Genomics xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Genomics

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

Computational prediction of functional abortive RNA in *E. coli*

Jeremy I. Marcus [a], Soha Hassoun [a,b], Nikhil U. Nair [b,*]

[a] Department of Computer Science, Tufts University, Medford, MA 02155, United States
[b] Department of Chemical and Biological Engineering, Tufts University, Medford, MA 02155, United States

ARTICLE INFO

Article history:
Received 30 September 2016
Received in revised form 24 February 2017
Accepted 22 March 2017
Available online xxxx

Keywords:
Noncoding
RNA
Transcription
E. coli
Abortives
Regulation

ABSTRACT

Failure by RNA polymerase to break contacts with promoter DNA results in release of bound RNA and re-initiation of transcription. These abortive RNAs were assumed to be non-functional but have recently been shown to affect termination in bacteriophage T7. Little is known about the functional role of these RNA in other genetic models. Using a computational approach, we investigated whether abortive RNA could exert function in *E. coli*. Fragments generated from 3780 transcription units were used as query sequences within their respective transcription units to search for possible binding sites. Sites that fell within known regulatory features were then ranked based upon the free energy of hybridization to the abortive. We further hypothesize about mechanisms of regulatory action for a select number of likely matches. Future experimental validation of these putative abortive-mRNA pairs may confirm our findings and promote exploration of functional abortive RNAs (faRNAs) in natural and synthetic systems.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

Non-coding RNAs (ncRNAs) are RNA molecules that are not translated into proteins and have been shown to be involved in a variety of key cellular processes. They function as gene regulatory molecules and generally exert action by the occluding or exposing binding elements in mRNA. Due to their predictability and versatility, they have been used extensively in engineering synthetic biological systems. Well-described regulatory ncRNAs in bacterial include ribosomal RNA (rRNA), transfer RNA (tRNA), small RNA (sRNA), and anti-sense RNA (asRNA) [1–3].

Abortive RNA transcripts are a poorly-documented class of ncRNAs characterized by their small size and unique mechanism of generation during transcription. RNA transcription involves three basic stages: initiation, elongation, and termination. Once RNA polymerase (RNAP) binds to a DNA promoter during initiation, it repetitiously synthesizes and releases abortive transcripts while remaining bound to the promoter region in a process known as abortive cycling. This phenomenon has been observed to some extent in nearly all in vitro transcription reactions involving RNAPs from different species, and has also been detected *in vivo* in *E. coli* [4–7]. Different RNAPs generate abortive fragments of varying length; for example, human RNAP II and *E. coli* RNAP release transcripts of up to 8 and 15 nt (nucleotides), respectively [8,9]. It has been estimated that only 1 out of every 10 to 100 transcription reactions initiated by RNAP results in successful transition to the elongation phase [10,11]. As a result, abortive initiation cycling leads to the accumulation of short abortive RNA transcripts. Short single-stranded unstructured RNA fragments tend to be unstable and are degraded quickly; but they can form weak, transient complexes with complementary nucleotide sequences [12]. For these reasons, it was considered unlikely that abortive transcripts could serve a functional role.

However, a recent study in the T7 bacteriophage identified a role for abortive transcripts in antitermination at the $T\phi 10$ terminator [13]. During early stages of infection, late gene expression is repressed by this rho-independent terminator. In late lifecycle, accumulation and binding of abortive transcripts to the upstream leg of $T\phi 10$ was shown to prevent hairpin formation and subsequently prevent termination. This resulted in read-through and expression of genes downstream of the terminator. The inherent time lag between initial gene expression and the accumulation of sufficiently high concentrations of abortive transcripts resulted in delayed expression of the downstream genes, which were speculated to be instrumental in T7 phage lifecycle.

