Stochastic modeling of enhancer molecular configurations

German Demidov

St. Petersburg Academic University, 8/3 Khlopina Str., St. Petersburg, 194021 Russia; and

Parseq Lab, 16 Birzhevaya Liniya, St. Petersburg, 199053 Russia

gdemidov@parseq.pro

Vitaly V. Gursky

Ioffe Institute, 26 Politekhnicheskaya Str., St. Petersburg, 194021 Russia; and

St. Petersburg Polytechnic University, 29 Politekhnicheskaya Str., St. Petersburg, 195251 Russia

gursky@math.ioffe.ru

Eukaryotic gene expression is regulated via *cis*-regulatory modules (enhancers) in DNA, consisting of transcription factor binding sites (TFBSs) and their clusters. These modules can be hundreds bases long and kilobases away from the basal promoter of the target gene. The precise molecular mechanisms of how enhancers control gene expression remain unknown for many systems. A promising approach for modeling eukaryotic gene expression is based on the thermodynamic models [1, 2]. These models utilize the statistical thermodynamics formalism to connect the probabilities of various molecular configurations that the enhancer may take with the probability of the target gene activation and start of transcription. The enhancer molecular configuration in this formalism is a particular combination of free and bound TFBSs on the enhancer, and the probabilities of all possible configurations are assumed obeying the Boltzmann distribution. The distribution depends on the binding energy for specific TFBSs, calculated by using the positional weight matrix (PWM) approach.

The use of the Boltzmann distribution for the enhancer states provides an approximation that neglects the history dependence of these states. This history involves the dynamical processes of transcription factor molecule transport to the DNA, non-specific binding to the DNA and unbinding from it, one-dimensional sliding of the molecule along the DNA, stochastic hops between distal parts of the sequence, the interaction of the molecule with other molecules already bound to their specific sites, etc. These dynamical processes provide paths to specific asymptotic states of the enhancer, in which specific TFBSs are occupied and the enhancer approaches a quasi-equilibrium. Therefore, taking into account the formation of the enhancer molecular configuration as the dynamical process should bias the resulting distribution of

enhancer asymptotic states, which is then connected to the probability of transcription activation.

In this study, we present a stochastic model for the dynamical formation of enhancer configurations and test it on the regulatory region of gap gene *giant* in *Drosophila*, one of the segmentation genes controlling the embryo development, and transcription factors Giant and Kruppel. As the simplest approximation, we have developed a model that considers transcription factors as segments that can bind to DNA and the PWM-scores as an estimation of their binding strengths. The model parameters include concentrations of transcription factors, different rates of binding and sliding along the DNA depending on the number of molecules, coordinates of open chromatin, etc. We have estimated parameter values from experimental data when possible (like average sliding length) and used reasonable approximations otherwise (like a sigmoidal function for probabilities of various elementary processes) [3, 4]. Based on simulation results, we identify regimes where the model predicts deviation from the Boltzmann distribution for the probabilities of possible enhancer configurations.

The study was partially supported by the RFBR grant 13-01-00405.

1. J.M.Dresch, M.Richards, A.Ay (2013) A primer on thermodynamic-based models for deciphering transcriptional regulatory logic, *Biochimica et Biophysica Acta*, **1829:**946–953.

2. K.N.Kozlov et al. (2014) Sequence-based model of gap gene regulatory network, *BMC Genomics*, **15**(Suppl 12):S6.

3. E.F.Koslover, M.A.Diaz de la Rosa, A.J.Spakowitz (2011) Theoretical and computational modeling of target-site search kinetics in vitro and in vivo, *Biophysical Journal*, **101:**856–865.

4. E.G.Marklund et al. (2013) Transcription-factor binding and sliding on DNA studied using micro- and macroscopic models, *PNAS*, **110**(49):19796–19801.