Budgeting for cellular expression: how to calculate protein expression costs?

Maud HOFMANN[1], Rafael MONTENEGRO[1], Olivier BORKOWSKI[1], Manish KUSHWAHA[1]

1 Institut Micalis, INRAe, Domaine de Vilvert, 78352 Jouy-en-Josas, France

Abstract

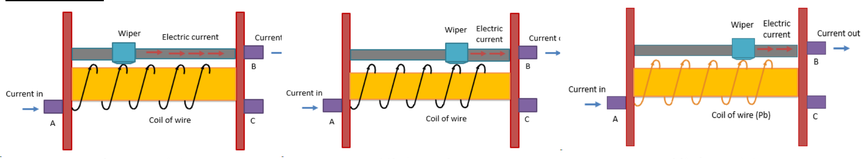
With the rapid advancement of synthetic biology, engineering robust genetic circuits has become increasingly challenging. As synthetic genetic constructs grow in size and complexity, they impose a greater burden on host cells, risking construct loss due to negative selection. This burden arises from a competition for gene expression resources [1, 2, 3], particularly in transcription and translation [4, 5].

To address this problem, we have established a system which allows us to precisely quantify the expression cost of any arbitrary coding sequence in a direct and sequence-specific manner. The coding sequence will be cloned in between two sites encoding inteins (protein-level ligases) [6], which are located inside a GFP sequence. After translation, the inteins are spliced out, together with the inserted peptide. Consequently, the two parts of the GFP protein are reassembled and can fluoresce. By comparing the resulting fluorescence of this construct containing an insert sequence to a « no insert » control, we can quantify the insert's expression cost. Using this system, we will test inserts with different sequence-specific properties, and also vary gene expression parameters, such as transcription initiation rate and translation initiation rate.

In doing so, we will collect a wide dataset which, once incorporated into a mathematical model, will allow us to reliably predict the gene expression cost of any arbitrary gene. This quantification will help us address the issue of a cellular burden caused by a heterologous construct and establish the rules necessary to design an efficient genetic circuit with a low burden.

How to measure gene expression cost?