A similar novel gene regulation mechanism has yet to be identified outside of the T7 bacteriophage. Further investigation of regulatory roles of abortive transcripts in other organisms requires systematic identification of abortive transcripts and their putative targets. *E. coli* is one of the most well-studied model organisms in genetics; thus, there is a wealth of available information describing its genome and regulatory mechanisms. This makes *E. coli* an appropriate choice for exploring the regulatory roles of abortive transcripts. However, the *E. coli* genome of approximately 4.6 million base-pairs, is much larger than the 39,937 base-pair T7 bacteriophage genome. Large quantities of genetic content can be prohibitive for experimentally conducting genome-wide searches for novel regulatory mechanisms. Predictive computational

* Corresponding author.
E-mail address: nikhil.nair@tufts.edu (N.U. Nair).

http://dx.doi.org/10.1016/j.ygeno.2017.03.003
0888-7543/© 2017 Elsevier Inc. All rights reserved.

ARTICLE IN PRESS

models can help expedite the process by focusing the experimental search space onto a manageable subset of the genome.

In this study, we utilized computational methods to predict locations in the *E. coli* genome where abortive fragments might perform some functional role in regulation at the transcriptional- or translational-level. We identified matches occurring in functionally relevant genomic features, such as terminators and ribosomal binding sites, and ranked these matches using quantitative free energy calculations and subjected them to statistical tests to assess their significance relative to random hits. Here we suggest mechanisms of regulatory control for three of the abortive fragments returned by our analysis.

2. Results

2.1. Energetics of abortive RNA-mRNA binding

Abortive initiation in *E. coli* generally results in the release of 2 to 15 nt long abortive fragments [14,15]. We chose to focus on fragments of lengths 4–15 nt, as shorter sequences are less likely to be able to exact a physiologically relevant effect beyond transcriptional priming [16]. We assumed that all lengths of all abortives within our specified range are produced, and therefore performed our initial Watson-Crick base-pairing search over the entire range for each of the 3780 known transcriptional units in the *E. coli* genome. An additional search performed using wobble base-pairing rules was run under the same conditions. For each abortive match site found, we calculated the standard Gibbs free energy of hybridization ($\Delta G^\circ_{binding}$) as a measure of physiological relevance. Free energy has previously been shown to be strongly and linearly correlated with the ability of abortive transcripts to disrupt the function of an intrinsic terminator [13]. A more negative free energy value implies that a reaction is more spontaneous, implying that a given abortive fragment is more likely to bind strongly and effect a regulatory function. Therefore, we could utilize free energy as a quantitative measure of an oligomer's potential to exert function, providing a basis for comparison between individual matches. Fig. 1 shows the distributions of $\Delta G^\circ_{binding}$ values for all matches found using both matching paradigms over all abortive RNA lengths.

In evaluating the functional significance of abortive fragments, we limited our analysis to those matches that occurred within two types of known functional genetic regions – rho-independent terminators and ribosomal binding sites (RBSs). Tables 1 and 2 present summaries of the filtered data obtained using Watson-Crick and wobble base-pairing respectively. Fig. 2a and b display the distribution of free energy values calculated for abortive matches located in RBSs and terminating regions, respectively.

The existence of a minimum free energy requirement to achieve effective antitermination or anti-translation by antisense hybridization has not been shown. However, Lee et al. demonstrated that antitermination of abortives at the T7 $\phi 10$ terminator was directly correlated to the free energy of hybridization [13]. They demonstrated that abortives with free energy of hybridization <−7 kcal/mol are able to antiterminate with >60% efficiency and those with <−9 kcal/mol are able to antiterminate with >80% efficiency. As comparison, *trans*-regulation by sRNA has been shown to be weak when hybridization energies are <−10 kcal/mol even with the aid of Hfq [17]. Thus, faRNA seem to be able to execute their function under less favorable thermodynamic conditions compared to sRNA. This may be the case because abortive fragments are produced in *cis* configuration, often in large excess compared to productive transcript, and their high localized concentration may be able to compensate for the weaker hybridization energies [18].

2.2. Statistical analyses of putative faRNA binding locations

When searching for binding sites for target sequences within a genomic sequence, one would expect to find those sites randomly scattered throughout. We expected that, on average, matches occur by random chance and are not functionally relevant. For shorter-length abortives (4–7 nt), the frequencies of actual exact matches are only slightly (<20%) more than those expected by random chance (Table 3).

