Gene 508 (2012) 221–228

Contents lists available at SciVerse ScienceDirect

Gene

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

Short Communication

Under-representation of intrinsic terminators across bacterial genomic islands: Rho as a principal regulator of xenogenic DNA expression

Anirban Mitra [a], Valakunja Nagaraja [a,b,*]

[a] Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore, India
[b] Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India

ARTICLE INFO

Article history:
Accepted 30 July 2012
Available online 4 August 2012

Keywords:
Horizontal gene transfer
Intrinsic termination
Hairpin
RNA polymerase
Pathogenicity

ABSTRACT

Two transcription termination mechanisms – intrinsic and Rho-dependent – have evolved in bacteria. The Rho factor occurs in most bacterial lineages, and has been hypothesized to play a global regulatory role. Genome-wide studies using microarray, 2D-gel electrophoresis and ChIP-chip provided evidence that Rho serves to silence transcription from horizontally acquired genes and prophages in *Escherichia coli* K-12, implicating the factor to be a part of the "cellular immune mechanism" protecting against deleterious phages and aberrant gene expression from acquired xenogenic DNA. We have investigated this model by adopting an alternate *in silico* approach and have extended the study to other species. Our analysis shows that several genomic islands across diverse phyla have under-representation of intrinsic terminators, similar to that experimentally observed in *E. coli* K-12. This implies that Rho-dependent termination is the predominant process operational in these islands and that silencing of foreign DNA is a conserved function of Rho. From the present analysis, it is evident that horizontally acquired islands have lost intrinsic terminators to facilitate Rho-dependent termination. These results underscore the importance of Rho as a conserved, genome-wide sentinel that regulates potentially toxic xenogenic DNA.

1. Introduction

Transcription involves synthesis of RNA by RNA polymerase (RNAP) on a DNA template and is functionally divided into initiation, elongation and termination (von Hippel, 1998). The last step i.e. termination involves stopping of elongation, release of the RNA and dissociation of the RNAP machinery (Richardson and Greenblatt, 1996). In bacteria, termination functions by two mechanisms — intrinsic and factor-dependent (Peters et al., 2011; Santangelo and Artsimovitch, 2011). At intrinsic terminators (ITs), termination is effected by the sequence and structural features of the hairpin and the U-trail of the nascent RNA (Epshtein et al., 2007). In contrast, factor-dependent termination predominantly involves the Rho protein which shows little preference for any specific sequence or structure on the RNA and the template DNA for its activity (Ciampi, 2006; Richardson, 2002). Rho seems to be the major termination factor for genes that do not have an IT downstream. It has been speculated that several genes are likely targets for Rho *in vivo*, although only few have been characterized (Ciampi, 2006). Historically, termination has received relatively lesser attention than the first two steps of transcription. In the recent post-genomics era its regulatory importance in the context of the whole cell is being understood (Cardinale et al., 2008; Peters et al., 2009).

An outcome of the large-scale sequencing and annotation of genomes is the "pangenome" concept. It is now understood that horizontal gene transfer has played a pivotal role in the evolution of prokaryotes (Boto, 2010; Boyd et al., 2009; Juhas et al., 2009; Ochman et al., 2000). In a nutshell, horizontal gene transfer (HGT) is the acquisition of DNA from the environment and its integration into the genome of the recipient species. The genes would be inherited by the daughter cells, even though they were not transmitted "vertically". Such genes or gene clusters (henceforth, generically referred to as genomic islands (GI)), code for various protein(s) with myriad functions. Their acquisition can result in "quantum leaps" by bacterial genomes (Boto, 2010; Nakamura et al., 2004). However, un-concerted expression of any recently-acquired gene(s) or expression of toxic proteins from bacteriophages (Canchaya et al., 2004) (Casjens, 2003) can have disastrous effects on cellular homeostasis of the host. Hence, after entering a genome, most GIs are repressed by silencing mechanisms that act at different stages of the gene expression process (Navarre et al., 2007).

