Developmental and Comparative Immunology 81 (2018) 105–115

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

The evolutionary conservation of the bidirectional activity of the *NWC* gene promoter in jawed vertebrates and the domestication of the *RAG* transposon

Lukasz Sniezewski [a], Sylwia Janik [a], Agnieszka Laszkiewicz [a], Michal Majkowski [a], Pawel Kisielow [a, b], Malgorzata Cebrat [a, *]

[a] *Laboratory of Molecular and Cellular Immunology, Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Weigla 12, 53-114 Wroclaw, Poland*

[b] *Laboratory of Tumor Immunology, Hirszfeld Institute of Immunology and Experimental Therapy, Polish Academy of Sciences, Weigla 12, 53-114 Wroclaw, Poland*

ARTICLE INFO

Article history:
Received 8 August 2017
Received in revised form
21 November 2017
Accepted 21 November 2017
Available online 22 November 2017

Keywords:
RAG transposon
NWC gene
Evolution of the adaptive immune system
Bidirectional promoter

ABSTRACT

The *RAG-1* and *RAG-2* genes form a recombinase complex that is indispensable for V(D)J recombination, which generates the diversity of immunoglobulins and T-cell receptors. It is widely accepted that the presence of *RAGs* in the genomes of jawed vertebrates and other lineages is a result of the horizontal transfer of a mobile genetic element. While a substantial amount of evidence has been gathered that clarifies the nature of the *RAG* transposon, far less attention has been paid to the genomic site of its integration in various host organisms. In all genomes of the jawed vertebrates that have been studied to date, the *RAG* genes are located in close proximity to the *NWC* gene. We have previously shown that the promoter of the murine *NWC* genes exhibits a bidirectional activity, which may have facilitated the integration and survival of the *RAG* transposon in the host genome. In this study, we characterise the promoters of the *NWC* homologues that are present in the representatives of other jawed vertebrates (*H. sapiens*, *X. tropicalis* and *D. rerio*). We show that the features that are characteristic for promoters as the hosts of a successful transposon integration (in terms of the arrangement, bidirectional and constitutive activity and the involvement of the Zfp143 transcription factor in the promoter regulation) are evolutionarily conserved, which indicates that the presence of *RAG* genes in jawed vertebrates is a direct result of a successful transposon integration into the *NWC* locus.

1. Introduction

RAG-1/*RAG-2* genes encode the proteins that form the recombinase complex, which is indispensable in the process of V(D)J recombination. The V(D)J recombination is a mechanism that generates a vast diversity of immunoglobulin and T-cell receptors, and thus it forms the basis for the functioning of an adaptive immune system that is characteristic for jawed vertebrates (Mombaerts et al., 1992; Oettinger et al., 1990; Schatz et al., 1989; Shinkai et al., 1992). It is widely accepted that *RAG* genes have emerged through the integration of a mobile genetic element (Fugmann, 2010; Thompson, 1995); however, the exact nature of such a transposon and the moment of its integration with the genome(s) of the host(s) organisms has been the subject of long and intensive research. The transposonal origin of *RAG* genes is supported by a considerable amount of crucial evidence concerning their function and composition that has emerged since their discovery. Similarities between the RAG-mediated processes and a cut-and-paste transposition include: a) similarities between the terminal inverted repeats (TIRs) and the recognition signal sequences (RSSs) that flank the immunoglobulins and T-cell receptor minigenes; b) the mechanism of the DNA cleavage reactions; and c) the compact structure of the *RAG* genes and their locus (Lieber, 2010; McBlane et al., 1995; Ramsden et al., 1994; Rooney et al., 2004; van Gent et al., 1996). Furthermore, the potential precursor

Abbreviations: TSS, transcription start site; RSS, recognition signal sequences; *hsNWC*, *Homo sapiens NWC* gene homologue; *xtNWC*, *Xenopus tropicalis NWC* gene homologue; *drNWC*, *Danio rerio NWC* homologue; DBD, DNA binding domain; TE, transposable element; DLR, Dual Luciferase Reporter Assay.

* Corresponding author.
E-mail address: cebrat@iitd.pan.wroc.pl (M. Cebrat).

https://doi.org/10.1016/j.dci.2017.11.013
0145-305X/

elements for modern *RAGs* have been identified as the transposases of the *Transib* family, which share a sequence that is similar to the core region of *RAG-1*, and show striking similarities in the transposition process to the one controlled by the *RAGs* (Kapitonov and Jurka, 2005). However, unlike the *RAG* genes, the Transib transposons contain only the *RAG-1*-like gene. Thus, the first step of the assembly of the *RAG* transposon would have to be the acquisition of a *RAG-2*-like gene by a *Transib* element. The origin of the *RAG-2* remains unknown, but the ability of the modern RAG-2 protein to interact with the Transib transposases suggests that the ancestral RAG-2 protein could also have interacted with the ancestral Transib protein and potentially enhanced its activity (Fugmann, 2010). In addition to jawed vertebrates, the *RAG-1/2* locus has been found in echinoderms (*Strongylocentrotus purpuratus*, *Lytechinus variegatus*, *Patiria miniata*) and cephalochordates (*Branchiostoma belcheri*) (Fugmann et al., 2006; Huang et al., 2016; Kapitonov and Koonin, 2015). In both groups, the *RAG* locus resembles the one that is found in jawed vertebrates with the *RAG-1*-like and *RAG-2*-like genes located close to each other in a tail-to-tail orientation. Also, the *S. purpuratus RAG-1*-like and *RAG-2*-like genes are coordinately expressed, and the encoded proteins interact with each other. However, there is no evidence that the protein complex is able to mediate the DNA cleavage or transposition (Fugmann et al., 2006). In contrast to *S. purpuratus*, the lancelet *ProtoRAG* superfamily of transposons represents functional transposable elements that contain both *RAG-1*-like and *RAG-2*-like genes (Huang et al., 2016). The similarities between the *ProtoRAG* and *RAG-1/RAG-2* sequences and functions are extensive and are indicative for the evolutionary relationship between both complexes. More information concerning the evolutionary history of the *RAG* transposon can be found in a review by Carmona and Schatz (2017).

