

RESEARCH DESCRIPTION

Marko Djordjevic

Tel: (614) 614-688-3198

Fax: (614) 247-6643 231

E-mail: mdjordjevic@math.ohio-state.edu

The Ohio State University

Mathematical Biosciences Institute

1735 Neil Avenue

Columbus, OH 43210

My research is broadly in the area of quantitative systems biology, biophysics and bioinformatics. Most of my current work aims to quantitatively understand how genes are expressed, and how this process is regulated. One direction in my research is to use theoretical methods (e.g. thermodynamics, statistical mechanics, dynamical systems) to model biological systems and processes, which is often referred to as quantitative systems biology. My other research direction is bioinformatics, i.e. developing methods that allow analyzing large amounts of data assembled in molecular biology.

The above research directions are most often complementary to each other, since the development of bioinformatic methods can be greatly aided by physical modeling of the underlying biological processes. Conversely, an ability to accurately analyze complex data that became available after the genomic revolution in biology is essential for developing realistic quantitative theories of biological systems. Finally, my theoretical/computational research is also aimed at contributing to the experimental design, i.e. at proposing experiments that would generate optimal data-set for further analysis. To accomplish a close interaction between theory and experiment, I have been intensively collaborating with experimental biology labs. Below is given a brief summary of my relevant research projects.

1. *Quantitative modeling of transcription initiation by bacterial RNA polymerase.* RNA polymerase (RNAP) is a central enzyme in cell, which is responsible for gene transcription. As a first step of transcription initiation, RNAP binds to double stranded DNA and opens the two strands of DNA, which is referred to as the open complex formation. By using the methods from theoretical biophysics, I developed the first quantitative model of the open complex formation by bacterial RNAP, which establishes an explicit relationship between the rate of transcription initiation and physical properties of promoter sequence and promoter-RNAP interactions. Comparison of the model with both biochemical measurements and genomics data leads to a very good agreement with the experiments, with no free parameters used in model testing. This agreement strongly supports both our quantitative model and a

qualitative mechanism on which the model is based. In a current work in progress, I am using the model to investigate kinetic properties of promoter sequences on the whole genome scale (M Djordjevic and R Bundschuh, *Biophys. J.*, in press, 2008; M Djordjevic, to be submitted to *IET Systems Biology*, (invited paper)).

2. *Modeling of gene regulatory circuits.* Restriction modification (R-M) gene switches are rudimentary immunological systems in bacteria, which are often mobile, thus propagating through bacterial populations and impacting genome evolution on a global scale. To achieve regulation of the R-M system, genes in the switch are under the control of a transcription factor (the control protein) whose gene is also part of the system. In collaboration with experimentalists, I used thermodynamics modeling in order to develop a mathematical description of how the genes in the system are regulated by the control protein, which showed a very good agreement with the experimental data. I further used dynamical system modeling in order to connect the proposed architecture of the R-M system with its presumed *in-vivo* biological function, as well as to elucidate design principles behind the switch architecture. (E Bogdanova, M Djordjevic, I Papapanagiotou, T Heyduk, G Kneale, K Severinov, *Nucl. Acids Res.*, in press, 2008).
3. *Functional versus non-functional transcription factor binding sites in genome.* The goal of the project is to address how transcription factors (proteins that bind to DNA and regulate gene expression) distinguish between functional and non-functional binding sites in the vast amount of genomic DNA. To address this issue two lines are followed: First, a bioinformatic analysis of genome sequences is performed, by using the protein-DNA interaction parameters that are inferred from a novel algorithm that we developed. Second, I performed a quantitative analysis of ChIP-chip experiments, which allows detecting genome-wide binding patterns of transcription factors. These two approaches should allow determining to what degree protein-DNA binding energy and physical binding of transcription factors to genomic DNA is related with the regulation of gene expression in higher eukaryotes (M Djordjevic and E Grotewold, in preparation).
4. *Modeling and data analysis of in-vitro selection experiments.* SELEX (Systematic Evolution of Ligands by EXponential enrichment) is an *in-vitro* selection procedure that allows the extraction, from an initially random pool of DNA, of those oligomers with high affinity for a given DNA-binding protein. We addressed what is a suitable experimental and computational procedure to infer parameters of transcription factor–DNA interaction from SELEX experiments. I developed a novel quantitative model of SELEX, and showed that the standard experimental procedure is unsuitable for obtaining accurate interaction parameters. However, from the model follows that a modified experiment in which chemical potential is fixed through different rounds of the experiment allows robust generation of an appropriate dataset. Based on the quantitative model, we proposed a novel bioinformatic algorithm for SELEX data

analysis. (M Djordjevic, AM Sengupta, *Phys. Biol.* **3**:13 2006; M Djordjevic, *Biomol. Eng.* **24**(2):179, 2007 (invited review paper)).

5. *Bacteriophage gene expression strategies*. An increasingly large number of bacteriophage (bacterial viruses) genomes are being sequenced each year. We address what is an efficient experimental and computational procedure to analyze gene expression strategies of these newly sequenced bacteriophages. On the example of a specific bacteriophage (Xp10), I developed a novel method of data analysis which combines kinetic modeling of the infection process with analysis of gene expression data and bioinformatic analysis of bacteriophage genome sequence. The method was subsequently generalized to other virulent bacteriophages, and we expect that genomic array experiments, together with the methods of data analysis that we developed, will provide an efficient way to analyze dynamics of gene expression for newly sequenced phages. (E Semenova, M Djordjevic, B Shraiman, K Severinov, *Mol Microbiol.* **55**:764, 2005; M Djordjevic, E Semenova, B Shraiman and K Severinov, *Virology* **354**(2):240, 2006; A Sevostyanova, M Djordjevic, K Kuznedelov, T Naryshkina, M Gelfand, K Severinov and L Minakhin, *J. Mol. Biol.* **366**(2):420, 2007).
6. *Transcription factor binding site detection*. We proposed a novel, biophysics based algorithm, for the supervised detection of TF binding sites. The method classifies potential binding sites by explicitly estimating the sequence-specific binding energy and the chemical potential of a given TF. In contrast with the widely used information-theoretic weight matrix method, our approach correctly incorporates saturation in the transcription factor/DNA binding probability. This results in a significant reduction in the number of expected false positives, and in the explicit appearance - and determination - of a binding threshold. The algorithm was applied to examine relationship between TF binding specificity and degree of pleiotropy. (AM Sengupta, M Djordjevic and BI Shraiman, *PNAS* **99**(4):2072 2002; M Djordjevic, AM Sengupta and BI Shraiman, *Genome Res.* **13**(11):2381, 2003).