Multiscale modeling of gene regulatory networks across evolutionary timescales by integrating "top-down" and "bottom-up" approaches

Pavel Novichkov[1]*, David J Reiss[2]*, Marcin Joachimiak[1], Alexey Kazakov[1], Serdar Turkaslan[2], **Inna Dubchak**[1], **Adam P Arkin**[1], and **Nitin S Baliga**[2]

[1]Lawrence Berkeley National Laboratory, Berkeley, CA; and [2]Institute for Systems Biology, Seattle, WA

http://enigma.lbl.gov/ENIGMA_InformaticsKnowledgebase.html
http://baliga.systemsbiology.net/enigma

Project Goals: Environmental conditions and their associated cellular states dynamically shape regulatory responses. As the environment changes, the solution space for regulation of the same gene (or set of genes) moves between different states depending on the combinatorial interplay between transcription factors, their ligands, and binding sites. Such responses are encoded in gene regulatory networks by specific cis-regulatory signals encoded in the genome. We have developed a novel strategy for reverse engineering microbial gene regulatory networks that times context-specific system-scale regulatory phenomena directly to mechanism at single nucleotide-level resolution.

As a first step toward this goal we have integrated the data-driven and genomic-driven regulon discovery approaches developed within the ENIGMA team. The user can start from an automatic gene regulatory network inference conducted with the ***cMonkey*** integrated biclustering algorithm. ***cMonkey*** generates biclusters (predicting putative regulons with corresponding putative condition-specific cis-regulatory motifs) which can then be directly input to ***RegPredict*** for detailed comparative genomics analysis in order to verify evolutionary conservation of transcription factor binding sites (TFBS) and to refine the regulon content. ***RegPredict*** automatically suggests the most appropriate set of closely related genomes required for comparative analysis. Finally, the refined regulon can be submitted to the ***RegPrecise*** database, a community repository of manually curated regulons.

We have used this strategy to build transcriptional regulatory network models for ENIGMA keystone organisms *D. vulgaris* Hildenborough and *M. maripaludis*. The workflow was tested on several regulons (Rex, Fur, FliA, PerR, SahR) from *D. vulgaris* Hildenborough. For example, ***cMonkey*** predicted the Fur regulon and its corresponding TFBS, which was then refined and 19 additional gene members added using the comparative genomics approach of ***RegPredict***. Based on this model, we are also developing a framework for automated network inference plugged into a Network Portal, which includes a suite of analysis and visualization tools.

The data-driven systems approach has proven powerful for automatic reconstruction of mid- to large-size regulons. However, it is more challenging to identify small regulons due to their small size, a property that limits both gene expression data comparisons and sequence motif detection. Our novel method enables us to tackle the small regulon

portion of the regulatory network reconstruction problem. The method uses a new biclustering engine along with heuristics designed from properties of known regulons to identify small regulon patterns in phylogenetic profiles.

This work conducted by ENIGMA was supported by the Office of Science, Office of Biological and Environmental Research, of the U. S. Department of Energy under Contract No. DE-AC02-05CH11231.