Optimal Protein Reclamation with MaxMass

Making Room for Bioproduction via Intelligent Reduction of the Host Proteome

Neeraj Kumar[1], Jeremy Zucker[1], Jeremy Teuton[1], Ernesto Nakayasu[1], Joseph Cottam[2], and Robert Egbert[1]
[1]Biological Science Division, [2]Visual Analytics, Pacific Northwest National Laboratory, Richland WA.

Introduction/Motivation:

Macromolecular crowding and ribosomal capacity are biophysical constraints on the ability of microbial cell factories to produce proteins of interest at sufficient titer, rate and yield to become economically viable [1,2]. To address these issues directly, the Cellular Capacity Optimization project aims to iteratively knock out nonessential regions of the genome that code for highly expressed proteins in order to reclaim that protein capacity for engineered proteins. In order to rationally guide this process, we have developed MaxMass, an algorithm for iterative reduction of the host proteome without replacement. MaxMass automatically calculates which genes are essential for the current environment and iteration, and uses these essential genes as constraints on the size of the knockout region [3], while optimizing for maximum protein mass reclamation.

Methods:
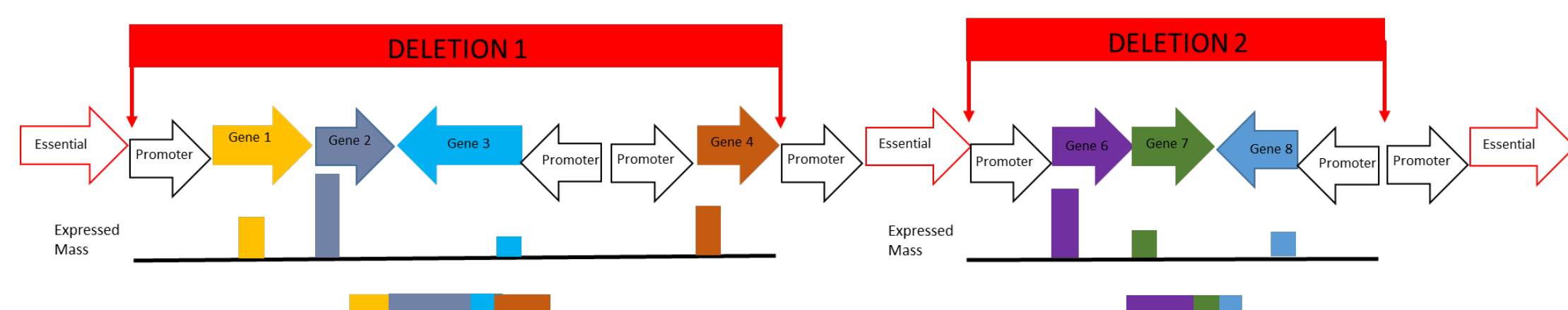
Max Mass Objective

The MaxMass algorithm contains the following objective function:

$$\max \sum_{k \in K} y_k \sum_{i=0}^{k-1} m_i - \sum_{k \in K} x_k \sum_{i=0}^{k-1} m_i$$

where the variables are

$x_k = \begin{cases} 1, & \text{if gene or promoter } k \text{ is the first gene or promoter within the deleted segment} \\ 0, & \text{otherwise} \end{cases}$

and

$y_k = \begin{cases} 1, & \text{if gene or promoter } k \text{ is immediately after the end of the deleted segment} \\ 0, & \text{otherwise} \end{cases}$

where $\sum_{k \in K} x_k = 1$

and $\sum_{k \in K} y_k = 1$

MaxMass constraints

$\sum_{j \in J} S_{i,j} v_j = 0$	$i = 1, \ldots, N$	Steady state conservation of metabolite i requirement
$\sum_{k \in K} y_k = 1$		Ensure only one gene knockout per iteration
$\sum_{j=1}^{k} x_j - \sum_{j=1}^{k} y_j = z_k$	$k = 1, \ldots, K$	Ensures all genes between start and end area also knocked out
$v_{biomass} \geq f \cdot v_{biomass, max}$		Requires that the KO must be capable of growing at the specified fraction of theoretical growth rate
$z_g = \prod_{p \in P} z_p$	$p = 1, \ldots, P$	Ensures that a gene is not functional if all of its promoters are deleted