Rheostat



- Low resistance
- High current

- High resistance
- Low current

- High resistance
- Low current

Heterologous Gene Expression

- Short sequence
- Low burden

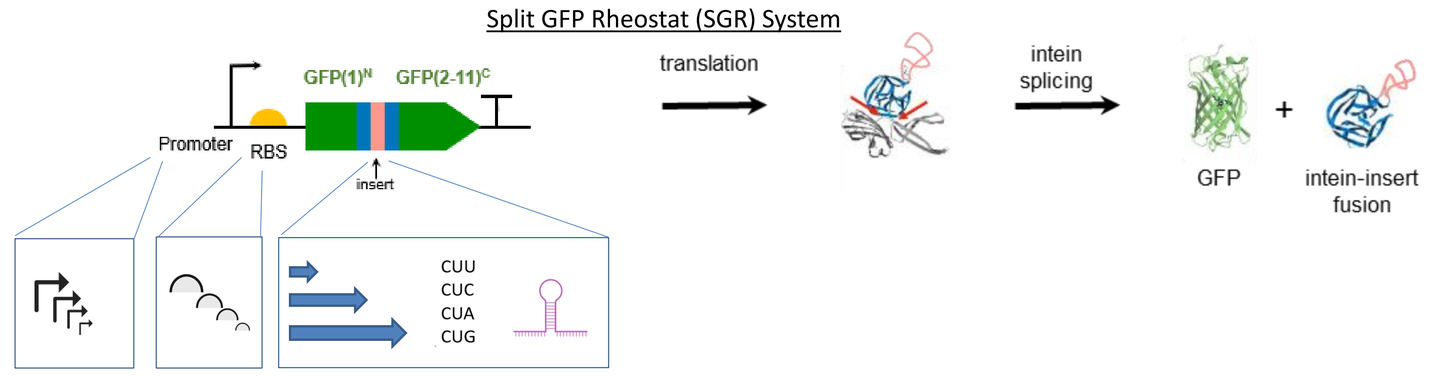
- Long sequence
- High burden

CDS length

CUU
CUC
CUA
CUG

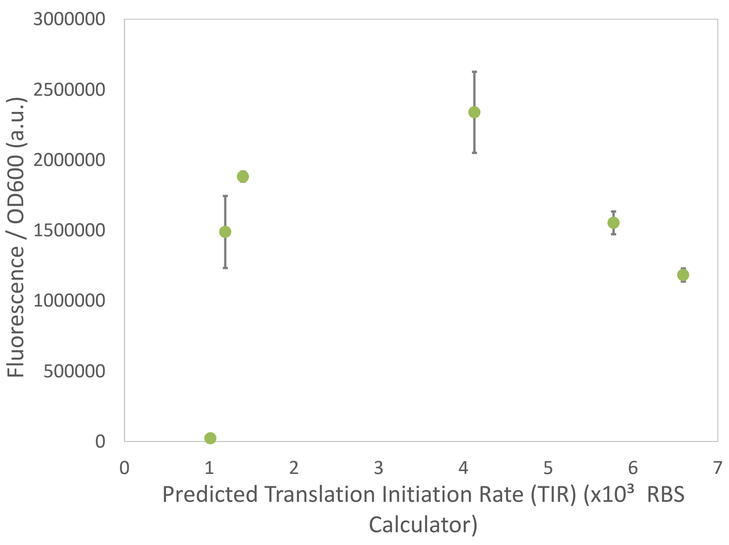
CDS composition

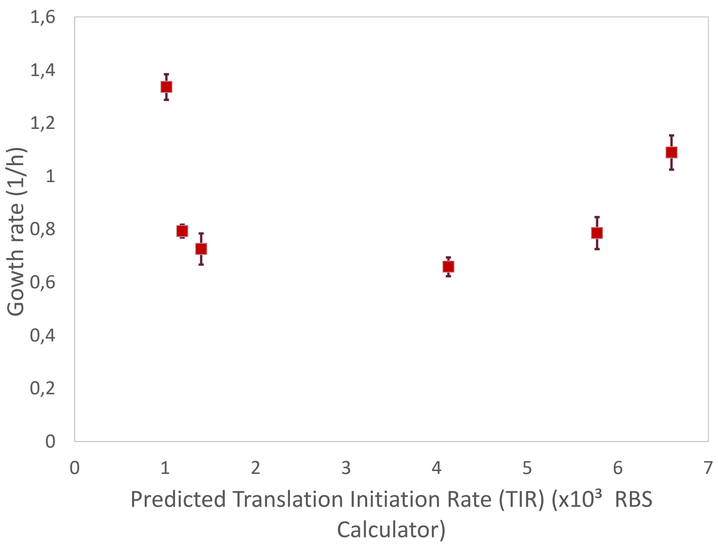
Split GFP Rheostat (SGR) System

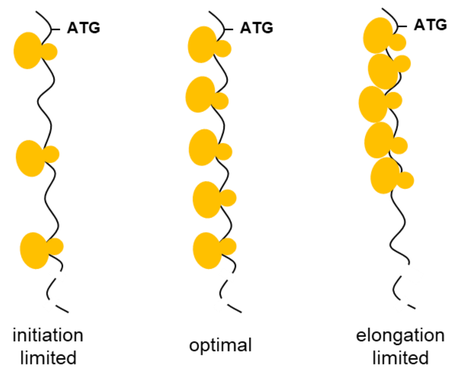


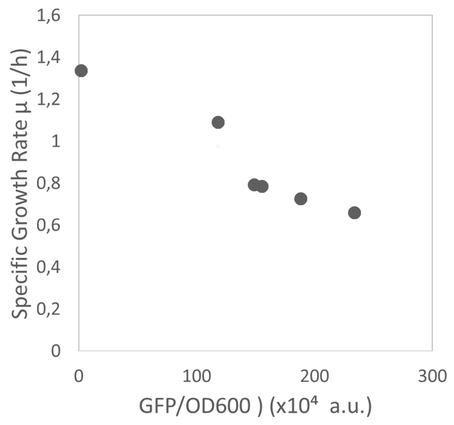
- Quantification of gene expression cost through intein splicing
- Variation of gene expression parameters, as well as CDS length and composition

