ALTERNATIVE TRANSCRIPTION
WITHIN PROCARYOTIC GENES PREDICTED
BY PROMOTER-SEARCH SOFTWARE

Shavkunov K.S.¹, Masulis I.S.¹, Matushkin Yu.G.², Ozoline O.N.¹
¹ Institute of Cell Biophysics, RAS, Pushchino, Moscow region, Russia; ² Institute of Cytology and
Genetics, SB RAS, Novosibirsk, 630090, Russia
* Corresponding author: e-mail: ozoline@icb.psn.ru

Key words: Escherichia coli, alternative transcription, promoters

SUMMARY

Motivation: Mapping of putative promoters within entire genome of Escherichia coli
by means of pattern-recognition software PlatProm revealed several thousands of sites
having high probability to perform promoter function. Along with the expected promoters
located upstream from coding sequences PlatProm identified several hundred of very
similar signals within coding sequences. Many of them may initiate transcription from the
sense strand thus permitting synthesis of shortened mRNA products, not expected
a priori in bacterial cells.

Results: Here we discuss possible functional significance of intragenic promoters,
estimate predictive capacity of our software in vivo and in vitro and provide
experimental evidences that at least one promoter predicted within coding sequences is
transcriptionally active.

Availability: Coordinates of predicted transcription start points for alternative
transcription are available by request (ozoline@icb.psn.ru).

INTRODUCTION

Genome-wide scanning by PlatProm revealed 709 genes, which have potential internal
promoters with a propensity to produce shortened RNA products from the sense strand
(Brok-Volchasinski et al., 2005). At least 46 of them may initiate synthesis of RNAs
previously detected in the fraction of small RNAs extracted from bacterial cells (Vogel et
al., 2003) and considered as products of mRNA degradation. Basically, the presence of
intragenic promoters may be required to intensify downstream transcription of neighboring
genes (if they have proper orientation) or trap RNA polymerase near real promoters (if they
are located at the beginning of the gene (Huerta, Collado-Vides, 2003)). However many
predicted promoters lie far from the 5′-end of gene, while the nearest downstream genes
have opposite orientation. In these cases internal promoters may be required to express
alternative proteins or antisense RNAs to the products of neighboring gene. We, therefore,
verified this possibility using available software (ORF Finder and RNA Structure), which
allowed identifying open reading frames (ORFs) and characterizing folding propensity of
putative RNA product. The scores of the transcription signals found within such genes were
compared with known promoters and transcription activity of the promoter, predicted in the
middle of the htgA gene was verified experimentally.
METHODS AND ALGORITHMS

The search for alternative ORFs was done using ORF Finder (www.ncbi.nlm.nih.gov). Transcription terminators were found on the basis of the following criteria: 5–10 bp G/C-rich stem, 3–8 bases loop, free energy < -7 kcal/mol, ≥ 4U downstream of the stem (Argaman et al., 2001). Folding propensity of potential RNA products was estimated by means of RNA Structure algorithm supplied with thermodynamic scoring system (http://rna.chem.rochester.edu).

Transcription activity of predicted promoters in vivo was tested using the total fraction of cellular RNAs isolated from cells taken during exponential and stationary growth phases. cDNA copies of target products were obtained by primer extension using RevertAid M-MuLV reverse transcriptase (Fermentas) and 32P-labeled gene-specific primers. cDNA products were separated from substrates using electrophoresis in 8 % polyacrylamide gel in the presence of 8M urea and visualized by radioautography.

Potassium permanganate footprinting was performed as described (Zaychikov et al., 1997). RNA polymerase – promoter complexes were formed at 36 °C in buffer, containing 50 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 0.1 mM DTT, 10 mM MgCl2, 50 mM NaCl and BSA (5 mg/ml). RNA polymerase was reconstituted from individual subunits, as suggested by Fujita and Ishihama (1996).

