expression of TdT early in development results in an absence of N regions and limited diversity (Flajnik and Rumfelt 2000b). These N regions, combined with the use of all three D elements in most rearrangement events, yield long and highly diverse CDR3s, consistent with the proposal that an NAR primary repertoire is utterly dependent upon CDR3 (Diaz et al. 1998). As mentioned, studies of skate Ig also showed N-region diversity from the earliest stages in which rearrangements occur (Miracle et al. 2001).

While the NAR primary repertoire in shark neonates and adults is entirely CDR3-based, there are structural constraints limiting CDR3 amino acid composition. CDR3-encoded Cys residues were found to be under strong selection in both adult and young sharks, and tertiary structure modeling predicts that these CDR3-encoded cysteines form disulfide bridges with the germline-encoded non-canonical cysteines in CDR1 for type-2 and FR2 for type-1 NAR. Like camelid species that also have antibodies with single-chain V domains, these atypical disulfide bridges must be critical for the stabilization of NAR proteins. Our models also predict that the non-canonical disulfide bridges formed in type-1 NAR place CDR2 in close proximity to CDR3, while for type-2 NAR, CDR1 and CDR3 are predicted to form a contiguous surface. The striking result of having many replacement substitutions generated by the somatic hypermutation mechanism in the CDR predicted to border CDR3, suggests that NAR non-canonical Cys residues profoundly influence the structure of the antigen binding surface. When four cysteines are present in a type-1 CDR3, it is likely that two of the cysteines form a disulfide bridge within the CDR3 loop.

While our data suggest that N-region addition, but not somatic hypermutation, plays a critical role in the NAR repertoire of shark pups, we were surprised to find expression of a third *NAR* locus (type 3). The expression and selection of this form is different than NAR types 1 and 2, as there is little N-region addition, usage of two rather than three D elements, and the resulting CDR3s are nearly identical in length and amino acid composition. It is likely that the cluster organization of the *NAR* loci, as for all of the shark *Ig* loci, allows for a "chimeric" neonatal repertoire of NARs, with highly diverse CDR3 in some loci, but low CDR3 diversity characteristic of type-3 clones. The significance of the highly conserved type-3 NAR in the neonatal repertoire is not known, but it is possible that it may reflect an evolutionary-selected specificity or it could regulate the development of NAR^+ cells. It is unquestionable, however, that whatever the significance, a stringent selection mechanism (self ligand?) has evolved to ensure that type-3 NAR molecules are of certain CDR3 length and amino acid composition.

We have recently uncovered a form of IgM, called IgM_{1gj}, which is entirely germline-joined and expressed in a similar fashion as type-3 NAR (Rumfelt et al. 2001). Both NAR type 3 and IgM_{1gj} are found as the predominant forms of the specific antigen receptor in all tissues

early in development, but high expression is perpetuated only in the epigonal organ, a primary lymphoid tissue, during adult life. It is not known whether some cells producing these receptors are continually generated in the epigonal organ throughout life or rather the epigonal organ also serves as a repertoire for cells producing all types of secreted antigen receptors. There is evidence from the skate that partially/totally germline-joined receptors have a transcriptional advantage early in development, but are extinguished later in life (Miracle et al. 2001; also in nurse shark, Ellen Hsu, personal communication). We believe that such a transcriptional advantage permits selection of some IgM/NAR clusters with expression of 'advantageous' receptors before the adaptive immune system kicks in.

Acknowledgements This work was supported by NIH grant RR06603.

References

Asarnow DM, Kuziel WA, Bonyhadi M, Tigelaar RE, Tucker PW, Allison JP (1988) Limited diversity of gamma/delta antigen receptor genes of Thy-1$^+$ dendritic epidermal cells. Cell 55:837–847

Bogue M, Candeias S, Benoist C, Mathis D (1991) A special repertoire of alpha:beta T cells in neonatal mice. EMBO J 10:3647–3654

Bork P, Holm L, Sander C (1994) The Immunoglobulin fold. Structural classification, sequence patterns and common core. J Mol Biol 242:309–320

Brady RL, Edwards DJ, Hubbard RE, Jiang J-S, Lange G, Roberts SM, Todd RJ, Adair JR, Emtage JS, King DJ, Low DC (1992) Crystal structure of a chimeric Fab' fragment of an antibody binding tumor cells. J Mol Biol 227:253–264

Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve NS, Jiang J-S, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography and NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D54:905–921

Desmyter A, Transue TR, Arbabi Ghahroudi M, Dao-Thi M-H, Poortmans F, Hamers R, Muyldermans S, Wyns L (1996) Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. Nat Struct Biol 3:803–811

Diaz M, Greenberg AS, Flajnik MF (1998) Somatic hypermutation of the new antigen receptor (NAR) gene in the nurse shark does not generate the repertoire: Possible role in antigen-driven reactions in the absence of germinal centers. Proc Natl Acad Sci USA 95:14343–14348

Diaz M, Velez J, Singh M, Cerny J, Flajnik MF (1999) Mutational pattern of the nurse shark antigen receptor (NAR) gene is similar to that of mammalian Ig genes and to spontaneous mutations in evolution: the translesion synthesis model of somatic mutation. Int Immunol 11:825–833