Validation of the SGR System



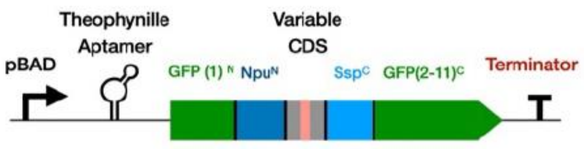


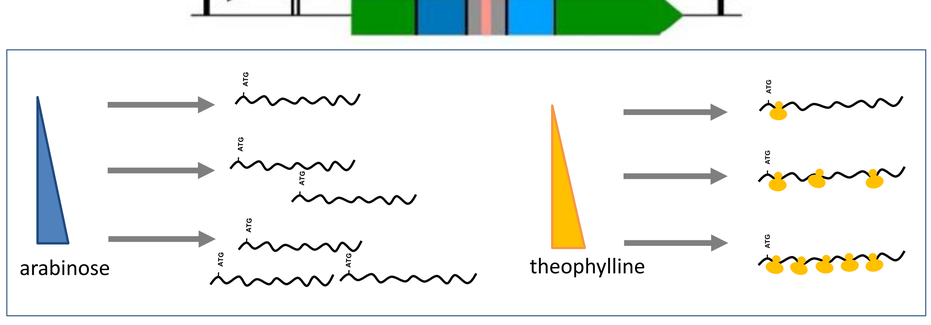




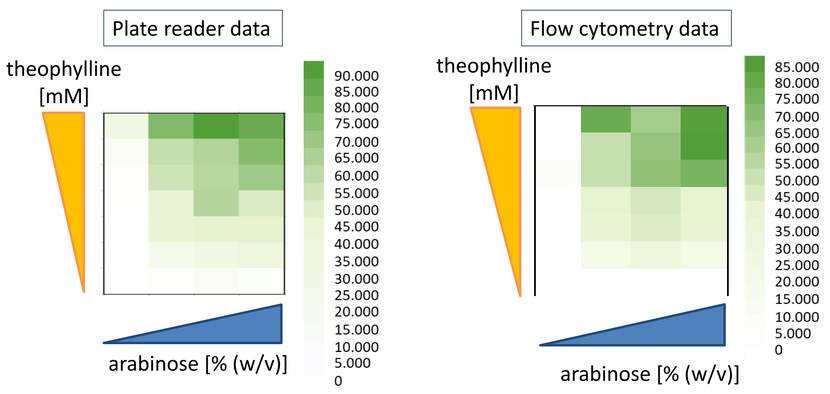
- There is an ideal RBS, for which GFP expression reaches a maximum.
- With further increase of the TIR, expression reaches a plateau.
- There is an inverse relationship between GFP production and growth rate.

Inducible SGR System

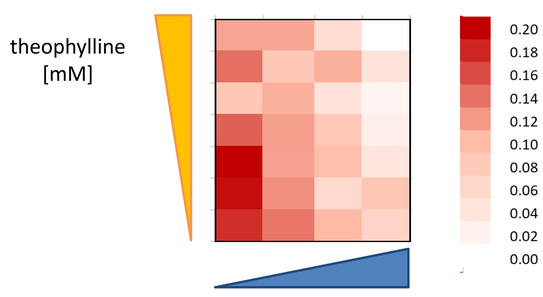




Average GFP/OD600 (a.u.)



Specific Growth Rate µ [1/h]



Conclusions

- GFP as a reporter through splicing allows "direct" measurement of insert cost.
- independent control of translation initiation and elongation rate
- inducible control of transcription and translation initiation rates and validation of dose-response

Next steps

- Characterization of insert sequences ("best" vs. "worst" codons or increasing vs. decreasing codon optimization)
- Data incorporation into a mathematical model, to enable reliable prediction of the gene expression cost of any arbitrary gene.

Centre
Île-de-France - Versailles-Grignon

Contact:
maud.hofmann@inrae.fr
https://www.linkedin.com/in/maud-hofmann

References

[1] Tuller, T., Waldman, Y.Y., Kupiec, M., and Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. Proc Natl Acad Sci USA 107, 3645–3650.
[2] Borkowski, O., Bricio, C., Murgiano, M., Rothschild-Mancinelli, B., Stan, G.-B., and Ellis, T. (2018). Cell-free prediction of protein expression costs for growing cells. Nat Commun 9, 1457.
[3] Ceroni, F., Algar, R., Stan, G.-B., and Ellis, T. (2015). Quantifying cellular capacity identifies gene expression designs with reduced burden. Nat Methods 12, 415–418.
[4] Li, G.-W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. Cell 157, 624–635.
[5] Marquez-Zavala, E., Utrilla, J. (2023). Engineering resource allocation in artificially minimized cells: Is genome reduction the best strategy? Microb. Biotech. 16, 990–999.
[6] Gogarten, J. P., Senejani, A. G., Zhaxybayeva, O., Olendzenski, L., Hilario, E. (2002). INTEINS: Structure, Function, and Evolution. Annu. Rev. Microbiol. 56, 263–87.