DETECTING HAIRPINS IN 3′-UNTRANSLATED REGIONS OF HIGHLY EXPRESSED GENES IN ACTINOBACTERIA

Lyubetskaya E.V., Seliverstov A.V.*, Lyubetsky V.A.
Institute for Information Transmission Problems, RAS, Moscow, Russia,
*Corresponding author: e-mail: slvstv@iitp.ru

Key words: transcription termination, bacteria, DNA relaxation, highly expressed genes

SUMMARY

Motivation: In bacterial genomes antiparallel genes are common within one chromosome. At least in cases when one of the genes is highly expressed, their shared 3′-untranslated region should be involved in transcription termination. At the same time, gene transcription very often entails the formation of DNA supercoils at this site.

Results: Very long hairpins are found in some Actinobacteria downstream of highly expressed genes. But these hairpins do not resemble known types of terminators involved in expression regulation. We suppose that they are involved in DNA relaxation and an uncharacterized termination mechanism.

INTRODUCTION

In Actinobacteria, gene expression is typically regulated on translation level through overlapping of ribosome binding site (Seliverstov *et al*., 2005), whereas in gamma- and alpha-proteobacteria the regulation follows classical attenuation scenario (Vitreschak *et al*., 2004). In most well studied bacteria, like *Escherichia coli* and *Bacillus subtilis*, operons end with transcription terminators, GC-rich hairpins with adjacent poly-U tract downstream. However, such terminators are rare or absent altogether from some bacterial taxa, like e.g. Cyanobacteria and *Mycobacterium* (Washio *et al*., 1998; Unniraman *et al*., 2002). One might suppose that transcription termination in Actinobacteria involves alternative secondary structures of double-stranded DNA.

In this work we sought for DNA structures putatively responsible for termination in 3′-untranslated regions of highly expressed genes encoding tRNA, elongation factors and some important proteins.

Structures found in 3′-untranslated regions of these genes allowed for better defining operon boundaries and predicting highly transcribed DNA regions (this problematic was discussed in detail in (Ishchukov *et al*., 2004)). Particularly, results of the search algorithm were used to reveal expressed paralogs.

RESULTS

Bacterial genomes were obtained from GenBank. Long hairpins are found in 3′-untranslated regions of genes encoding tRNA and some proteins, being especially abundant in intergenic spaces between antiparallel genes, with one of them coding for tRNA. For example, in *Propionibacterium acnes* long hairpins are found downstream of

tRNA-Ala (*ppa2421*), tRNA-Arg (*ppa2413*), tRNA-Arg (*ppa2189*), tRNA-Asn (*ppa2422*), tRNA-Glu (*ppa2432*), tRNA-Lys (*ppa0181*), tRNA-Lys (*ppa1961*), tRNA-Met (*ppa2423*), tRNA-Phe (*ppa2454*), tRNA-Pro (*ppa2428*), tRNA-Thr (*ppa2412*). In *Corynebacterium efficiens* long hairpins are found downstream of tRNA-Ala, tRNA-Arg, tRNA-Asp, tRNA-Leu, tRNA-Pro, tRNA-Ser. Moreover, such hairpins are found downstream of other six highly expressed protein-coding genes in *P. acnes*, six such genes in *C. efficiens* and five protein-coding genes in *Mycobacterium bovis*.

A part of our data is presented in Table 1:

Table 1. The numbers of hairpins with length equal or higher than L for leader regions (1), regions of converging located genes (2), coding regions (3), regions of divergently located genes (4) are shown in second, third, forth and fifth columns, respectively

L	(1)	(2)	(3)	(4)	L	(1)	(2)	(3)	(4)
Corynebacterium efficiens					Mycobacterium bovis				
25	2	1	0	0	25	2	12	1	1
23	2	2	0	0	23	3	15	1	1
20	6	16	0	0	20	4	17	2	1
17	23	37	5	1	17	12	21	6	2
15	44	57	13	6	15	17	22	28	3
10	182	121	960	27	10	188	37	1499	37
Corynebacterium glutamicum					Mycobacterium leprae				
25	0	1	0	0	25	5	0	2	1
23	0	3	1	0	23	5	0	2	1
20	2	16	1	0	20	8	1	4	1
17	12	40	4	0	17	10	1	12	1
15	29	59	8	0	15	14	1	27	2
10	221	133	617	29	10	111	4	482	8
Corynebacterium diphtheriae					Mycobacterium avium				
25	0	2	0	0	25	1	1	0	0
23	1	6	0	0	23	3	1	0	0
20	2	14	0	0	20	4	4	0	0
17	14	32	2	0	17	8	8	8	0
15	24	48	7	0	15	17	19	24	3
10	137	76	497	23	10	257	46	2197	66
Propionibacterium acnes					Mycobacterium tuberculosis				
25	0	0	0	0	25	4	11	3	1
23	0	0	0	0	23	5	14	3	1
20	0	2	1	0	20	7	17	3	2
17	3	17	2	0	17	17	20	6	3
15	8	27	3	0	15	24	21	24	5
10	101	63	571	17	10	202	36	1519	30

