gcFront: a tool for determining a Pareto front of growth-coupled cell factory designs

Laurence Legon[1,2], Christophe Corre[2], Declan G. Bates[1,*] and Ahmad A. Mannan[1,*]

[1]*Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry, UK*

[2]*Warwick Integrative Synthetic Biology Centre, School of Life Sciences, University of Warwick, Coventry, UK*

*** Corresponding authors**: ahmad.mannan@warwick.ac.uk, d.bates@warwick.ac.uk

ABSTRACT

Motivation: A widely applicable strategy for developing evolutionarily robust cell factories is to knock out (KO) genes or reactions to couple chemical synthesis with cell growth. Genome-scale metabolic models enable their rational design, but KOs that provide growth-coupling (gc) are rare in the immense design space, making searching difficult and slow, and though several measures determine the utility of those strains, few drive the search.

Results: To address these issues we developed a software tool named gcFront - using a genetic algorithm it explores KOs that maximise key performance objectives: cell growth, product synthesis, and coupling strength. Our measure of coupling strength facilitates the search, so gcFront not only finds a gc-design in minutes but also outputs many alternative Pareto optimal gc-designs from a single run - granting users freedom to select designs to take to the lab.

Availability and Implementation: gcFront, with documentation and a workable tutorial, is freely available at GitHub: https://github.com/lLegon/gcFront, the repository of which is archived at Zenodo, DOI: 10.5281/zenodo.5557756 (Legon *et al.*, 2021).

Supplementary Information: Supplementary notes and data files are available at Bioinformatics online.

INTRODUCTION

Genome-scale constraint-based models (GSMs) are used to explore gene or reaction knockouts (KOs) that rewire cell metabolism for chemical overproduction (Maia *et al.*, 2016). A promising strategy for enabling robust chemical synthesis seeks KO combinations that couple production with cell growth so product synthesis becomes obligatory at maximum cell growth rate (Ohno *et al.*, 2014; Feist *et al.*, 2010). Though KOs can disrupt metabolism, resulting in poorer performance than predicted, growth-coupling allows the selection of higher producing phenotypes through adaptive laboratory evolution, which can even find strains with synthesis and growth rates near the predicted optimal (Tokuyama *et al.*, 2018). Moreover, designs with stronger coupling, where smaller drops in synthesis strongly impair growth, could make synthesis more robust, both with respect to evolution by creating stronger selection pressure to conserve optimal synthesis; and to a heterogeneous growth environment by conserving high synthesis in suboptimally-growing cells. Growth-coupling is a widely applicable strategy (von Kamp and Klamt, 2017), but KOs enabling this are rare, making the search for them difficult and slow (Ohno *et al.*, 2014). Also, fast growth, high synthesis, and strong coupling are desirable characteristics of a strain design, the search for which we address as a multiobjective optimization problem. To solve it, we developed a software tool named gcFront that determines KO sets that predict phenotypes maximal in these three performance objectives, for any chemical and host of interest. Moreover, our proposed measure of coupling strength facilitates the search through the design space, so that gcFront outputs many Pareto optimal designs from a single run, in reasonable time.

THE gcFRONT WORKFLOW

gcFront works in MATLAB, with dependencies on the COBRA toolbox (Heirendt *et al.*, 2019) for analysis of a compatible GSM; and the MATLAB Global Optimization toolbox for solving the multiobjective optimization problem (Supplementary Note 1A). The workflow, detailed in Supplementary Note 1B and Supplementary Fig. 1, entails four fundamental steps.

Inputs. Two interactive windows allow the user to define the GSM and target metabolite product or its exchange reaction, and optional inputs (Supplementary Table 1), like maximum number of KOs and search time.

Preprocessing. To reduce the sets of KOs and search space, dead reactions are removed, unbranched pathways are lumped into composite reactions, and single KOs essential *in silico* for growth or synthesis are excluded.

Solving the optimization problem. To determine growth-coupled (gc-)designs, gcFront solves the multiobjective optimisation problem defined in Supplementary Note 2A. Our measure of coupling strength shapes the search landscape; it defines weak and strong coupling but also distinguishes between uncoupled designs (Supplementary Note 2B, Supplementary Fig. 2a). It assigns higher values to KOs that reduce the cost to growth for increases in the maximum allowable synthesis, thus driving bias to gc-designs (Supplementary Fig. 2b) to ease the search.

