**TITLE:** DECONSTRUCTION OF REGULATORY COMPLEXITY IN ESCHERICHIA COLI THROUGH SYSTEMIC BIOLOGY APPROACHES

**AUTHORS:** BIANCA NAOMI DE CARVALHO, RAFAEL SILVA-ROCHA

**INSTITUTION:** SYSTEMS AND SYNTHETIC BIOLOGY LAB, CELL AND MOLECULAR BIOLOGY DEPARTMENT, FMRP-USP (AVENIDA BANDEIRANTES 3900, PRÉDIO CENTRAL, SALA 22, RIBEIRÃO PRETO – SP, BRAZIL)

**ABSTRACT:**

*Escherichia coli* is the organism best investigated from the molecular point of view. For this organism, there are a large number of data referring to the mechanisms of regulation of gene expression that have been generated in the last decades. There are specialized databases that compile this information in a systematic way and manually cured, serving as a starting point for many studies in this organism. The objective, therefore, is to make use of this information through the use of computational approaches to map natural complex promoters of *E. coli* and to contrast the topological relationships of different promoters with the expression profiles of the target genes. For this, information on the regulatory interactions between transcription factors and target promoters in *E. coli* were extracted from the RegulonDB database version 9. That is, extraction of architectural information from promoters such as data about name of transcription factors and corresponding promoters, position of each promoter, site of interaction of transcription factors, as well as the effect activation or repression of these factors on target genes, make it possible to select naturally complex promoters of *E. coli*. In this sense, those promoters regulated by 5 or more transcription factors were selected to analyze the dynamics of gene expression with similar promoter architectures. We used python scripts to extract the information from the data collected from *E. coli*. it was possible to obtain a total of 227 complex promoters using the aforementioned criterion. The architectures of promoters regulated by 5 or more transcription factors were further analyzed and decoded using *ad hoc* scripts in the R platform. Thus, the list of selected promoters was mapped and group with the expression profiles of the target genes extracted from the Colombos Commons gene expression database, version 3. The latter has more than 4,000 experimental conditions analyzed and normalized to the more than 4,500 *E. coli* genes based on microarray and RNAseq data. Thus, through the analysis of the correlation between the architecture of natural complex promoters and the final levels of expression of the target genes, it was possible to identify which promoters grouped by architecture control genes that are grouped by the levels of expression, as well as the cases where this rule is not obeyed. For those cases where promoter clusters do not result in clusters of gene expression, the potential mechanisms related to the lack of correlation will be investigated in more detail, thus allowing deconstructing the regulatory complexity in this model organism.

**Keywords:** architecture, clusters, complex promoters, expression.

**Financial support:** This work was funded by FAPESP (award number 2018/03857-3).