Gene Protein Reaction constraints

Single gene catalyzes reaction: $k \rightarrow v_j$

$(1-z_k) \cdot LB \leq v_j \leq (1-z_k) \cdot UB$

Enzyme dimer complex catalyzes reaction: $(k_1 \text{ AND } k_2) \rightarrow v_j$

$(1-z_{k_1}) \cdot LB \leq v_j \leq (1-z_{k_1}) \cdot UB$
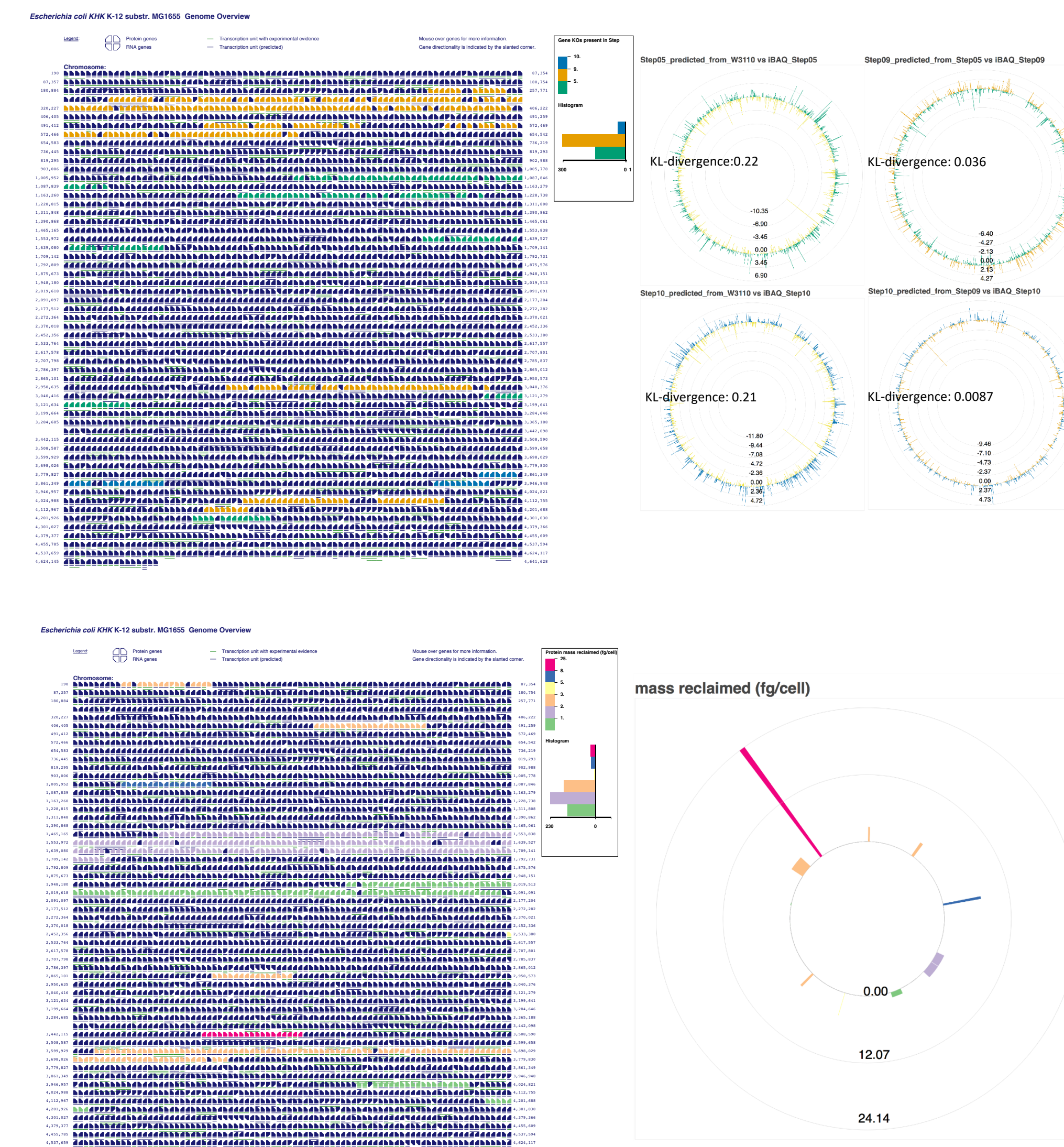$(1-z_{k_2}) \cdot LB \leq v_j \leq (1-z_{k_2}) \cdot UB$