RESULTS AND DISCUSSION

PlatProm identified 709 genes containing strong promoter-like signals, from which shortened RNA products potentially may be synthesized. Most of them are expressed as independent transcription units or are the last genes of operons, thus suggesting that some internal promoters may be required to transcribe new genes in intergenic loci or intensify the expression of properly oriented downstream genes. The average size of intergenic regions flanking 3′-ends of such genes is slightly smaller (119 bp) than throughout the whole genome (~150 bp). In general that argues against the first assumption but is in line with the second one. Thus orientations of downstream genes were examined and it was found that two neighboring genes have similar direction in 318 cases.

Transcription signals within remaining 391 genes may be required to synthesize RNAs with a capacity to encode shortened proteins or untranslated RNA products. That is why the sequences located downstream from predicted promoters were further analyzed to reveal alternative ORFs. For this purpose we used nucleotide sequences spanning from predicted promoters to the ends of genes and increased the length of each sequence by 150 bp downstream so as to take into consideration possible terminators located in intergenic regions. Shortened ORFs were found within 305 genes and in 175 cases they are supplied with suitable ribosome binding site (at least 4 matches to AGGAGGT). This set of internal promoters probably has the highest heuristic significance. Fig. 1 represents their scores in comparison with scores of 328 known bacterial promoters, which were absent in compilation used to generate weight matrices of PlatProm. One can see that there are many real promoters with low scores, however in most cases the values of S deviate from the background level for more than 3 Std. Surprisingly we found that the distribution of S has two well pronounced maxima (4.5 ≤ S ≤ 5.5 and 7.5 ≤ S ≤ 9.5). It could be speculated that corresponding promoters are subjected to different types of regulation. For instance, the set of weaker promoters may require transcription activators for maximal activity, while stronger promoters may be constitutive or their functionality may depend on repressors. In any case, for predictive mapping we used only strong transcription signals (4 Std higher than background level). Fig. 1 demonstrates that distribution of scores for predicted internal promoters also has maximum; positioning of this maximum is the same as in the case of real promoters and there is a fraction of very strong transcription signals (S ≥ 13.5, or 6 Std
Part 1

higher than background). That means that the set of predicted promoters have some features similar with real promoters. All of them can not be ascribed to any known gene.

Fig. 2 exemplifies such situation within gene *htgA*. It encodes positive regulator for promoters recognized by σ32 (heat shock regulon) and may be expressed from the σ32-dependent promoter, located 82 bp upstream of the initiating codon of *htgA* (unrecognized by PlatProm) or from the weak σ70-specific promoter, situated 114 bp upstream of ORF. *HtgA* lies between *yaaH* and *yaal*, transcribed from the opposite strand and fully overlaps with the putative gene b0011. At least 3 promoter-like sites are predicted in this region. The strongest one most probably controls transcription of *yaaH*. Others were found within *htgA* and are probably required to produce both antisense or alternative RNA products in respect to genes *htgA* and *b0011*, as well as an alternative ORF found at the end of *htgA*.

*Figure 1.* Distribution of scores (S) for 328 bacterial promoters (open circles) and 175 internal promoters (black circles), having a propensity to produce alternative mRNAs. Each point represents the number of nucleotide sequences, having scores within an interval S-0.5 ÷ S. An average S for non-promoter DNAs, estimated by PlatProm was -4.85. Arrows indicate levels of S, which are 3 and 4 standard deviations (Std) larger than this value. Only signals with S ≥ 4 Std were used for predictive mapping. Both curves were smoothened using running window 3.

*Figure 2.* Schematic representation of the chromosome locus, containing gene *htgA*. Solid black lines and arrows drown above or below X axis show positioning of genes in respect to the initiating codon of *htgA* and respective direction of transcription. Bars represent promoters predicted on both strands. Open rectangle indicates location of alternative ORF, while zigzag lines show putative RNA products, which may be synthesized between the predicted promoter and the first r-independent terminator.