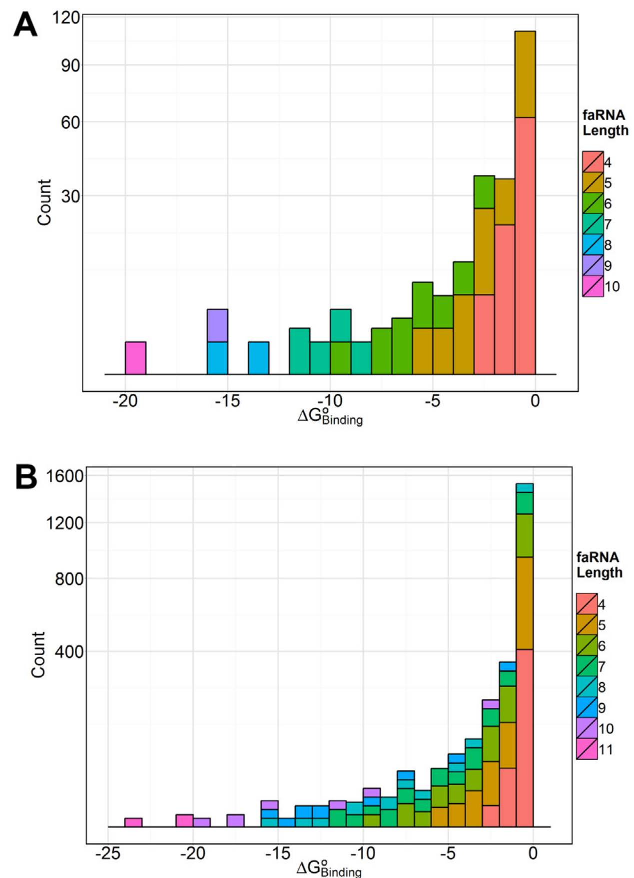

Fig. 1. Distribution of computed $\Delta G^\circ_{binding}$ values for predicted complexes between abortive initiation fragments and mRNAs. We only report complexes which occur within the same transcriptional unit from which the abortive initiation fragment was generated. RNA binding free energy calculations performed using UNAFold *hybrid2.pl*, with temperature range 0–100 °C, 1.0 μM concentrations of both strands. Counts displayed on a square-root scale for visual clarity. Colored bars correspond to the length of the nucleotide match.

(A) $\Delta G^\circ_{binding}$ distribution for complexes exhibiting exact Watson-Crick (A-U, G-C) base-pair matching.
(B) $\Delta G^\circ_{binding}$ distribution for complexes exhibiting exact base-pair matching under a wobble base-pairing paradigm (A-U, G-C, G-U).

Table 1
Abortive RNA hybridization sites within mRNA as matched by Watson-Crick base-pairing.

Abortive fragment length	Transcription units containing ≥ 1 match	Total match count (chance of finding[a])	Matches within terminators (chance of finding[a])	Matches within RBSs (chance of finding[a])
4	3448	26,290 (696%)	89 (2.36%)	6 (0.159%)
5	2386	6784 (180%)	26 (0.688%)	2 (0.0529%)
6	1097	1789 (47.3%)	11 (0.291)	0 (0%)
7	386	453 (12.0%)	5 (0.132%)	0 (0%)
8	123	129 (3.41%)	2 (0.0529%)	0 (0%)
9	42	42 (1.11%)	1 (0.0265%)	0 (0%)
10	14	14 (0.370%)	1 (0.0265%)	0 (0%)
11	6	6 (0.159%)	0 (0%)	0 (0%)
12	1	1 (0.0265%)	0 (0%)	0 (0%)
13	0	0 (0%)	0 (0%)	0 (0%)
14	0	0 (0%)	0 (0%)	0 (0%)
15	0	0 (0%)	0 (0%)	0 (0%)

[a] Chance of finding is % probability of finding an abortive of that length with a complimentary sequence within the same transcript = n/3780 × 100%.

Download English Version:

https://daneshyari.com/en/article/5590071

Download Persian Version:

https://daneshyari.com/article/5590071

Daneshyari.com