Post-processing and output. On termination (conditions in Supplementary Table 1), many Pareto optimal KO sets are found from a single run. Genes/reactions that do not affect performance if ignored are removed from each set, and the Pareto front of designs and their performance is output to an interactive plot, a table in the command window, and a .csv file. Users can select designs they deem suitable for their chemical and host of interest, based on bespoke combinations of the performance metrics. A tutorial is given in Supplementary Note 3.

COMPARATIVE PERFORMANCE ASSESSMENT

To test gcFront's performance, we compared it to other MATLAB-based procedures that identify gc-designs, including RobustKnock (Tepper and Shlomi, 2009) as implemented in OptPipe (Hartmann *et al.*, 2017); gcOpt (Alter and Ebert, 2019); FastPros (Ohno *et al.*, 2014); and OptGene (Patil *et al.*, 2005) as implemented in COBRA (Heirendt *et al.*, 2019). We ran each for 6 hours, saved the gc-designs found and the time they needed to find their first gc-design, repeating this three times for gcFront and OptGene because of the stochastic nature of searching with a genetic algorithm. For a fair comparison, we ran each algorithm using the *E. coli* GSM model iML1515 (Monk *et al.*, 2017), for non-essential reaction KOs (*in silico* and based on (Goodall *et al.*, 2018)), for synthesis of succinate, tyrosine, and pyruvate, as example products (detailed in Supplementary Notes 4 and Fig. 1). gcFront found the first gc-design in 38% less time than gcOpt for succinate synthesis, 98% less time than RobustKnock for tyrosine synthesis, and orders of magnitude less time than the other methods and products (Fig. 1a, Supplementary Data). Its power was especially apparent when searching for designs of tyrosine and pyruvate synthesis – still finding designs in minutes despite these designs, of at least six KOs, being rarer vs three KOs found for succinate (Supplementary Fig. 3). Furthermore, though the single gc-design found with other methods lay near the Pareto front of gc-designs from gcFront, gcFront offered many designs that achieved at least higher coupling strength (Fig. 1b, Supplementary Data).

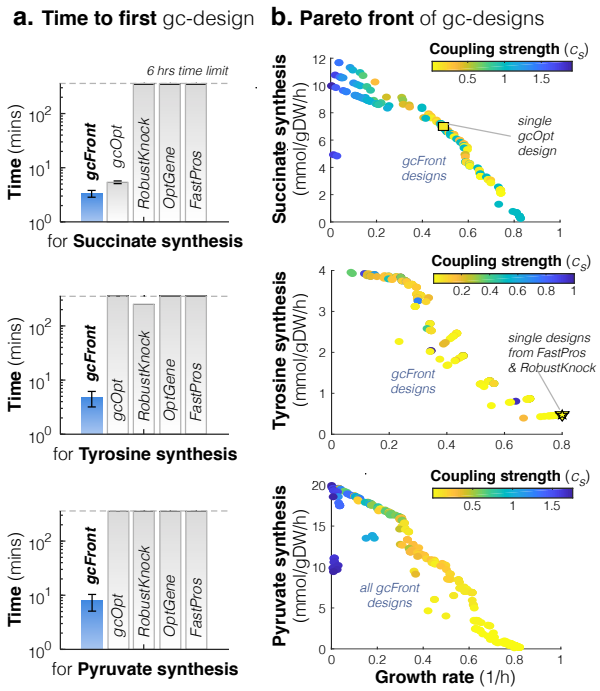

Fig. 1. gcFront finds many Pareto optimal growth-coupled designs, faster, and with superior performance vs other algorithms. The speed and designs found from 6-hour runs of gcFront were compared to those of RobustKnock, gcOpt, FastPros, and OptGene (see Supplementary Note 4) on a MacBook Pro (2.3 GHz Quad-Core Intel core i5 processor, 8GB 2133 MHz LPDDR3 RAM). Designs were based on KOs of only non-essential, gene-associated reactions, for the synthesis of three example products: succinate, tyrosine, and pyruvate from the *E. coli* iML1515 GSM model, in aerobic, minimal media with glucose. (**a**). Time to identify the first gc-design from each procedure. Due to the stochastic nature of searching using the genetic algorithm in OptGene and gcFront, the average (bars) and standard deviation (error bars) of times are reported from three runs (N=3, ±SD). (**b**). Pareto fronts of all gc-designs found from three 6-hour runs.

