Developmental and Comparative Immunology 37 (2012) 176–183

Contents lists available at SciVerse ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

Somatic hypermutation of TCR γ V genes in the sandbar shark

Hao Chen [a], Harris Bernstein [b], Parvathi Ranganathan [a], Samuel F. Schluter [a,*]

[a] Department of Immunobiology, College of Medicine, University of Arizona, Tucson, AZ 85724, USA
[b] Department of Cell Biology and Anatomy, College of Medicine, University of Arizona, Tucson, AZ 85724, USA

ARTICLE INFO

Article history:
Received 4 August 2011
Revised 29 August 2011
Accepted 31 August 2011
Available online 8 September 2011

Keywords:
T cell receptor
Somatic hypermutation
Adaptive immune system
Evolution
Shark

ABSTRACT

In a recent publication we demonstrated that somatic hypermutation occurs in the V region of the TCR γ gene of the sandbar shark (*Carcharhinus plumbeus*). We hypothesize that similar mechanisms are used to generate somatic mutations in both immunoglobulin and TCR γ genes of the sharks. Two distinct patterns of mutation occur, single nucleotide mutations (point mutations) and mutations comprising 2–5 consecutive bases (tandem mutations). Our data indicates that point mutations occur by a mechanism similar to that of somatic hypermutation in immunoglobulin genes of mammals, whereas tandem mutations may be generated by an error-prone DNA polymerase with terminal deoxynucleotidyl transferase (TdT)-like activity. Shark hotspot motifs identical to those of higher vertebrates were identified. We confirm that, as in immunoglobulin of sharks and higher vertebrates, highly significant targeting of AID activity to the classical DGYW/WRCH motif occurs in somatic hypermutation of sandbar shark TCR γ V genes. Our analysis suggests that the purpose of somatic mutations in shark TCR γ V-regions is to generate a more diverse repertoire in γ/δ receptors, rather than receptors with higher affinity.

1. Introduction

Somatic hypermutation (SHM) in mammals is responsible for generating antibody populations that have higher affinity to antigens, a process called affinity maturation. In the current model, activation induced cytidine deaminase (AID) has been shown to be the key player for SHM. The first step is AID mediated conversion of cytidine (C) to uridine (U). A basic characteristic of AID activity in mammals is targeting of C residues in DGYW/WRCH "hot spot" motifs (Dorner et al., 1997). For C/G base pairs, a mutation is generated via base excision repair (BER) in which uracil DNA glycosylase (UNG) removes the U base. UNG mediated cleavage of the U base from the DNA backbone results in a non-instructive abasic site. Next, error prone polymerases are recruited to the abasic site to replace the C in the original C/G base pairs. Mutations at A/T base pairs are generated by mismatch repair (MMR) reactions. The recruitment of error-prone polymerase η (Delbos et al., 2007) to the DNA lesion leads to mutations in the excised region.

Although T cell receptors are structurally similar to immunoglobulins (Igs) and utilize similar mechanisms to generate receptor diversity, they are believed to be incapable of SHM (Vitetta et al., 1991). We have recently shown that SHM occurs in sandbar shark TCR γ V regions, generating single nucleotide mutations (point mutations) and mutations comprising 2–5 consecutive mutated bases (tandem mutations). SHM has also been found in nurse shark T cell receptors α and γ (Criscitiello et al., 2010), and recently in camel T cell receptor δ genes (Antonacci et al., 2011). Here we perform a comprehensive study of SHM in TCR γ V regions and conclude that sandbar shark TCR γ V regions and Ig genes of nurse shark have similar mutation patterns, supporting our hypothesis that shark B cells and γ/δ T cells utilize similar mutation mechanisms. In addition, analysis of point mutations indicates that sharks and higher vertebrates share similar mechanisms of SHM. Analysis of the tandem mutations, which are unique to sharks, suggests that an error-prone DNA polymerase with terminal deoxynucleotidyl transferase (TdT)-like activity may be responsible for the generation of these unique mutations. Furthermore, it appears that the purpose of somatic mutations in shark TCR γ V-regions is to generate a more diverse repertoire in γ/δ receptors, rather than receptors with higher affinity.

2. Materials and methods

2.1. Animal

Sandbar shark spleen was provided by Dr. Carl Luer (Mote Marine Laboratories, Sarasota, FL). All the genomic DNA and mRNA were collected from a single animal.

2.2. Genomic and cDNA sequences of sandbar shark TCR γ gene

In a previous report (Chen et al., 2009), we sequenced the genomic TCR γ locus of sandbar shark (Accession No. FJ854492) and a

* Corresponding author. Address: Department of Immunobiology, College of Medicine, 1501 N Campbell Ave., P.O. Box 245221, Tucson, AZ 85724, USA.
E-mail address: schluter@u.arizona.edu (S.F. Schluter).

0145-305X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2011.08.018

total of 73 TCR γ V region cDNA clones (Accession Nos. FJ854417–FJ854491). cDNA sequences were obtained using the 5′ Rapid amplification of cDNA ends (RACE) technique. Three C region specific primers were used to rule out potential bias in amplification of V regions.

2.3. Germline DNA and cDNA sequences alignment

Alignments were performed using Mac Vector 11 (MacVector, Inc.) and MAFFT (Katoh et al., 2002) software.

2.4. Calculation of mutability indices

The mutability index is the observed number of mutations of a specific nucleotide divided by the expected number of mutations of that nucleotide. The expected number of mutations was derived by determining the frequency of the nucleotide within the sequenced TCRGV cDNA database multiplied by the total number of observed mutations within the database. A mutability index value of 1.00 would indicate the effects of random mutations.