Although mechanisms that control initiation and repression of transcription in GIs have been studied, the importance of transcription termination at genomic islands was noticed only recently. In *Escherichia coli*, the global regulatory role of Rho has emerged from two studies using either microarray or ChIP-chip approaches, (Cardinale et al., 2008; Peters et al., 2009). Such studies have unambiguously shown

Abbreviations: GI, genomic islands; IT, intrinsic terminator; HGT, horizontal gene transfer; RNAP, RNA polymerase.

* Corresponding author at: Department of Microbiology and Cell Biology, Indian Institute of Science, Bangalore-560012, India. Tel.: +91 80 22932598; fax: +91 80 23600668.
E-mail address: vraj@mcbl.iisc.ernet.in (V. Nagaraja).

0378-1119/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.07.064

that in *E. coli* Rho-dependent termination is important in suppressing aberrant expression from the genomic islands including prophages.

In this manuscript, we propose that the suppression of transcription in GIs could be a universally conserved function of Rho. We show that Rho-dependent termination indeed seems to play a similar role in regulating expression at GIs across diverse bacterial phyla. Furthermore, based on the experimental understanding about the mechanism of interactions of Rho with the nascent RNA and RNAP (Dutta et al., 2008), we suggest that the lesser density of ITs in GI could actually facilitate Rho-dependent termination.

2. Materials and methods

The program GeSTer can recognize both canonical and non-canonical intrinsic terminators. The mode of action of GeSTer has been described earlier (Mitra et al., 2009, 2010; Unniraman et al., 2002). All genome sequences were downloaded in their GenBank format from NCBI (ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/). Sequence information about GIs was obtained from available literature, and from the NCBI lists for individual genomes. Once GeSTer has identified all the ITs for a genome, we analyzed the intrinsic terminator-content in the GIs of that genome. For a given GI, the density of ITs (D_{IT}) was calculated as the [(number of ITs identified)/(number of genes)] × 100. Similarly, the genomic D_{IT} = [(number of ITs identified in genome)/(number of genes in genome)] × 100.

To ascertain which transcription units (multigenic operon or single-gene) had an IT downstream, the gene at the 3′ end of a multigenic operon was identified from the DOOR database, and the GeSTer results for that genome were analyzed to see if that 3′ terminal gene had an IT after its stop codon. The HGT (IT/TU)% was calculated as (number of ITs in the GI)/(number of transcription units in the GI). The total number of transcription units in the genome was calculated from the genome-specific statistics available at the DOOR site (http://csbl1.bmb.uga.edu/OperonDB_10142009/displayspecies.php).

Genomic (IT/TU)% is calculated as (number of ITs in the genome)/(number of transcription units in the genome).

3. Results and discussion

3.1. Rationale for the experimental design

A salient result of the microarray studies in *E. coli* K-12 is that, when Rho action was inhibited by the antibiotic Bicyclomycin, the transcription of several GIs (known as K-islands in *E. coli* K-12 MG1655) significantly increased (Cardinale et al., 2008). These studies also revealed an under-representation of ITs in the same K-islands. Yet another study, treatment of *E. coli* K-12 MG1655 with sublethal dosage of Bicyclomycin followed by ChIP-chip analysis showed several regions on the chromosome where RNAP could localize only in presence of Bicyclomicin (Peters et al., 2009). The inference was that Bicyclomycin specifically inhibited Rho in these cells, thus allowing RNAP to transcribe into regions where Rho would have caused termination in absence of the antibiotic (Peters et al., 2009). These genomic regions, named Bicyclomycin Sensitive Regions (BSRs), are thus sites where Rho-dependent termination would normally occur. The study identified 23 BSRs which were downstream of K-12-specific genes (belonging to K-islands) or prophage DNA. We analyzed the IT profile of these BSRs and found that they have an under-representation of ITs and hairpins. Of the 23 BSRs that are downstream of the GIs, there was not a single IT or even a stable hairpin-forming sequence in 16 (70%) of them (Supplementary Table S1). Thus, ITs are under-represented in those regions of *E. coli* genome where Rho is functioning. In fact, the scarcity of ITs seems to have been compensated by the action of Rho (Cardinale et al., 2008). Hence, Rho is most likely to terminate transcription at the ends of genes where ITs are absent as these are the only mechanisms of termination known in bacteria. This would mean that the intrinsic D_{IT} of the GIs of any genome could be a pointer of Rho activity at such genomic islands. In other words, if the D_{IT} of GI(s) is lower than the D_{IT} of the whole genome, then Rho-dependent termination is probably an important mode of regulation in these GI(s). Hence, we selected representative genomes from different phyla and classes, for which information about GIs was available, and analyzed their IT profiles using the algorithm, GeSTer, which detects both canonical and non-canonical ITs (Mitra et al., 2009; Mitra et al., 2010; Unniraman et al., 2002). If the assumption that GIs across bacteria have extensive Rho-dependent termination is correct, we should observe a consistent trend of decreased presence of ITs in GIs in different species. Our sample included well characterized prophages, cryptic phages and other kinds of GIs.