Two Isozymes catalyze reaction: $(k_1 \text{ OR } k_2) \rightarrow v_j$

$(2-z_{k_1}-z_{k_2}) \cdot LB \leq v_j \leq (2-z_{k_1}-z_{k_2}) \cdot UB$
$LB \leq v_j \leq UB$

Complex Gene Protein Reaction: $(k_1 \text{ AND } k_2) \text{ OR } (k_1 \text{ AND } k_3) \rightarrow v_j$

$(2-z_{k_1}-z_{k_2}) \cdot LB \leq v_j \leq (2-z_{k_1}-z_{k_2}) \cdot UB$
$(1-z_{k_2}) \cdot LB \leq v_j \leq (1-z_{k_2}) \cdot UB$

Results:

mass reclaimed (fg/cell)

Protein expression of previous strains tends to overestimate subsequent KO protein expression

To evaluate the ability of the algorithm to predict protein reclamation, we used proteomics data obtained from the KHK collection of knockouts [4]. As a baseline first step, we wished to assess how well the protein distribution of the previous step was able to predict the protein distribution of subsequent steps by assuming that the only expression change between steps were the genes that were knocked out.

MaxMass predicts 27% protein reclamation with 10 knockouts

Applied MaxMass using the protein expression from wild-type E. coli strain W3110 under minimal media [5] in order to predict which gene knockouts would result in the largest amount of protein reclamation. The results of MaxMass are displayed in the figures above. We predict that 27% of the total protein mass (48 fg protein/cell out of 180 fg total protein/cell) can be reclaimed after 10 knockouts.

Discussion/Future Work:

- Proteome prediction based on previous measurements overestimated the actual proteome measurement, but immediately prior steps were better estimators of proteome than wild-type. Interestingly, wild-type protein expression is a better estimator of Step10 (10% net error) than Step05 (20% net error).
- Knocking out the most highly expressed proteins can result in new highly expressed proteins.
- To predict which proteins must be overexpressed in the knockout, future work will focus on extending the metabolic model with RNA and Protein expression mechanisms, known as ME-models. We will apply MaxMass to other organisms such as *Pseudomonas flourescens, Shewanella oneidensis* MR-1 and *E. coli* Nissle.

References

1. Rivas, Germán et al. (2016) Macromolecular Crowding *In Vitro, In Vivo*, and In Between. *Trends in Biochemical Sciences* , Volume 41 , Issue 11 , 970 – 981
2. Jewett, M. C., Miller, M. L., Chen, Y., & Swartz, J. R. (2009). Continued protein synthesis at low [ATP] and [GTP] enables cell adaptation during energy limitation. *Journal of Bacteriology, 191*(3), 1083–91. https://doi.org/10.1128/JB.00852-08
3. Wang, L., & Maranas, C. (2017). MinGenome: An in silico top-down approach for the synthesis of minimized genomes. *ACS Synthetic Biology*, acssynbio.7b00296. https://doi.org/10.1021/acssynbio.7b00296
4. KHK Collection https://shigen.nig.ac.jp/ecoli/strain/
5. Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., … Heinemann, M. (2016). The quantitative and condition-dependent Escherichia coli proteome. *Nature Biotechnology, 34*(1), 104–110. https://doi.org/10.1038/nbt.3418

ABOUT Pacific Northwest National Laboratory

The Pacific Northwest National Laboratory, located in southeastern Washington State, is a U.S. Department of Energy Office of Science laboratory that solves complex problems in energy, national security, and the environment, and advances scientific frontiers in the chemical, biological, materials, environmental, and computational sciences. The Laboratory has been managed by Ohio-based Battelle since 1965.

Neeraj Kumar
Pacific Northwest National Laboratory
Richland, WA 99352
(509) 372-6902
Neeraj.kumar@pnnl.gov

U.S. DEPARTMENT OF ENERGY