

Combined sequenced-based model of the *Drosophila* gap gene network

Dymova A.V., Kozlov K.N.

Peter the Great St. Petersburg Polytechnic University, Politekhnikeskaya St., 29, St. Petersburg, Russia

e-mail: dymova.arina@mail.ru, kozlov_kn@spbstu.ru

Gursky V.V.

Ioffe Institute, Politekhnikeskaya St., 26, St. Petersburg, Russia

e-mail: gursky@math.ioffe.ru

Samsonova M.G.

Peter the Great St. Petersburg Polytechnic University, Politekhnikeskaya St., 29, St. Petersburg, Russia

e-mail: m.samsonova@spbstu.ru

Transcription is vital for normal functioning of all organisms, as it is the most significant regulation level of gene expression. An interest in detailed analysis of transcriptional regulation grows due to new large-scale data acquisition techniques. There is a large amount of experimental data available about the *Drosophila* segment determination system. The gap gene system implements the most upstream regulatory layer of the segmentation gene network. It receives inputs from long-range protein gradients encoded by maternal coordinate genes and establishes discrete territories of gene expression. In this process the gap gene cross-regulation plays important role. The formation of gap gene expression domains is a dynamic process: the domains do not form in one place, but instead in the posterior half of the embryo they shift anteriorly during cleavage cycle 14. Despite our expanding knowledge of the biochemistry of gene regulation, we lack a quantitative understanding of this process at a molecular level.

We applied the systems-level approach to study gap gene network in *Drosophila*. A combined sequence-based model of gap gene regulatory network controlling segment determination in the early *Drosophila* embryo was developed. We modified the recent model from [1] in three ways. First, we narrowed down the spatial domain of the model by considering only the trunk region of blastoderm from 35% to 92% of embryo length along the A-P axis. This allowed us to combine the data on protein concentration from FlyEx database [2] and on mRNA concentration from SuperFly database [3] in the fitting procedure. Second, we used

the same BTM parameter and the same range for short range repression for all targets and TFs. Third, we allowed several activator molecules to contact BTM simultaneously.

The state variables of the model are the concentrations of mRNAs and proteins encoded by four gap genes *hb*, *Kr*, *gt*, and *kni*. The binding sites for 8 TF – the products of *hb*, *Kr*, *gt*, *kni*, *bcd*, *tll*, *cad*, and *hkb* genes in the potential regulatory region from 12Kbp upstream to 6Kbp downstream of TSS for each gene were predicted in [1] according to [4]. We modified thermodynamic approach originally proposed in [5]. The target gene expression at mRNA level is proportional to the promoter occupancy that is determined by the concentration of TFs. The dynamics mRNA and protein concentrations is described by the system of differential equations with the delay parameter that accounts for the average time between events of transcription initiation and corresponding protein synthesis.

The promoter occupancy of every target gene for every embryo nucleus at every point of time was calculated in the form:

$$E_i^a(t) = \frac{\sum_{\sigma} W_i^a(\sigma, t) Q^a(\sigma)}{\sum_{\sigma} W_i^a(\sigma, t) Q^a(\sigma) + \sum_{\sigma} W_i^a(\sigma, t)} \quad (1)$$

where σ is a molecular configuration of the regulatory region for gene a , $Q^a(\sigma)$ is the statistical weight of the interaction between TFs and bound basal transcriptional machinery (BTM), and $W_i^a(\sigma, t)$ is the statistical weight of configuration σ for nucleus-time coordinate (i, t) , that depends on the concentration v_i^b of all TFs regulating gene a in nucleus i at time moment t . The system of differential equations considers both mRNA and protein synthesis.

The equation for mRNA concentration $u_i^a(t)$ of target gene a in nucleus i included production, diffusion and decay terms. The equation for protein concentration $v_i^a(t)$ described protein synthesis, diffusion and degradation.

$$du_i^a / dt = R_u^a E_i^a(t) + D_u^a(n)[(u_{i-1}^a - u_i^a) + (u_{i+1}^a - u_i^a)] - \lambda_u^a u_i^a \quad (2)$$

$$dv_i^a / dt = R_v^a u_i^a(t - \tau_v^a) + D_v^a(n)[(v_{i-1}^a - v_i^a) + (v_{i+1}^a - v_i^a)] - \lambda_v^a v_i^a \quad (3)$$

where n is the cleavage cycle number, R_v^a and R_u^a are maximum synthesis rates, D_v^a , D_u^a are the diffusion coefficients, and λ_v^a and λ_u^a are decay rates for protein and mRNA of gene a . The parameter τ_v^a is the delay parameter.