The length of a hairpin is the number of nucleotides in its shoulders. The numbers of hairpins with length equal or higher than L for leader regions (1), trailer regions (2), coding regions (3), regions of divergently located genes (4) are shown in second, third, forth and fifth columns, respectively. Besides, the Table 1 shows only hairpins with loops shorter than 15 nucleotides and with only one internal loop 2 nucleotides or less in length. Moreover, the left shoulder was not allowed to contain regions complementary to those of the hairpin loop.

Hairpins of 18–27 bp length (called abnormally long hairpins) are seldom found in some genomes (results shown for *P. acnes*). For each gene, its leader region was defined as a region no more than 300 bases in length and not crossing the bounds of neighbor genes. Transcription initiation site was not considered and is usually unknown. In the *P. acnes* genome, mass searches for long hairpins without bulges in 5′-untranslated regions of up to 300 bp length upstream of all genes contained in GenBank annotation resulted in detecting four hairpins with stem size exceeding 18 bp. Hairpins with stems longer than 28 bp were not detected in

intergenic spaces of this genome. Two hairpins were found with 27 bp, one – with 22 bp and one – with 18 bp-long stems. Here the first two are described.

A hairpin with a 4 bases-long loop and 27 bp-long stem without bulges was detected in 3′-untranslated region immediately following the stop codon of elongation factor G at an 8-base distance. Downstream of the hairpin the gene of transmembrane protein PPA1874 is located. Both genes are of considerable length.

The other hairpin is confined in between genes *ppa1754* and *ppa1753* encoding the alpha subunit of highly expressed succinyl-CoA synthetase and a putative transmembrate protein, respectively.

DISCUSSION

In bacterial genomes antiparallel genes are common within one chromosome. At least in cases when one of the genes is highly expressed, their shared 3′-untranslated region should be involved in transcription termination, which is probably mediated by the found hairpins. For instance, tRNAs genes are highly expressed because of intensive usage of their products in the cell.

Besides, gene transcription entails formation of DNA supercoils, also in the 3′-untranslated region, which are conventionally thought to be relaxed by topoisomerases. Although, in intergenic regions, with at least one gene highly expressed, an alternative process might be involved in DNA relaxation with the use of detected hairpins.

In other words, in 3′-untranslated region of a highly expressed gene (especially if it belongs to a pair of antiparallel genes) one might expect to find a pair of hairpins forming the so called “cross” on two DNA strands.

Comparative analysis of hairpins in orthologs of close species reveals high divergence of their primary structure and high conservation of topology, which implies severe functional constraints imposed on the hairpin secondary structure.

These hairpins do not resemble known types of terminators involved in expression regulation. Indeed, the Rho-independent terminator typically contains a U-rich region, and the Rho-protein binding site has a UC-rich region lacking hairpins. None of these is found in or nearby the detected hairpins.

ACKNOWLEDGEMENTS

The authors are grateful to L.Y. Rusin and A.G. Vitreschak for discussion and help. This study was supported by ISTC grant No. 2766.

REFERENCES

Ishchukov I.M. *et al*. (2004) A new algorithm for recognizing the operon structure of procaryotes. *Proceedings of the Fourth International Conference on Bioinformatics of Genome Regulation and Structure*, Novosibirsk, pp. 73–76.

Seliverstov A.V. *et al*. (2005) Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiol*., **5,** 54.

Unniraman S. *et al*. (2002) Conserved economics of transcription termination in eubacteria. *Nucl. Acids Res.*, **30**, 675–684.

Vitreschak A.G. *et al*. (2004) Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS Microbiol Lett*., **234**, 357–70.

Washio T. *et al*. (1998) Analysis of complete genomes suggests that many prokaryotes do not rely on hairpin formation in transcription termination. *Nucl. Acids Res.*, **26**, 5456–5463.