BGRS’2006
Activity of two intrinsic promoters located on the top strand of *htgA* was verified experimentally. They form two clusters and may provide RNA products 92–108 and 62–66 nt long. Primer 2 (Fig. 2) was used for the reaction of reverse transcription to detect the expected RNA products in total fraction of cellular RNAs. They were isolated from cells at exponential and stationary growth phases. At least three short RNAs: 92, 69 and 67 nt long were detected in addition to longer products, originated from upstream promoters (Fig. 3a). Their abundance does not depend on the growth phase. Although sizes of these products are very close to the expected ones, some of them may be products of mRNA degradation. That is why we used the potassium permanganate footprinting technique to answer the question whether RNA polymerase forms open promoter complexes in these regions (Fig. 3b).

*Figure 3.* Experimental verification of promoter activity by means of primer extension (a) and potassium permanganate footprinting (b). (a): Primer 2 and 1 ng of cellular RNA were used to obtain cDNA copies. cDNA products were separated on polyacrylamide gel (8%) and visualized by radioautography. Arrows on the right indicate observed shortened RNAs. (b): PCR amplified DNA fragment (Primers 1 and 2 shown in Fig. 1) was used to study an RNA polymerase binding capacity to predicted promoters. Complexes were formed as described in Methods and Algorithms. RNA polymerase – promoter ratio was 1:4(M:M). Marks “-” and “+” denote samples, containing free DNA fragment and DNA-protein complexes, respectively. Arrows on the right indicate bands representing the specific modification of unpaired thymines. Both gels were calibrated by standard G-specific ladder of another DNA fragment. Ciphers on the left reflect the sizes of indicated fragments.
Modifying only unpaired thymines, the potassium permanganate provides an excellent opportunity to reveal transcriptional bubble and, therefore, to localize specific RNA-polymerase binding site(s) on DNA. The data obtained clearly indicate that in vitro open complexes are really formed and the binding site is located near the cluster of predicted transcription start points with high scores (92–108 bp upstream of primer 2). There are no any reactive thymines near weaker transcription signals (62–66 bp far from the primer) thus indicating that RNA polymerase selects more strong promoter site, while two products detected in the reaction of primer extension may result from RNA decay. Fig. 4 shows the nucleotide sequence of the region containing active promoter.

The observed pattern of reactivity against potassium permanganate allows a possibility that RNA polymerase can initiate transcription from all three predicted start points in the cluster (genomic coordinates are: 11090, 11099 and 11102 on the + strand). The major product observed in vivo (Fig. 3a) is, however 6–3 nt shorter than expected in the case if RNAs are initiated from promoter having almost perfect -35 and -10 elements (98 and 95 bp from primer), and more pronounced transcription bubble. Weaker upstream promoter with a capacity to give 107 nt RNA (initiated from position 11090) also binds RNA polymerase; although in vivo the product of exactly this length was observed only upon longer exposition. In any case, the internal promoter predicted within gene htgA is active. RNA transcribed from this promoter may encode a 31 amino acids long polypeptide, with ORF shifted on 1 position in respect to mRNA of htgA. This product has no sequence homology with any other known protein. Alternatively 158 nt long RNA transcribed between the verified promoter and the first ρ-independent terminator may function as antisense RNA to mRNA of hypothetical protein b0011. Free energy of folding for this transcript (-57 kcal/M) is typical for small regulatory RNAs of this length.

Figure 4. Nucleotide sequence of the predicted internal promoter. Lower case letters indicate predicted start points of transcription. Suitable -35 regions are underlined; -10 regions are shown by larger font letters. Adenines complementary to thymines, modified by potassium permanganate are double underlined. Ciphers above the sequence indicate the expected length of the product originated from marked position.

Taken together we conclude that PlatProm may be used as a tool predicting novel transcripts in the genome of E. coli.

ACKNOWLEDGEMENTS

The studies are supported by the RFBR(03-04-48339) and RFBR-naukograd (04-04-97280).

REFERENCES


