Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Immunology 42 (2005) 1243–1249
Molecular Immunology
www.elsevier.com/locate/molimm

Short communication

Analysis of recombination signal sequences in zebrafish

Nadia Danilova*

Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Received 27 October 2004
Available online 5 January 2005

Abstract

Recombination signal sequences (RSS) from immunoglobulin and TCRα genes of zebrafish were analyzed in comparison with RSS from human and species-specific features were revealed. In contrast to human RSS, in zebrafish RSS from both V_H and TCRα genes the last nonamer position is not conserved. On the contrary, the fourth nonamer position, which is not conserved in human or mouse is conserved in zebrafish. The 12 bp spacers from human and zebrafish RSS contain 9 bp motif resembling nonamer sequence. Spacers in zebrafish 23 bp RSS from both immunoglobulins and TCRα contain 7 bp motif also resembling nonamer sequence while corresponding human sequences do not contain analogous motif. RSS are recognized by RAG1 protein, which also has specific features in teleost suggesting co-evolution of RAG1 with corresponding RSS.

Keywords: RAG1; RSS; Immunoglobulin; V_H; TCRα

1. Introduction

Receptors of the adaptive immune system are encoded in genome as multiple segments: variable (V), diversity (D), joining (J) and constant (C). Functional genes are created by recombination of these gene segments generating a virtually unlimited variety of immune receptors. Recombination activating proteins encoded by *RAG*1 and *RAG*2 genes perform critical DNA recognition and cleavage functions in V(D)J recombination (Gellert, 2002). The recombination is targeted by specific DNA sequences, known as recombination signal sequences (RSSs), which flank the coding segments (Fig. 1). RSS consists of a conserved heptamer (consensus 5′-CACAGTG) and nonamer (consensus 5′-ACAAAAACC) separated by a moderately conserved spacer of either 12 or 23 base pairs. The presence of RSS is both necessary and sufficient to direct RAG mediated recombination on artificial substrates.

V(D)J recombination is strongly regulated in a cell-, locus- and stage-specific manner (Swanson, 2004). It takes place only in lymphoid cells. B cells rearrange only immunoglobulin genes while T cells rearrange only T cell receptors. In both types of cells rearrangement takes place only at specific stages of cell development. Recombination follows the 12/23 rule, which dictates that it occurs between two RSSs with different spacer lengths.

At the first phase of recombination the RAG proteins assemble on RSS and nick the DNA at the border between the RSS heptamer and the gene segment. The resulting 3′-hydroxyl group on the nicked strand performs a nucleophilic attack on the opposite strand of DNA forming a phosphodiester bond and hence creating a DNA hairpin at the terminal end of the coding sequence. In the next step, the 12 and 23 signal ends are joined heptamer to heptamer in a precise junction. In contrast, joining of coding ends is imprecise due to addition or deletion of nucleotides before ligation, which serves to further diversify the repertoire of antigen specific receptors. Many other proteins cooperate with RAGs in the recombination process (Gellert, 2002).

The nature of recognition of RSS by RAG proteins remains poorly understood. Although both RAG1 and RAG2 are necessary for recombination, sequence-specific DNA binding sites were found only in RAG1 (Spanopoulou et al., 1996).

* Tel.: +1 617 253 6705; fax: +1 617 253 8699.
E-mail address: ndanilov@mit.edu.

0161-5890/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.molimm.2004.11.022

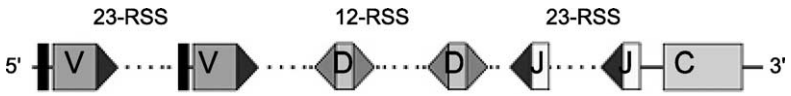

Fig. 1. RSS flank the coding segments of immunoglobulin and TCR genes. Components of the immunoglobulin heavy chain gene are shown. V_H segments have RSS with 23 bp spacer at the 3′end; D_H sequences have RSS with 12 bp spacer and J_H segments have 23-RSS.

