Computational Biology and Chemistry 49 (2014) 7–13

Contents lists available at ScienceDirect

Computational Biology and Chemistry

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

Research Article

Gene expression regulation of the PF00480 or PF14340 domain proteins suggests their involvement in sulfur metabolism

Vassily A. Lyubetsky*, Semen A. Korolev, Alexandr V. Seliverstov, Oleg A. Zverkov, Lev I. Rubanov

Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Russia

ARTICLE INFO

Article history:
Received 11 November 2013
Received in revised form 6 January 2014
Accepted 9 January 2014
Available online 22 January 2014

Keywords:
Mycobacteria
Sulfur metabolism
Sulfur-containing compounds transport
Cysteine
Rho-dependent transcription attenuation
Classic transcription attenuation

ABSTRACT

The paper studies proteins with domains PF00480 or PF14340, as well as some other poorly characterized proteins, encoded by genes associated with leader peptide genes containing a tract of cysteine codons. Such proteins are hypothetically regulated with cysteine-dependent transcription attenuation, namely the Rho-dependent or classic transcription attenuation. Cysteine is an important structural amino acid in various proteins and is required for synthesis of many sulfur-containing compounds, such as methionine, thiamine, glutathione, taurine and the lipoic acid. Earlier a few species of mycobacteria were predicted by the authors to have cysteine-dependent regulation of operons containing the *cysK* gene. In *Escherichia coli* this regulation is absent, and the same operon is regulated by the CysB transcription activator. The paper also studies Rho-dependent and classic transcription regulations in all annotated genes of mycobacteria available in GenBank and their orthologs in Actinomycetales. We predict regulations for many genes involved in sulfur metabolism and transport of sulfur-containing compounds; these regulations differ considerably among species. On the basis of predictions, we assign a putative role to proteins encoded by the regulated genes with unknown function, and also describe the structure of corresponding regulons, predict the lack of such regulations for many genes. Thus, all proteins with the uncharacterized Pfam domains PF14340 and PF00480, as well as some others, are predicted to be involved in sulfur metabolism. We also surmise the affinity of some transporters to sulfur-containing compounds. The obtained results considerably extend earlier large-scale studies of Rho-dependent and classic transcription attenuations.

1. Introduction

We study putative proteins encoded by genes associated with leader peptide genes containing a *tract of cysteine codons*, among them proteins with the PF00480 (ROK) and PF14340 (DUF4395) domains. We hypothesize that gene expression of such proteins is regulated with cysteine-dependent transcription attenuation, namely the Rho-dependent or classic transcription attenuation. Cysteine is an important structural amino acid in various proteins and is required for synthesis of many sulfur-containing compounds, such as methionine, thiamine, glutathione, taurine and the lipoic acid.

Studies of the Rho-dependent and classic attenuation regulations mediated by the concentration of tryptophan date back to the pioneer works by C. Yanofsky (Konan and Yanofsky, 2000; Yanofsky, 1981). These regulations are characterized by an open reading frame of the leader peptide gene (further referred to as the *leader gene*) that possesses a short tract of regulatory codons. The regulation mechanism is based on the transcription and translation coupling, a common phenomenon in prokaryotes.

In the case of classic attenuation the leader gene is separated from a downstream structural gene (or an operon) by a relatively short spacer region, which transcript sequence forms one or several terminator hairpins and often contains very short poly(U) (U-rich) regions positionally associated with the hairpins' 3′-shoulders. A complex of a terminator and its neighboring poly(U) is called the intrinsic *terminator*. The formation of the terminator hairpins is regulated by other (one or several) hairpins located upstream and called *antiterminators*, or by the ribosome attached to leader gene's codons.

