Journal of Theoretical Biology 301 (2012) 83–94

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

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

Dynamics of transcription of closely spaced promoters in *Escherichia coli*, one event at a time

Leonardo Martins [a,1], Jarno Mäkelä [b,1], Antti Häkkinen [b], Meenakshisundaram Kandhavelu [b], Olli Yli-Harja [b,c], José M. Fonseca [a], Andre S. Ribeiro [b,*]

[a] Faculdade de Ciências e Tecnologia Universidade Nova de Lisboa, Monte da Caparica, 2829-516 Caparica, Portugal
[b] Computational Systems Biology Research Group, Department of Signal Processing, Tampere University of Technology, P.O Box 553, FI-33101 Tampere, Finland
[c] Institute for Systems Biology, 1441N 34th St, Seattle, WA 98103-8904, USA

ARTICLE INFO

Article history:
Received 24 October 2011
Received in revised form
8 February 2012
Accepted 13 February 2012
Available online 20 February 2012

Keywords:
Closely spaced promoters
Stochastic dynamics
Transcription start sites
Open complex formation

ABSTRACT

Many pairs of genes in *Escherichia coli* are driven by closely spaced promoters. We study the dynamics of expression of such pairs of genes driven by a model at the molecule and nucleotide level with delayed stochastic dynamics as a function of the binding affinity of the RNA polymerase to the promoter region, of the geometry of the promoter, of the distance between transcription start sites (TSSs) and of the repression mechanism. We find that the rate limiting steps of transcription at the TSS, the closed and open complex formations, strongly affect the kinetics of RNA production for all promoter configurations. Beyond a certain rate of transcription initiation events, we find that the interference between polymerases correlates the dynamics of production of the two RNA molecules from the two TSS and affects the distribution of intervals between consecutive productions of RNA molecules. The degree of correlation depends on the geometry, the distance between TSSs and repressors. Small changes in the distance between TSSs can cause abrupt changes in behavior patterns, suggesting that the sequence between adjacent promoters may be subject to strong selective pressure. The results provide better understanding on the sequence level mechanisms of transcription regulation in bacteria and may aid in the genetic engineering of artificial circuits based on closely spaced promoters.

1. Introduction

Genes of *Escherichia coli* differ widely in expression kinetics (Taniguchi et al., 2010) due to, among other factors, the diversity of promoter sequences driving their expression. The regulation of expression levels is usually exerted during transcription initiation, a highly complex, multi-stepped process that starts with the binding of the RNA polymerase (RNAp) to the promoter region, followed by DNA unwinding and stabilization of the closed complex, assembly of the clamp/jaw on downstream DNA, formation of the open complex, and promoter escape (Browning and Busby, 2004; Saecker et al., 2011; Hsu, 2002). Only after, can RNA synthesis begin.

In vitro studies of the kinetics of initiation of several promoters in *E. coli* suggest that this process can have up to three sequential steps: formation of a closed complex, isomerization, and formation of the open complex (Saecker et al., 2011; Buc and McClure, 1985; McClure, 1985). Their duration varies between promoters, even when the sequences only differ slightly (Lutz et al., 2001; Singh et al., 2011), and are tightly regulated by repressing and activating molecules, whose bindings are, in general, in the promoter region.

The genome of *E. coli* contains various sites with closely spaced transcription start sites (TSSs). The geometry of these promoters with closely spaced TSSs can be tandem (same direction of elongation), divergent (directions of elongation are opposite, in a back-to-back fashion), or convergent (directions of elongation are opposite, in a front-to-front fashion) (McClure, 1985; Beck and Warren, 1988). Other sources of diversity between these promoters with closely spaced TSSs are the distance, in number of nucleotides, between the two TSSs and the location of the transcription factor binding sites (TFBS). Here, isolated TSSs are referred to as unidirectional promoters.

A recent survey suggests that approximately 15% of the promoters in *E. coli* are closely spaced (Gama-Castro et al., 2010). The same configurations have been found in quantity in other organisms, also exhibiting structural diversity at various levels (Beck and Warren, 1988; Häkkinen et al., 2011).

The kinetics of expression of genes driven by closely spaced promoters remains relatively unstudied, particularly in prokaryotes. In these organisms, it is yet unknown to what extent is the

---

[*] Corresponding author. Tel.: +358 408 490 736; fax: +358 331 154 989.
E-mail addresses: andre.sanchesribeiro@tut.fi, andre.ribeiro@tut.fi (A.S. Ribeiro).

[1] Equal contributions.

0022-5193/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2012.02.015

production of RNA molecules affected by possible interference between RNA polymerases (RNAps). In Sneppen et al. (2005) an analytical model was proposed to study the impact of transcriptional interference on mean expression levels. The model includes three mechanisms of interference, namely, occlusion (passing RNAps block access to the promoter), collisions between elongating RNAps, and "sitting duck" interference. Given the features of their model, the analysis focused on study of mean expression rates and, regarding this feature of the kinetics, the results agreed with measurements from convergent promoters (Callen et al., 2004).

While there is little study of the effects of proximity of TSSs in prokaryotes, there is some information available from studies in eukaryotes. In Wang et al. (2011) the dynamics of a stochastic model of closely spaced promoters was analyzed. The model used accounted for chromatin remodeling by switching the promoter state between ON and OFF via stochastic, first order reactions. Gene expression was modeled as a single step event. The results suggest that the orientation and distance between TSSs affect noise in RNA and proteins numbers. It is noted that this model assumes *a priori* that distance and other topological features (not modeled explicitly) have effects on the correlation between the expression of the two genes, thus on both genes' mean expression and noise, rather than determining these effects from the structure.