2.5. Determination of hotspots

We used published sequences of nurse shark light chains (Lee et al., 2002) to identify hotspot motifs. The sequence database contained 81 mutated cDNA light chain V region sequences transcribed from two genomic loci. As the genomic sequences differed by only four nucleotides and no mutations occurred at these positions, the sequences were combined for this analysis. Hotspots were derived using program CLUSTERM (Katoh et al., 2002; Rogozin et al., 2001), which calculates a threshold by analyzing the frequency distribution derived from a mutation spectrum. Sites in the genomic sequence were identified as hotspots if the number of cDNA sequences with a mutation at the corresponding position was equal or greater than the threshold.

2.6. Statistical analysis

Except as noted, statistical analyses were performed using 2×2 contingency tables. Probability values were calculated using the Fisher exact test or χ^2 analysis. The student t test was used for the comparison of the mutation frequency among individual domains. Probability (p) values < 0.05 were considered statistically significant. Analyses were performed using SAS (SAS Institute, Inc.) or Prism 5 (GraphPad Software, Inc.).

2.7. Assessment of antigen selection/affinity maturation from mutation profiles

The presence of antigen selection and affinity maturation was determined by comparing the R/S ratios between CDR and FR regions using $\chi 2$ analysis. The focused binomial test (Hershberg et al., 2008), a new robust method that takes into account many systemic errors, was also used.

3. Results

3.1. Somatic hypermutation in sandbar shark TCR γ V genes

Despite the low identity between human and shark TCR γ V genes, 11 amino acid sites are conserved throughout evolution (Fig. 1A), and these enabled us to accurately identify the CDRs and FRs in the shark V genes. The amino acid sequences of individual V genes share less than 50% identity with each other, and there is a major difference in the size of CDR1 (5–9 amino acids) and CDR2 (6–8 amino acids) among individual V genes. Alignment of the cDNA sequences with their parental genomic DNA sequence allowed us to analyze the mutation profiles of the sandbar shark TCR γ V genes. Almost half (33) of the 73 cDNA clones contain at least one mutation in their V regions (14 GV1 clones, 5 GV2 clones, 6 GV3 clones, 7 GV4 clones, and 1 GV5 clone), and 22 clones contain two or more mutations. Although somatic mutations occur in the CDR3 region, these are not included in this report since they are extremely hard to analyze due to extensive deletions and additions at the VJ junctions (Chen et al., 2009). Only sequences containing mutations are included in the following analyses, and GV5 sequences are excluded due their limited number. The alignment of amino acid sequences with their parental genomic sequence is shown in Fig. 1B (for GV1) and Supplementary Figures S1 (for GV2–4). The amino acid changes resulting from tandem mutations are underlined. The mutation rate considering only sequences that contained mutations was 0.018/bp, which is comparable to the mutation rate in Ig genes of mice (0.016/bp) and sharks (0.015/bp) (Lee et al., 2002). However, since the mutated sequences comprised about half of the total sequences, the overall mutation rate could be considered half this value. As also seen in shark Ig and the new antigen receptor (NAR) genes, two different patterns of mutation were identified in TCR γ; namely, point mutations and tandem mutations.

3.2. Single nucleotide mutation

Detailed analyses of point mutations and tandem mutation are shown in Table 1 and Supplementary Tables SI and SII. A total of 106 point mutations were identified, with a bias towards G and C bases. Overall, 64.2% of all point mutations were at G/C nucleotides (χ^2, $p = 0.011$), varying from 51.2% to 84.2% among individual V genes. Mutation indices (MI) of A, C, G, T are 0.76, 1.70, 1.16 and 0.59, respectively, with mutations of C bases significantly higher and mutations of T bases significantly lower than the expected number if there were no bias. In addition, a strong bias towards transition mutations was found. About 53.8% of all point mutations are transitions, which is significantly higher ($\chi 2$, $p = 0.002$) than the expected frequency of 33%.

3.3. Tandem mutation

In our analysis of TCR γ V genes, a total of 33 tandem mutations (involving 85 nucleotides) were found in 12 of the 33 cDNA clones that contained mutations. Tandem mutations comprised 44.5% of all mutations and varied in size from 2 to 5 nucleotides. No GC bias is present in the tandem mutations. Mutations at GC bases comprise only 45.9% of all the nucleotides in tandem mutations, varying from 42.9% to 50% among V genes. Second, in sharp contrast to a universal transition preference in SHM of Igs (Hinds-Frey et al., 1993; Lee et al., 2002) that also applies to NAR (Diaz et al., 1999), and TCR point mutations of sharks, no transition preference was found in tandem mutations of shark TCR γ V regions in this study. Overall, the transition to transversion ratio of tandem mutations is significantly lower than that of point mutations ($\chi 2$, $p = 0.008$). A total of 33 tandem mutations led to 43 amino acids replacements, or 1.3 replacements per tandem on average. Therefore, the replacement efficiency of tandem mutations is significantly higher than that of point mutations ($\chi 2$, $p = 0.001$). In the tandem mutations of TCR γ V genes, 18 tandem mutations (54.5%) contain di- and tri-nucleotide repeats (underlined in Table 2). Similar di-and tri-nucleotide repeat patterns occur in SHM of shark Ig light and heavy chain genes (Lee et al., 2002; Malecek et al., 2005). In comparison to their parental genomic sequences, the sequences of tandem mutations appear to be random.