Model fitting was carried out by differential evolution (DEEP) method [6] to minimize the combined objective function. This function is a sum of the residual sum of squared differences between the model output and data, weighted pattern generating potential and a penalty term, which limits a growth of regulatory weights. The weighted pattern generating potential was proposed in [7] to account not only for the magnitude of difference between model and data, but also for the direction of change.

Concentration profiles for mRNA and protein for 4 target genes – *hb*, *Kr*, *gt* and *kni* are presented in Fig. 1 together with experimental observations from FlyEx and SuperFly used for fitting.

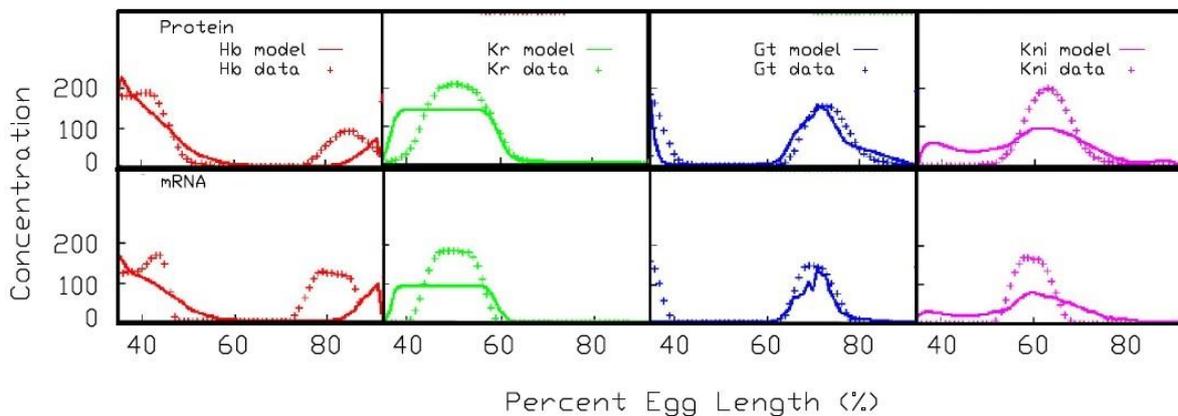


Figure 1. Model output for a representative network as compared to protein concentration profiles from the FlyEx database (upper row) and mRNA concentration profiles from the SuperFly database (bottom row). Results are shown for time class 6 in cleavage cycle 14A.

The model correctly reproduces the majority of characteristic features of gap gene expression patterns at protein and mRNA level. The reasons for the discrepancies between model output and experimental observations will be further investigated. The advantage of the presented model in comparison with the previous one [1] is the fact that the promoter occupancy is calculated using the model output for TF concentrations and not the experimental data.

References

1. Kozlov K, Gursky V, Kulakovskiy I, Samsonova M, Sequence-based model of gap gene regulatory network, *BMC Genomics* 15(Suppl 12):S6, 2014.
2. Pisarev, A., Poustelnikova, E., Samsonova, M., Reinitz, J.: FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucleic Acids Research* 37, 560–566 (2008).
3. Cicin-Sain D, Pulido AH, Crombach A, Wotton KR, Jiménez-Guri E, Taly JF, Roma G, Jaeger J. SuperFly: A comparative database for quantified spatio-temporal gene expression patterns in early dipteran embryos. *Nucleic Acids Res* 43:751-755. 01/01/2015. F.I.: 8.808.
4. I.V. Kulakovskiy and V.J. Makeev. Discovery of dna motifs recognized by transcription factors through integration of different experimental sources. *Biophysics*, 54(6):667–674, 2009.
5. He, X., Samee, M.A.H., Blatti, C., Sinha, S.: Thermodynamics-based models of transcriptional regulation by enhancers: the roles of synergistic activation, cooperative binding and short-range repression. *PLoS Comput. Biol.* 6(9) (2010).
6. Kozlov, K., Samsonov, A.: Deep – differential evolution entirely parallel method for gene regulatory networks. *Journal of Supercomputing* 57, 172–178 (2011)
7. Samee, M.A.H., Sinha, S.: Evaluating thermodynamic models of enhancer activity on cellular resolution gene expression data. *Methods* 62, 79–90 (2013)