Variations in RSS sequence are known to influence the efficiency of recombination. No RSSs have been found to recombine more efficiently than the consensus forms of the joining signals (Hesse et al., 1989; Ramsden and Wu, 1991). The heptamer as well as nonamer are thought to be conserved, especially the three heptamer bases closest to the recombination crossover site, and the fifth and sixth positions of the nonamer (Hesse et al., 1989; Ramsden and Wu, 1991). The role of other positions in hexamer and octamer as well as in spacer have been disputed; however, as recent studies suggest they may influence the recombination (Lee et al., 2003; Montalbano et al., 2003; Nadel et al., 1998; Olaru et al., 2003; Ramsden et al., 1994; Tillman et al., 2004). A program has been developed to predict RSS efficiency based on the whole RSS sequence (Cowell et al., 2002; Lee et al., 2003). This program was based on a database of mouse sequences (Cowell et al., 2003). It was applied successfully for identification of functional cryptic signals in human and mouse genomes (Cowell et al., 2003). However, another group applied this program in analysis of RSS of human Vκ genes and found a discrepancy between the prediction and experimental data: RSSs with close scores differed substantially in recombinogenic efficiency (Montalbano et al., 2003). These data suggest differences in RSS between species.

Previously, only a few attempts have been made to analyze the difference between RSS of different species (Glusman et al., 2001). It appears that RSS from different genes and/or species have been often pooled on the assumption that RSSs have the same pattern in all cases. This assumption might be based on the fact that the same protein, RAG1, recognizes RSS in all species. RAG1 is a very conserved protein, however, given that substitutions are present even in the core region of the RAG1 and that many RSS residues seem to be involved in binding, species-specific variations in RSS–RAG1 recognition site are possible. In support, some subtle differences have been noted in RSS at TCR Vα and Vβ genes from mouse and human (Glusman et al., 2001). Species-specific features may account for the inaccuracy of prediction of RSS recombinogenic efficiency in one species using model originally designed based on sequence motifs from another species.

The zebrafish is the first non-mammalian vertebrate whose T and B receptor genes' sequences have become available in numbers sufficient for statistical analysis as a result of zebrafish genome project. Analysis of RSS sequences from species phylogenetically distant from mammals may help to better understand the mechanisms of gene rearrangement. In this study, the analysis of RSS sequences of zebrafish in comparison with human RSS sequences indicates that RSS have both species-specific and gene-specific features.

2. Methods

2.1. Human RSS

There are 39 known expressed human V_H segments (accession number NT_026437). In addition, five putative functional V_H segments were found during genome sequencing. These 44 segments were used as a source of 23-RSS. The vast majority of human immunoglobulin heavy chain genes are pseudogenes and they had not been included in the analysis. There are 27 D_H gene segments in human genome (accession number NT_026437). They have 12-RSS on both sides; all those were included in the analysis.

2.2. Zebrafish RSS

From 47 V_H segments identified in zebrafish genome 33 have no defects and are presumed functional or expressed (Danilova et al., 2000) and (Danilova et al., 2004). RSS from these V_H segments were analyzed. Igζ and Igμ heavy chains have 2 and 5 D_H segments respectively flanked with 12-RSS. Since there were only 14 sequences from D_H segments available, they were combined with 20 sequences from light chains (Haire et al., 2000). A substantial part of zebrafish TCRα locus has been sequenced and annotated. We examined 23-RSS from 54 functional TCR Vα segments (accession numbers AL591481 and AL591399).

2.3. Sequence logos

To analyze RSS we created logos of sequences where the frequency of each nucleotide at a particular position is presented graphically (Schneider and Stephens, 1990). The height of each letter is proportional to its frequency and the letters are sorted so the most common one is on top. Maximal size, 2 bits, corresponds to 100% frequency of a given base at a given position.

RSSs were aligned and the sequence logo was created using program Delila distributed by Cambridge University by the authors (Schneider and Stephens, 1990). Heptamer and nonamer sequences were highlighted using Adobe Photoshop.