The study (Yanofsky, 1981) established a mechanism of classic attenuation regulation based on the mutually exclusive formation of overlapping terminator and antiterminator on mRNA sequence (transcription attenuation sensu Yanofsky). We proposed alternatives of this mechanism that often involve the formation of pseudoknots, triplexes, and additional hairpins. For each alternative, large-scale predictions were obtained with biological data

* Corresponding author at: IITP RAS, 19 Bolshoy Karetny Lane, Moscow 127994, Russia. Tel.: +7 910 4646917.
E-mail address: lyubetsk@iitp.ru (V.A. Lyubetsky).

http://dx.doi.org/10.1016/j.compbiolchem.2014.01.001
1476-9271/

(Lopatovskaia et al., 2010). In particular, classic attenuation mediated by concentrations of tryptophan, phenylalanine, histidine, threonine or ramified amino acids is predicted in large-scale analyses using bioinformatic tools (Lopatovskaia et al., 2010; Seliverstov et al., 2005); further references are provided in the cited works.

In the case of Rho-dependent attenuation regulation, the stop codon of the leader gene is associated with a short {C,U}-rich site called the *Rho binding site*, which is specifically recognized by the Rho factor. A pyrimidine-rich region may be considered as a putative Rho binding site, which however does not suffice to draw further conclusion. For the Rho factor, the ATPase activity was measured for different nucleotide compositions of mRNA (Kim and Patel, 1999). It was shown to be 3–4-fold higher with poly(C) compared to poly(U), and tens of thousands times lower if no mRNA is available. This and some other indirect evidence did not so far bring more clarity to the definition of the Rho binding site.

Rho-dependent transcription attenuation was first described for expression regulation of the tryptophanase gene *tna* in *Escherichia coli* (Konan and Yanofsky, 2000). This case remains the best studied to date. Another example (Richardson, 2002), a Rho binding site located closely to the start codon in the coding region of the *lac* operon in *E. coli*. For a few species of mycobacteria we earlier predicted the cysteine-mediated Rho-dependent transcription attenuation of the operons containing the *cysK* gene (Seliverstov et al., 2005). Note that *E. coli* lacks this regulation, and the same operon is regulated with an alternative mechanism mediated by the CysB transcription activator. The Rho transcription factor is known to be present in all mycobacteria, with the exception of *Mycobacterium africanum* GM041182.

Under the excess of a regulating amino acid the operon is active, because the Rho binding site is shielded by a ribosome. When the amino acid is in deficiency (e.g., during starvation), ribosomes do not occupy the Rho binding site on mRNA, and it becomes available to the Rho factor that terminates transcription even when the 5′-end of the structural operon has already been transcribed. Functioning of this mechanism can cause the opposite (decreasing instead of increasing) correlation depending on the distance between the regulatory codons and the Rho binding site. The Rho factor can also interact with riboswitches (Hollands et al., 2012). Elsewhere (Konan and Yanofsky, 2000; Yanofsky, 1981; Heery and Dunican, 1993; Lin et al., 1998), mentioned types of the regulations are corroborated experimentally.

In the case of classic transcription attenuation (Lopatovskaia et al., 2010; Seliverstov et al., 2005), the increase of amynoacyl-tRNA concentration is associated with the decrease of transcription continuation frequency caused by the RNA polymerase passing the putative transcription termination site of the structural gene (usually a U-rich tract within the regulatory region). This frequency is called the *transcription level* and is measured in unit fractions (or in percent) as $1-p(c)$, where c is the amynoacyl-tRNA concentration, and $p(c)$ – the frequency of transcription termination (DNA·RNA-duplex dissociation) usually occurring on a U-rich tract. In regulations of catabolism, transport, etc., the transcription level increases with the increase of amynoacyl-tRNA concentration. The dependency $p(c)$ is estimated according to a model of classic attenuation (Lyubetsky et al., 2006, 2007; Rubanov and Lyubetsky, 2007). This model is systematically applied herein to predict regulation on the basis of a noticeable change of the structural gene transcription level $1-p(c)$.