Du Pasquier L, Wilson M, Greenberg AS, Flajnik MF (1998) Somatic mutation in ectothermic vertebrates: Musings on selection and origins. Curr Top Microbiol Immunol 229:199–216

Feeney AJ (1990) Lack of N-regions in fetal and neonatal mouse immunoglobulin V-D-J junctional sequences. J Exp Med 172:1377–1390

Flajnik MF, Rumfelt LL (2000a) The immune system of cartilaginous fish. Curr Top Microbiol Immunol 248:249–270

Flajnik MF, Rumfelt LL (2000b) Early and natural antibodies in non-mammalian vertebrates. Curr Top Microbiol Immunol 252:233–240

Greenberg AS, Avila D, Hughes A, McKinney EC, Flajnik MF (1995) A new antigen receptor gene family that undergoes rearrangement and extensive somatic diversification in sharks. Nature 374:168–173

Guex N, Peitsch MC (1997) Swiss-model and the Swiss-PDB viewer: an environment for comparative protein modelling electrophoresis. Electrophoresis18:2714–2723

Hardy RR, Hayakawa K (2001) B cell development pathways. Annu Rev Immunol 19:595–621

Kearney JF, Won WJ, Benedict C, Moratz C, Zimmer P, Oliver A, Martin F, Shu F (1997) B cell development in mice. Int Rev Immunol 15:207–241

Kokubu F, Litman R, Shamblott MJ, Hinds K, Litman GW (1988) Diverse organization of immunoglobulin VH gene loci in a primitive vertebrate. EMBO J 7:3413–3422

Kormanik GA (1988) Time course of the establishment of uterine seawater conditions in late-term pregnant spiny dogfish (*Squalus acanthias*) J Exp Zool 137:443–456

Kraulis PJ (1991) A program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 24:946–950

Lee SS, Tranchina D, Ohta Y, Flajnik MF, Hsu E (2002) Hypermutation in shark immunoglobulin light chain gene results in contiguous substitutions. Immunity 16:571–582

Litman GW, Anderson MK, Rast JP (1999) Evolution of antigen-binding receptors. Annu Rev Biochem 17:109–147

Merritt EA, Bacon DJ (1997) Raster 3D: Photorealistic Molecular Graphics. Methods Enzymol 277:505–524

Miracle AL, Anderson MK, Litman RT, Walsh CJ, Luer CA, Rothenberg EV, Litman GW (2001) Complex expression patterns of lymphocyte-specific genes during the development of cartilaginous fish implicate unique lymphoid tissues in generating an immune repertoire. Int Immunol 13:567–580

Nguyen VK, Desmyter A, Muyldermans S (2001) Functional heavy-chain antibodies in Camelidae. Adv Immunol 73:261–296

Nuttall SD, Krishnan UV, Hattarki M, De Gori R, Irving RA, Hudson PJ (2001) Isolation of the new antigen receptor from wobbegong sharks, and use as a scaffold for the display of protein loop libraries. Mol Immunol 38:313–326

Nuttall SD, Krishnan UV, Doughty L, Nathanielsz A, Ally N, Pike RN, Hudson PJ, Kortt AA, Irving RA (2002) A naturally occurring NAR variable domain binds the Kgp protease from *Porphyromonas gingivalis*. FEBS Lett 516:80–86

Ohta Y, Okamura K, McKinney EC, Bartl S, Hashimoto K, Flajnik MF (2000) Primitive synteny of vertebrate major histocompatibility complex class I and class II genes. Proc Natl Acad Sci USA 25:4712–4717

Peitsch MC (1996) PROMOD and Swiss-Model: Internet-based tools for automated comparative protein modelling. Biochem Soc Trans 24:274–279

Roux KH, Greenberg AS, Greene L, Strelets LA, Avila D, McKinney EC, Flajnik MF (1998) Structural analysis of the nurse shark (new) antigen receptor (MAR): molecular convergence of NAR and unusual immunoglobulins. Proc Natl Acad Sci USA 95:11804–11809

Rumfelt LL, Avila D, Diaz M, Bartl S, McKinney EC, Flajnik MF (2001) A shark antibody heavy chain encoded by a nonsomatically rearranged VDJ is preferentially expressed in early development and is convergent with mammalian IgG. Proc Natl Acad Sci USA 98:1775–1780

Smith RF, Wiese BA, Wojzynski MK, Davidson DB, Worley KC (1996) BCM Search Launcher: an integrated interface to molecular biology database search and analysis services available on the World Wide Web. Genome Res 6:454–462

Sokol RL, Rohlf FJ (1981) Biometry, 2nd edn. Freeman, New York, pp 691–720

Thompson JD, Higgins DG, Gibson TJ (1994) Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Wagner SD, Neuberger MS (1996) Somatic hypermutation of immunoglobulin genes. Annu Rev Immunol 14:441–457

Williams AF, Barclay AN (1988) The immunoglobulin superfamily – domains for cell surface recognition. Annu Rev Immunol 6:381–405

Xu JL, Davis MM (2000) Diversity in the CDR3 of V(H) is sufficient for most antibody specificities. Immunity 13:37–45