Another work studied the effects of genetic and epigenetic properties of promoters on expression variability in budding yeast (Woo and Li, 2011) from genome-wide datasets of gene expression and nucleosome occupancy. The authors suggest that, for this organism, divergent TSSs tend to have lower expression variability than tandem TSSs and that this variability, for both configurations, tend to decrease with decreasing distance. These results are somewhat in disagreement with those of Wang et al. (2011) for distances between TSSs shorter than 300 nucleotides. Finally, in Ebisuya et al. (2008) experimental evidence was reported that, in eukaryotes, transcription initiation appears to exert a "ripple effect", that is, the induction of expression of a gene tends to stimulate the expression of neighbor genes. The authors suggest that this mechanism may be advantageous for coordinated expression of genes participating in similar functions. Since chromatin dynamics and DNA methylation may play a role in this effect, and given several other differences in the mechanisms of gene expression, it is unknown if the effects of TSSs proximity is similar in prokaryotes and eukaryotes, although it is likely that the proximity does play a role in the expression kinetics of closely spaced genes in both cases. Due to that, models must be made at the nucleotide level so as to study, among other features, interferences in the expression dynamics.

Here, using the delayed stochastic simulation algorithm (delayed SSA) (Roussel and Zhu, 2006) to drive the dynamics, we model promoters at the nucleotide level and simulate the kinetics of transcription, one RNAp at a time. One of the novelties of this study, allowed by the model used, is the quantification of the effects of changes in promoters, at the nucleotide level, on the kinetics of RNA production. Also, we account for the duration of rate limiting steps at the TSS, such as isomerization and the open complex formation, which varies from one event to the next.

Following the description of the model and comparison of its predictions to measurements, we first study the kinetics of the binding of RNAps to the promoter sequence. Next, we study the dynamics of RNA production as well as the degree of correlation between consecutive choices of directions of elongation as a function of the geometry. Finally, we study the kinetics of expression and its regulation by repression by occlusion as a function of the positioning of the TFBS, among other variables. In the end, we present our conclusions and address the following questions: what are the effects of the rate limiting steps at the TSS, the closed and open complex formations, on the kinetics of RNA production for various promoter configurations? Are there abrupt changes in the kinetics of RNA production with nucleotide distance between TSSs? To what extent does the proximity between two TSSs correlate the dynamics of RNA production under their control?

2. Methods

Transcription in prokaryotes is both a stochastic process (Arkin et al., 1998) and sparse in time (Taniguchi et al., 2010; McClure, 1985), which imposes the use of Monte Carlo methods to simulate it, such as the stochastic simulation algorithm (SSA) (Gillespie, 1977). Additionally, the process of initiation contains several rate limiting steps. Usually, there appear to be two major rate limiting steps, the closed and the open complex formations (McClure, 1985; Lutz et al., 2001). The first includes the finding of the promoter region and diffusion of the RNAp along the DNA template until reaching the TSS and forming of the closed complex. The second includes a few isomerization steps until the open complex is formed (deHaseth et al., 1998).

Given the above, we use the delayed SSA (Roussel and Zhu, 2006) to drive the kinetics of the models since, unlike the original SSA (Gillespie, 1977), it allows delayed events. In these, once the reaction occurs and the reactants are removed from the vessel of reactions, the products are kept on a wait list for a predetermined amount of time, and only after are made available for reactive events. To implement the delayed SSA and simulate the models described below we use the simulator SGNSim (Ribeiro and Lloyd-Price, 2007). An example of the implementation of one the models that can be simulated by SGNSim is provided in supplementary material. All models are described in detail in the supplementary material as well.

The models of promoters, namely, their nucleotide structure, as well as various kinetic rate constants, are extracted from measurements. We extracted the sequences of known and predicted divergent and convergent closely spaced promoters in *E. coli* from the RegulonDB database (version 7.0) (Gama-Castro et al., 2010). The distributions of nucleotide length between TSSs are shown in Fig. 1. The bulk of the distribution is below 200 nucleotides in length (88.8% for convergent and 61.8% for divergent). Mean distances are 108.4 and 225.7 for convergent and divergent promoters, respectively. In all models below, the distances between TSS are set within these realistic intervals.

The model of transcription, the set of possible reactions and events, along with the stochastic rate constants are described in supplementary material. The first step towards the production of an RNA molecule is the binding of the RNAp to the DNA template. The RNAp can then diffuse one nucleotide at a time in a direction chosen initially at random. Provided long time intervals, it is believed that diffusion RNAps can change direction. However, these changes are not common events, that is, for short distances (of tenths of nucleotides) evidence suggests that the direction of diffusion does not change (Sakata-Sogawa and Shimamoto, 2004; Gorman and Greene, 2008). Nevertheless, we note that this assumption (if wrong) likely does not affect the results significantly, due to the much higher speed of diffusion (600 nuc/s) and disassociation ($0.3\ s^{-1}$) in comparison with the other possible events, such as those at the TSS and during elongation.

The RNAp can unbind from the template at any step. If there are multiple RNAps on the template, there can be collisions. In that case, one of the RNAps falls off the template (randomly chosen). In convergent promoters, elongating and diffusing RNAps can collide (Callen et al., 2004). Since the elongating RNAp is more