3.2. GIs of other E. coli strains are poor in ITs

The importance of Rho-dependent termination in GIs of *E. coli* was based primarily on experiments in *E. coli* K-12 MG1655. In particular, the paucity of ITs in GIs was shown only for the K-islands of *E. coli* K-12 (Blattner et al., 1997; Cardinale et al., 2008). At first, we ensured that the results reported for the K-islands of *E. coli* K-12 could also be obtained using GeSTer. Tabulation of the ITs in 42 K-islands (Cardinale et al., 2008) showed that indeed, there was ~50% reduction in D_I. D_{IT} in these GIs was only 21.9% as compared to the whole genomic D_{IT} of *E. coli* K-12 of 41.7%. Thus, although we had used a different algorithm, these results were consistent with the previous study. Next, we considered another "model" strain, *E. coli* 0157:H7 EDL933, which also houses several GI, collectively called O-islands (OIs). As with *E. coli* K-12, the OIs of this genome also show enhanced transcription after bicyclomycin treatment (Cardinale et al., 2008). Hence, the IT profile of 11 OIs—OI-7, 8, 9, 35, 36, 43, 44, 45, 47, 48 and 50 (consisting of a total of 616 genes i.e.11.8% of genome) of *E. coli* 0157:H7 EDL933 was analyzed. The major criterion for selecting these OIs was that they all were relatively large GIs. The largest among them, OI-43, encoded for 106 genes while the smallest, OI-35, contained 15 genes. Additionally, in order to assess the regions annotated as resident phages, we selected a prophage (OI-45) and four representative cryptic phages. Out of 616 genes from these 11 O-islands, only 135 genes have an IT immediately downstream. Thus, as observed in *E. coli* K-12, the number of IT is distinctly lower (D_{IT} = 21.9%) in these islands as compared to the genomic D_{IT} of 36.6% (Fig. 1A). A closer examination into the IT profiles of the individual OIs showed that large stretches of genes are devoid of any ITs. Also, as reported in *E. coli* K-12, we note that many genes occur in series on the same strand and most of these genes, including the gene at the 3′ end of the series, often lack ITs (Cardinale et al., 2008). If these serial gene clusters are operons, then it seems likely that they lack an IT downstream. In addition, ITs are absent for most of the genes that are at the 5′ or 3′ ends of the OIs. Lack of identifiable ITs hints at the possibility that Rho-dependent termination is probably the major termination mechanism in these OIs.

The genomes of two other strains of *E. coli* – enteropathogenic *E. coli* 234869 (Iguchi et al., 2009) and uropathogenic *E. coli* CFT073 (Lloyd et al., 2007) – code for several experimentally characterized pathogenicity islands. The total number of GI genes identified in *E. coli* 234869 is 493. Besides prophages, these GIs also include the LEE island that has been implicated in virulence. Similarly, the CFT073 strain houses the well-known islands – PAI-II, PAI-III and PAI-CFT073-serX – that encode a total of 299 genes (Lloyd et al., 2007, 2009). The D_{Is} of these islands show that there is a similar decrease in abundance of ITs. The D_{IT} of the islands were 19.9% and 20.1% for strains 234869 and CFT073 respectively i.e. between 50 and 58% of the genomic values (Fig. 1B, Supplementary Fig. 1A). A detailed analysis of the two islands — PAI-II from strain CFT073 and LEE from strain 234869 for the presence of ITs in relation to the genomic