Proteobacteria lack cysteine-dependent transcription attenuation (Lopatovskaia et al., 2010) and possess other regulation types (Kredich, 1992; Lynch et al., 1994). The regulation of sulfur metabolism substantially differs across proteobacterial lineages. In *E. coli*, certain pathways, including that of cysteine synthesis, are regulated by the CysB protein from the LysR family, which binds DNA close to the promoter. In proteobacteria *Salmonella typhimurium*, *E. coli* and *Klebsiella aerogenes* the transcription of the operons *cysPTWAM*, *cysK*, *cysJIH*, *cysDNC*, *sbp*, and the L-cysteine transport system is also activated by the CysB protein (Kredich, 1992; Lynch et al., 1994). In these proteobacteria, self-repression of transcription is described for the gene *cysB*. The CysB·DNA binding is sensitive to concentration of the *O*-acetylserine cysteine precursor but is not sensitive to that of sulfuric compounds. In *E. coli* and other γ-proteobacteria the methionine synthesis is known to be regulated by the MetJ factor, which suppresses transcription upon binding to S-adenosylmethionine (Saint-Girons et al., 1984; Augustus and Spicer, 2011). In some Gram-positive bacteria, including *Bacillus subtilis*, *Clostridium acetobutylicum*, and *Staphylococcus aureus*, the regulation of sulfur metabolism is mediated by the S-box riboswitch (Grundy and Henkin, 1998).

The Rho protein is a homohexamer. Each of its subunits contains two domains, one binding mRNA, and the other being an ATPase. Among the six identical subunits of the hexamer only two possess the ATPase activity due to the molecule's asymmetry and assembly-dependent conformational properties of adjacent subunits. In in vitro experiments the Rho protein binds a 78(C)-long region.

Remember that at the initial stage of cysteine synthesis the serine hydroxyl is acetylated by the *cysE*-encoded serine acetyltransferase. At the next stage, *O*-acetylserine reacts with hydrogen sulfide to form the cysteine. This reaction is catalyzed by the *cysK*-encoded cysteine synthase.

Cysteine is a donor of sulfur in taurine synthesis. In mycobacteria, genes *tauC* and *tauA* involved in the taurine transport system likely belong to the same operon, as their separating spacer is only about 10 nt long (with the exception of *Mycobacterium massiliense* str. GO 06).

Transcriptional repression of plastid genes *cysT* and *cysA* involved in sulfate transport was proposed in the Viridiplantae (Lyubetsky et al., 2013). In this case, a single-box repressor binding conserved motif with the consensus TAAWATGATT was found close to the promoters in many species of algae.

Both *cysK* and *cysE* genes usually belong to the same operon, which in *Mycobacterium avium* and *Mycobacterium leprae* also contains proteins with *unknown function*.

The ROK (Repressor, ORF, Kinase) domain PF00480 belongs to a diverse family of proteins that unites transcription factors (repressors *xylR* in *B. subtilis*, *Lactobacillus pentosus*, *Staphylococcus xylosus*, and *nagC* in *E. coli* that possess a helix-turn-helix DNA-binding motif absent from other members of the family), sugar kinases and uncharacterized proteins (Titgemeyer et al., 1994).

The Pfam domain PF14340 (DUF4395) of unknown function is frequently found across bacteria and eukaryotes. It possesses two conserved cysteic residues likely to be functionally important. The role of PF00480 and PF14340 proteins remained unclear.

Protein functions can be predicted on the basis of their regulation mechanisms. With this notion, we discuss the putative involvement of proteins in sulfur metabolism in Actinomycetales. To detect regulations, we performed a large-scale search for putative leader genes containing poly(Cys) regions in all currently available genomes of the Mycobacteriaceae (species of *Mycobacterium* and *Amycolicicoccus subflavus*) and Actinomycetales. Among many leader genes detected, the majority is associated with structural genes involved in sulfur metabolism, and some – with genes encoding less characterized proteins, including proteins with Pfam domains PF14340 and PF00480. The length of poly(Cys) tracts in leader genes reaches 8 (e.g., upstream of the protein YP_006570365.1[1] gene in *Mycobacterium smegmatis* str. MC2 155), and 4 upstream of the PF00480-domain protein genes.

[1] Hereafter, sequence/protein accession numbers refer to GenBank.