The research conducted so far has focused on the nature of *RAG* transposon itself. Far less attention has been paid to the integration site of the transposon in the genome, which constitutes the *NWC* gene locus in all jawed vertebrate genomes that have been subjected to an analysis so far. The *NWC* gene is a highly evolutionarily conserved gene (Cebrat et al., 2005) whose orthologues, in addition to those in vertebrates, can be found in numerous invertebrate species such as the sea urchin (*S. purpuratus*), starlet sea anemone (*Nematostella vectensis*), starfish (*Asterina pectinifera*), sea mussel (*Mytilus edulis*), sea snail (*Lottia gigantea*), stony coral (*Acropora millepora*) or the placozoan (*Trichoplax adhaerens*). In vertebrates, the *NWC* consists of 6–7 exons located upstream of the *RAG-2* gene and these two genes are convergently transcribed. The first, non-coding, exon of the *NWC* gene is located immediately upstream of the coding exon of the *RAG-2* gene (Cebrat et al., 2005). Previously, we characterised in detail the mechanisms that regulate the transcription of the murine *NWC* gene. The *NWC* gene transcription is regulated by the *NWC*'s own promoter in all non-lymphoid cells, and the level of transcription is the highest in the testis, which is the only organ that contains detectable levels of the NWC protein (Cebrat et al., 2008; Kasztura et al., 2016, 2009; Laszkiewicz et al., 2011). The minimal region of the *NWC* promoter consists of a 100 bp fragment located upstream of exon 1. The promoter lacks the TATA-box and Inr elements, but it is associated with a CpG island, and the *NWC* promoter is activated by the Zfp143 transcription factor which has two binding sites within the promoter (Laszkiewicz et al., 2012).

The integration of transposons into the host genome can be disadvantageous for the host, primarily due to the potential of the transposons to damage the host's genes during translocations within the genome (Orgel et al., 1980). Consequently, defence mechanisms have evolved that hamper the activity of the transposons, primarily by silencing their expression through epigenetic changes in the DNA structure (DNA methylation and/or histone modifications) (Lippman and Martienssen, 2004; O'Donnell and Boeke, 2007). Analyses of the relatively few cases in which the transposon-derived genes are still active in their hosts have shown that these transposons have integrated with the genome at specific locations. One such location is a DNA fragment that is located directly downstream of an existing gene. In such a case, later changes occur that create a hybrid gene which codes a protein with a modified function (when compared to that of the transposase and the original host protein). Another characteristic location is a head-to-head arrangement, whereby the transposon integrates itself upstream of an existing gene that is controlled by bidirectional promoters containing a CpG island (Kalitsis and Saffery, 2009). The open chromatin structure surrounding such promoter regions is thought to favour the integration and survival of the transposon-derived genes. The transcription of the transposon gene could be ensured by the activity of the host promoter and, most importantly, any attempt at silencing the transposon gene through the methylation of the CpG island of the bidirectional promoter would also hamper the expression of the host's gene. Furthermore, in contrast to the first scenario, the structure of the host's gene and the transposase gene is left intact, which allows both genes to perform their original functions. Because many CpG island-containing promoters control the housekeeping genes, their inactivation or any interference with their primary function could be fatal for the host organism (Anno et al., 2011; Trinklein et al., 2004). As was described above, the *NWC* gene and its promoter meet the criteria of a host gene for a transposon, as it is constitutively active and controlled by a CpG island (Cebrat et al., 2008; Laszkiewicz et al., 2012). Most importantly, we have also shown that in the mouse, the *NWC* promoter exhibits a bidirectional activity that drives a detectable transcription of the *RAG-2* coding exon in non-lymphoid tissues (Laszkiewicz et al., 2012). Therefore, it is possible that the integration and survival of the *RAG* transposon is facilitated by the bidirectional activity of the *NWC* promoter. To make this hypothesis acceptable, it would have to be shown that the bidirectional activity of the *NWC* promoter is an evolutionarily conserved feature. Thus, the aim of our study is to verify this hypothesis.

2. Materials and methods

All procedures using animals were reviewed and approved by the First Local Ethical Commission for Animal Experimentation, at the Institute of Immunology and Experimental Therapy (Wroclaw, Poland; permit number 81/2012).

2.1. RNA ligase-mediated rapid amplification of cDNA ends (5′-RACE)

For determination of the transcription initiation site, RNA ligase-mediated 5′-rapid amplification of cDNA ends (5′-RACE) was performed using the FirstChoice RLM-RACE kit (Ambion). After ligation of the 5′-RACE adapter, RNA was reverse-transcribed with a Random hexamer, oligo-dT or a *NWC* gene-specific primer (listed below). Nested PCR was performed using *NWC* gene-specific primers in conjunction with primers complementary to the 5′-RACE adapter. The resulting PCR products were cloned into the pGEM-T Easy vector (Promega) and subjected to sequencing. *RAG-2*-oriented 5′-RACE was performed on a RNA isolated from cells transfected with pGL3 vectors containing *RAG-2*-oriented promoter fragments. Firefly luciferase gene-specific primers were used on reverse transcription of those samples. Nested PCR was performed using firefly luciferase gene-specific primers in conjunction with primers complementary to the 5′-RACE adapter. The following primers were used:

NWC-oriented 5′-RACE: