MODELING CLASSIC ATTENUATION REGULATION OF GENE EXPRESSION IN BACTERIA

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

Key words: classic attenuation regulation, transcription elongation, RNA-polymerase termination, translation initiation and elongation, energy of secondary structure, transition between secondary structures

SUMMARY

Motivation: Attenuation regulation, particularly, in its classic form is well described on comparative genomic level using evidence from both datamining and experiment. Even being confronted with difficulties in choosing adequate parameter settings, developing a rigor and effective computer model of each attenuation type is a timely and important task. Such a model is prerequisite to *in-silico* choose between alternative hypotheses of any gene leader region, as well as to study attenuation mechanisms in conjunction and in comparison with the mechanism of protein-DNA interaction (repressor-activator).

Results: An effective computer model of classic attenuation regulation is developed. The model is based on rigor and explicit statements (viz., description of all correlations and parameter value settings), which provides for its greater accuracy to explain experimental data. Results of computations reveal qualitatively correct correlations between termination probability and amino acid concentration for leader regions with predicted attenuation. When applied to random sequences, the model produces correlation values varying around a certain constant, which indicates the lack of regulation.

MODEL

The approach is based on modeling RNA secondary structure in the regulatory region between the ribosome and RNA-polymerase, resonant equations of the RNA-polymerase inhibition by helices in this region (equation for F given bellow), modeling transcription and translation initiation and elongation. Microstate is a set of continuous fragments, referred to as *hypohelices*, of any non-continuable *helices* in the same region. The model describes transitions between microstates: decomposition and binding of *hypohelices* in the same region.

The rate constant of transition between microstates ω and ω' is calculated as follows:

$$K(\omega \to \omega') = \kappa \cdot \exp[\frac{1}{2}\Big((G_{loop}(\omega) + G_{hel}(\omega)) - (G_{loop}(\omega') + G_{hel}(\omega'))\Big)] .$$

Here $R \cdot T \cdot (G_{loop}(\omega) + G_{hel}(\omega))$ is free energy of RNA secondary structure (loops and helices) in microstate ω, R – *universal* gas constant, T – *absolute* temperature (for details ref. to Mironov, Lebedev, 1993).

Two microstates ω and ω' belong to one *macrostate* Ω if both ω and ω' are realized by identical diagram (for definitions of *diagram* and its *chord* ref. to Lyubetsky *et al*., 2006); intuitively it means that both $K(\omega \to \omega')$ and $K(\omega' \to \omega)$ are relatively large. It was our aim to achieve that transitions between two microstates ω and ω' within any macrostate Ω are fast, and those between any microstates ω and ω' from different Ω and Ω', respectively, are slow.

Absolute probabilities of transitions between microstates ω and ω' in macrostate Ω are inessential in our model. Instead, transitions in the set of all microstates ω in any macrostate Ω are required to produce Boltzmann-Gibbs stationary probability distribution:

$$p(\omega) = \frac{\exp\left(-(G_{loop}(\omega) + G_{hel}(\omega))\right)}{z(\Omega)}, \text{ where}$$

$$z(\Omega) = \sum_{\omega \in \Omega} \exp\left(-G_{loop}(\omega) - G_{hel}(\omega)\right).$$

Trivial averaging over all pairs of microstates ω in Ω and ω' in Ω' produces the following equation for the transition rate constant between macrostates Ω and Ω' that applies to both increase and decrease of macrostate by *one chord*: $K(\Omega \to \Omega') = \sum_{\omega \in \Omega} \sum_{\omega' \in \Omega'} p(\omega) \cdot K(\omega \to \omega')$. All other transitions between macrostates are null.

The rate constant of polymerase transition from one nucleotide to the next is calculated as $\nu(\Omega) = \overline{\lambda}_{pol} - F(\Omega)$, where Ω is a macrostate, and $F(\Omega)$ is effective decrease of the polymerase rate constant in s^{-1}. In the model, polymerase deceleration by hairpin ω is described as follows:

$F(\omega) = \frac{\delta}{L_1^2 \cdot (p(\omega) - p_0)^2 + 1} \cdot \exp\left(-\frac{r}{r_0}\right)$, where *r* is distance between the terminus of hairpin and the polymerase. *Parameters* L_1, p_0, r_0, δ depend on polymerase characteristics and value $p(\omega)$ – on hairpin ω. For a hairpin consisting of the handle and the loop, *p* is estimated from the equation: $\mathrm{tg}(p \cdot h) = \frac{2}{p \cdot l}$, $0 < p \cdot h < \frac{\pi}{2}$, where *h* is *handle length*, i.e. the number of its base pairs, and *l* is *loop length*. An analogous equation is used for an arbitrary hairpin.

The rate constant of the polymerase sliding within a T-rich region is estimated as $\mu(\Omega) = F(\Omega)/4$ (Yin *et al*., 1999).

On non-regulatory codons, the rate constant λ_{rib} of ribosome elongation by 1 nucleotide is $\overline{\lambda}_{rib} = 45 s^{-1}$. On regulatory codons, λ_{rib} depends on concentration *c* of aminoacyl-tRNA according to the Michaelis-Menten law: $\lambda_{rib}(c) = \frac{\overline{\lambda}_{rib} \cdot c}{c_0 + c}$.

To model obstacles in ribosome binding, we incorporated ribosome binding rate constant $K_0 = \lambda_0 \cdot \frac{d_{open}}{d_{\max}}$, where d_{open} is current value of the maximum number of open nucleotides in the Shine-Dalgarno sequence (provided that the start codon is open), $d_{\max}$ – the length of the sequence and λ_0 – translation initiation parameter.

Standard Monte-Carlo technique is used in modeling. For example, neighborhood of given state Ω, centered in Ω, is a set of all states Ω' with non-zero probability of transition from Ω to Ω' by both increase and decrease of macrostate Ω. If given neighborhood contains *n* states and corresponding transition rate constants are k_1, ..., k_n, the next state on the trajectory of transitions is determined by realizing random variable $i \to \frac{k_i}{\sum k_i}$.

The following parameter settings were chosen: $\kappa = 10^3$, r_0 within the range 2–8, $L_1 = 14.5$, $p_0 = 0.167$, $\delta = 25$, $\kappa = 10^3$ s^{-1}, $c_0 = 1$. "Sizes" of ribosome and polymerase are $s_0 = 12, s_1 = 5$.

The purpose of modeling was estimating function $p = p(c)$ of correlation between termination probability and concentration c of amino acid or concentration c of aminoacyl-tRNA synthetase for operon leader regions in bacteria. These estimated were also obtained for random sequences (see below). Function $p(c)$ was estimated with repeating the modeled process certain number of times (usually 10^3–10^4) under given c increment, and $p(c)$ was calculated as a fraction of times when termination occurred.

Computer assays were "positive" when all available regions with putative attenuation (using evidence e.g. from Vitreschak *et al*., 2004) were analyzed under fixed values of the above described parameters. The assays were "negative" under the same parameter settings when modeling was done with "*random*" *sequences* assembled from the leader peptide gene upstream of *trpE* in *Vibrio cholerae*, a U-rich terminator from the same leader region and a random sequence in four-letter alphabet of random length intercalating the two. Positive assays were expected to return approximately monotonous growth of function $p(c)$, while negative – to demonstrate its absence. All positive assays, except for tryptophan biosynthesis operons in *Streptomyces* spp., returned approximately monotonous growth (ref. to Results), and all negative – oscillations around different constants.

RESULTS AND DISCUSSION

Values in the Table were obtained by computing with our model on leader regions upstream of gene *trpE* in *Corynebacterium diphtheriae*, *Corynebacterium glutamicum*, *Agrobacterium tumefaciens*, *Bradyrhizobium japonicum*, *Rhodopseudomonas palustris*, *Rhizobium leguminosarum*, *Sinorhizobium meliloti*, *Escherichia coli*, *Vibrio cholerae*, and also *for* gene *trpS* in *Streptomyces avermitilis*. The results are in congruence with multiple alignments of corresponding leader regions, which are available in publications for actinobacteria (Seliverstov *et al*., 2005) and proteobacteria (Vitreschak *et al*., 2004).

For *C. glutamicum*, termination probability estimated in the model doubles under tryptophan concentration growth but still was very low. For some alpha-proteobacteria, modeled termination probability increases considerably: 48-fold in *R. palustris*, 7.6-fold in *S. meliloti*, and 16.6-fold in *V. cholerae*. The ranges decrease under κ growth, and in this sense their interpretation is unclear.

Some rows of the Table represent not strictly monotonous pattern. This might be accounted for by precision of modeling being below 0.01–0.02, which also depends on characteristics of the random seed generator. Classic attenuation is applicable within specific intervals of the c value that are determined individually for each gene and organism. Small size of such interval measured in the model in c / c_0 does not necessarily imply small physical values, e.g. in mM/l. Also, bacteria in favorable natural environment do not display strictly monotonous function $p(c)$. The results (presented partially) reveal

correlation $p(c)$ congruent on the qualitative level with presence of attenuation in most gene leader regions studied. All negative assays (data not shown) returned $p(c)$ values oscillating around certain constant.

Thus, our model can be used to predict the impact of point mutations in regulatory regions on attenuation regulation and to predict stability of this system during the course of evolution. It can also be incorporated into a broader non-linear model of bacterial metabolism with dynamic modeling of gene expression regulation. Another possible application of the model is prediction of attenuation regulation by modeling correlation between the enzyme activity and amino acid concentration for a *single sequence*, thus eliminating the need to analyze sequence profiles.

A method is proposed to objectively choose model settings on the basis of source data. The computer program offers high flexibility to vary all model parameters and correlations. The model was applied to biological data to assess its relative robustness against varying parameter settings and to obtain their estimates using Monte-Carlo approximations of typical stem lengths, macro- and microstate ratios, lengths of the ribosome and polymerase neighboring-state transition cycles, etc.

Table 1. Termination probability $p(c)$ against concentration c of triptophanyl-tRNA in various bacteria

Species	Concentration c										
	0.00	**0.05**	**0.10**	**0.15**	**0.20**	**0.25**	**0.30**	**0.35**	**0.40**	**0.45**	**0.50**
C. diphtheriae	0.34	0.34	0.39	0.46	0.50	0.54	0.53	0.53	0.53	0.52	0.54
C. glutamicum	0.05	0.06	0.08	0.10	0.10	0.09	0.09	0.09	0.10	0.10	0.10
S. avermitilis, trpS	0.06	0.13	0.21	0.26	0.28	0.29	0.30	0.30	0.32	0.32	0.30
A. tumefaciens	0.49	0.50	0.62	0.70	0.74	0.78	0.77	0.78	0.82	0.80	0.79
B. japonicum	0.19	0.20	0.24	0.26	0.28	0.26	0.26	0.27	0.26	0.26	0.26
R. leguminosarum	0.23	0.30	0.42	0.55	0.60	0.65	0.67	0.70	0.71	0.71	0.71
R. palustris	0.01	0.22	0.40	0.48	0.56	0.59	0.60	0.60	0.63	0.61	0.62
S. meliloti	0.07	0.11	0.23	0.37	0.43	0.49	0.48	0.51	0.50	0.53	0.51
E. coli	0.34	0.46	0.54	0.68	0.70	0.70	0.71	0.73	0.75	0.75	0.74
V. cholerae	0.05	0.16	0.39	0.57	0.70	0.74	0.77	0.77	0.80	0.79	0.81

ACKNOWLEDGEMENTS

The authors are grateful to A.A. Mironov and L.Y. Rusin for discussion and help. This work was partly supported by ISTC 2766.

REFERENCES

Lyubetsky V.A., Rubanov L.I., Seliverstov A.V., Pirogov S.A. (2006) Model of genes expression regulation in bacteria by means of formation of secondary RNA structures. *Mol. Biology*, **40**, 497–511.

Mironov A.A., Lebedev V.F. (1993) A kinetic model of RNA folding. *BioSystems*, **30**, 49–56.

Seliverstov A.V., Putzer H., Gelfand M.S., Lyubetsky V.A. (2005) Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiology*, **5**, 54.

Vitreschak A.G., Lyubetskaya E.V., Shirshin M.A., Gelfand M.S., Lyubetsky V.A. (2004) Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS Microbiology Letters*, **234**, 357–370.

Yin H., Artsimovitch I., Landick R., Gelles J. (1999) Nonequilibrium mechanism of translation termination from observations of single RNA polymerase molecules. *PNAS*, **96**, 13124 – 13129.