

“Genome-Scale reconstruction of the transcriptional and translational machinery in *Escherichia coli*: A knowledge-database and its mathematical formulation”

Ines Thiele¹, Neema Jamshidi¹, Ronan M.T. Fleming², Bernhard Ø. Palsson^{1*}

¹ Systems Biology Research Group, Bioengineering Department, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0412, USA.

² Department of Biochemistry, National University of Ireland, Galway, Ireland

* Email: bpalsson@bioeng.ucsd.edu

There is an exponentially growing literature reporting large “-omics” datasets that provide systems-level measurements for virtually all types of cellular components in a model organism. Concomitantly there is a need to assemble this information in the context of biological knowledge which can be used for quantitative calculations and predictions of cellular interactions. Metabolic reconstructions have demonstrated utility in serving this purpose in the study of metabolism. The underlying principles used to build these metabolic networks can be applied to other cellular functions, such as transcription and translation.

Knowledge-base: We present the first comprehensive, genome-scale assembly of gene-specific synthesis reactions for *E. coli*'s transcriptome and proteome. The reconstruction accounts for the 303 known gene products of the transcriptional and translational machinery, which together enable the synthesis of all *E. coli* genes. Information from more than 500 publications and 3 databases were collected to create this reconstruction, thus being the first complete knowledge-base for this important cellular function. We accurately captured the current status of knowledge in 27 subsystems, including rRNA and tRNA modification mechanisms and iron-sulfur-cluster biogenesis reactions, which have not been compiled elsewhere to date. The reconstruction process also resulted in a list of 34 unknown proteins for which no corresponding gene has been found in *E. coli*'s genome.

Modeling: Moreover, the reconstruction can be easily converted into a mathematical format, similar to metabolic network reconstructions, which allows the characterization of the synthesis machinery using well established mathematical tools. We validated the network properties by comparing calculated ribosome production with observed ribosome numbers and found excellent overlap. Since mRNA molecules are explicitly accounted for in the model, the integration of microarray data could be done on a quantitative level, which is not possible with the current formulation of metabolic networks. These constraints were used to calculate functional protein modules, that is, proteins which act together within the network in a stoichiometrically coupled manner. Interestingly, we found that these functional protein modules go beyond canonically defined subsystems hence providing a predictive tool for proteomics.

We view this reconstruction and model of the transcriptional and translational machinery of *E. coli* as a first step towards a new generation of metabolic networks that will not only account for enzymes within the reactions but also their synthesis reactions. Hence, a quantitative integration system for many of the “-omics” data such as gene expression, ChIP-chip, and proteomic data will ease the analysis of multiple data sets as well as it can be used as a predictive tool in metabolic and protein engineering and to understand principles in adaptive evolution.