Igor Zwir

University of Granada, Spain.

Mapping sequence to numbers: A quantitative model of promoter binding and gene transcription kinetics under DNA accessibility constraints

A major challenge in biology is to develop quantitative, predictive models of gene regulation that unfold over time in response to environmental changes. Promoters contain transcription factor binding sites differing in their affinity and accessibility, but little is understood about how these variables combine to generate a single fine-tuned, quantitative response. By using the targets of the PhoP DNA binding protein in Salmonella, we were able to quantify the relations between transcription factor input and expression output. We developed a model capable of capturing variable changes in in vitro measurements of kinetic constants of individual binding sites; combining them in promoters with multiple sites; and inferring the corresponding effective affinities that drive disparate in vivo kinetic binding behaviors of PhoP co-regulated promoters. The model faithfully reproduced the observed quantitative changes in terms of binding and transcription onset times and levels (that are not necessarily correlated) of the promoters that occurred upon altering the affinity of the transcription factor for its binding sites by the silencing effect of nucleotide association proteins that impedes transcription factors to access to the DNA. Furthermore, because in vivo binding and in vivo transcription kinetics where independently modeled, we were able to identify key cis-acting elements that are responsible for discrepant behaviors. Because the quantitative measurements are not always available for a given regulator, the quantitative parameters of this model were mapped to sequence motifs representing cis-acting elements arranged in particular promoter architectures, and safely replaced to achieve a stand-alone predictive model. This model, based on sequence analysis, was able to predict gene expression of PhoP regulated genes that were not previously implicated in the model construction, without measuring their biochemical parameters. Finally, the knowhow gained in this study can serve as a base to be applied in other systems and/or species with less detailed genome wide experiments.