DISCUSSION

gcFront can find a multitude of Pareto optimal growth-coupled designs for evolutionarily robust cell factories, from a single run, in a computationally efficient manner. Since each gc-design provides a different balance between the maximised objectives, the user is free to select designs with the balance they deem most suitable to the cell host and chemical product of interest, e.g. sacrifice growth for stronger coupling and synthesis, for instance for more robust pyruvate synthesis (Fig. 1b); or sacrifice synthesis for higher growth and stronger coupling, for instance for higher volumetric productivity with robust synthesis of succinate (Fig. 1b) – making it widely applicable to different contexts. It is also user friendly, but versatile – the interactive user interface means no coding is required, making gcFront easy to use out-of-the-box, yet because it is a function in the MATLAB environment it can be easily

integrated within a wider pipeline. For instance, it could complement pipelines automating the construction of synthetic metabolic pathways (Carbonell *et al.*, 2018) - serving as an input to envisioned robotic labs implementing KOs (Sinha *et al.*, 2018) and adaptive laboratory evolution (Radek *et al.*, 2017), for evolving optimally performing phenotypes.

FUNDING

This work was supported by EPSRC (EP/L016494/1) & BBSRC (BB/M017982/1).

AUTHOR CONTRIBUTIONS

LL, AAM: developed theory; LL: developed code, ran procedures, plotted results; AAM, DGB: designed and supervised the research; all authors: wrote the paper.

CONFLICT OF INTEREST

None declared.

REFERENCES

Alter,T.B. and Ebert,B.E. (2019) Determination of growth-coupling strategies and their underlying principles. *BMC Bioinformatics*, **20**, 1–17.

Carbonell,P. *et al.* (2018) An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Commun. Biol.*, **1**, 1–10.

Feist,A.M. *et al.* (2010) Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. *Metab. Eng.*, **12**, 173–186.

Goodall,E.C.A. *et al.* (2018) The essential genome of Escherichia coli K-12. *MBio*, **9**, e02096-17.

Hartmann,A. *et al.* (2017) OptPipe - a pipeline for optimizing metabolic engineering targets. *BMC Syst. Biol.*, **11**, 1–9.

Heirendt,L. *et al.* (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.*, **14**, 639–702.

121 von Kamp,A. and Klamt,S. (2017) Growth-coupled overproduction is feasible for almost all metabolites in five
122 major production organisms. *Nat. Commun.*, **8**, 15956.
123 Legon,L. *et al.* (2021) gcFront - a tool for determining a Pareto Front of growth-coupled cell factory designs.
124 *Zenodo*, doi:10.5281/zenodo.5557756
125 Maia,P. *et al.* (2016) In Silico Constraint-Based Strain Optimization Methods: the Quest for Optimal Cell Factories.
126 *Microbiol. Mol. Biol. Rev.*, **80**, 45–67.
127 Monk,J.M. *et al.* (2017) iML1515, a knowledgebase that computes Escherichia coli traits. *Nat. Biotechnol.*
128 Ohno,S. *et al.* (2014) FastPros: Screening of reaction knockout strategies for metabolic engineering. *Bioinformatics*,
129 **30**, 981–987.
130 Patil,K.R. *et al.* (2005) Evolutionary programming as a platform for in silico metabolic engineering. *BMC*
131 *Bioinformatics*, **6**, 1–12.
132 Radek,A. *et al.* (2017) Miniaturized and automated adaptive laboratory evolution: Evolving Corynebacterium
133 glutamicum towards an improved D-xylose utilization. *Bioresour. Technol.*, **245**, 1377–1385.
134 Sinha,H. *et al.* (2018) An automated microfluidic gene-editing platform for deciphering cancer genes. *Lab Chip*, **18**,
135 2300–2312.
136 Tepper,N. and Shlomi,T. (2009) Predicting metabolic engineering knockout strategies for chemical production:
137 Accounting for competing pathways. *Bioinformatics*, **26**, 536–543.
138 Tokuyama,K. *et al.* (2018) Application of adaptive laboratory evolution to overcome a flux limitation in an
139 Escherichia coli production strain. *Biotechnol. Bioeng.*, **115**, 1542–1551.
140