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]

Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry, UK | [1] School of Engineering; [2] School of Life Sciences

bioRxiv — THE PREPRINT SERVER FOR BIOLOGY

Preprint *(click here to access)*

Legon, L., Corre, C., Bates, D. G, Mannan, A. A (2021) bioRxiv, doi:10.1101/2021.10.12.464108

WARWICK — THE UNIVERSITY OF WARWICK — WISB — WARWICK INTEGRATIVE SYNTHETIC BIOLOGY CENTRE

UKRI Biotechnology and Biological Sciences Research Council — *grant BB/M017982/1*

UKRI Engineering and Physical Sciences Research Council — *grant EP/L016494/1*

Motivation

— **Genome-scale models** are used to **predict KOs** that will reroute cell metabolism for **chemical overproduction**.

— A promising strategy to **create cell factories with robust chemical synthesis** is do **KOs** that make **synthesis obligatory at high growth** — growth-coupling.

— This enable us to **evolve and select** KO mutants on growth, to **attain evolutionarily robust, high synthesis strains**.

Problem

— But **designs are rare** in the immense search space — making it **difficult and slow to find**.

— We developed gcFront - **a user-friendly tool** that **efficiently determines many KO sets** for growth-coupled synthesis.

Three key performance objectives of cell factories

[1] *Strong growth-coupled synthesis*

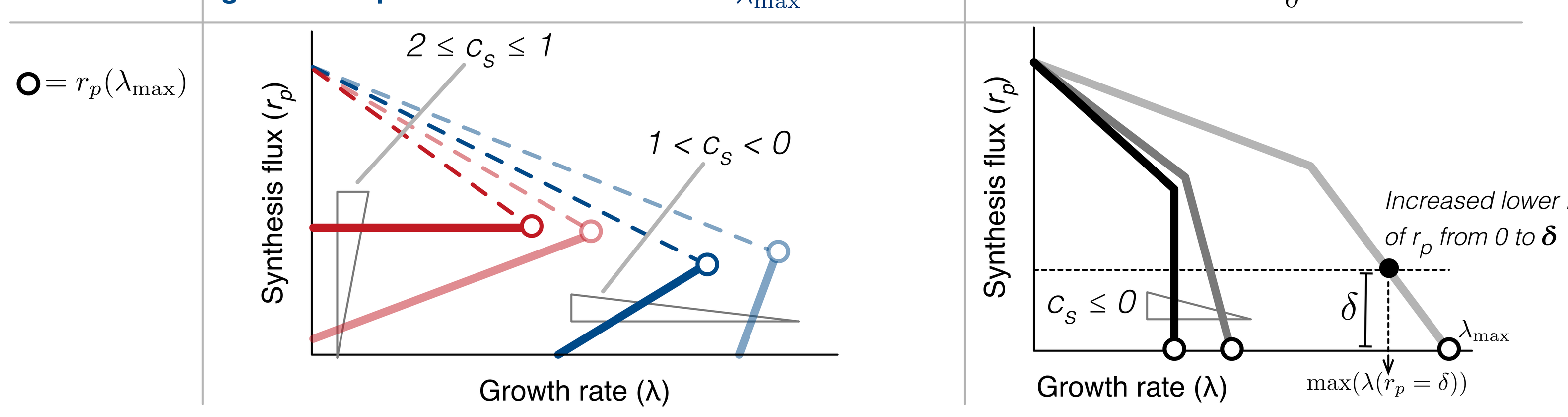
Coupling strength (c_s)

Strongly growth-coupled + **Weakly growth-coupled**:

$$c_s^{gc} = \min\left(1, \frac{\min(r_p(\lambda=0))}{\min(r_p(\lambda_{max}))}\right) + \frac{\lambda_{max} - \max(\lambda(r_p=0))}{\lambda_{max}}$$

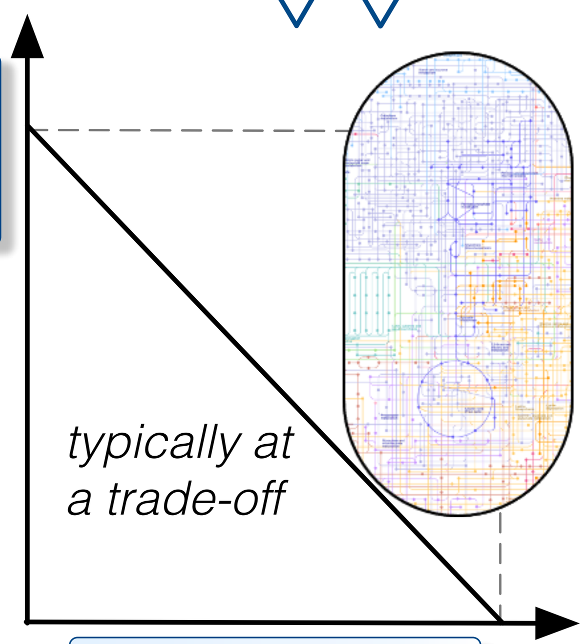
Non growth-coupled:

$$c_s^{nc} = \frac{\lambda_{max} - \max(\lambda(r_p=\delta))}{\delta}$$



[2] *High synthesis flux*

[3] *Fast growth*



Determining cell factory designs — A multiobjective optimization problem

$$\max_{\underline{r}}(J_1, J_2, J_3),$$
J_1 = growth rate, λ; J_2 = synthesis flux, r_p; J_3 = coupling strength, c_s

subject to

$S \cdot \underline{r} = \underline{0}$, and $\underline{k} \circ \underline{\text{lb}} \le \underline{r} \le \underline{k} \circ \underline{\text{ub}}$,

where $\underline{k}$ is vector defining rxn KOs, $k_i \in \{0,1\}$

gcFront - the workflow

Required prerequisites

MATLAB, Global Optimization Toolbox

COBRA Toolbox, LP solver: **glpk** or **Gurobi**

Human-required input

Compulsory

- Name curated and constrained COBRA-compatible GSM (i.e. GSMs from BiGG database) of host cell with product synthesis.
- Name target metabolite

***Optional** for GSM model*

- LP solver (e.g. glpk or Gurobi)
- KO genes or reactions (rxns)
- List of rxns or genes to exclude from KOs
- Max number of KOs
- Minimum growth rate

***Optional** for genetic algorithm*

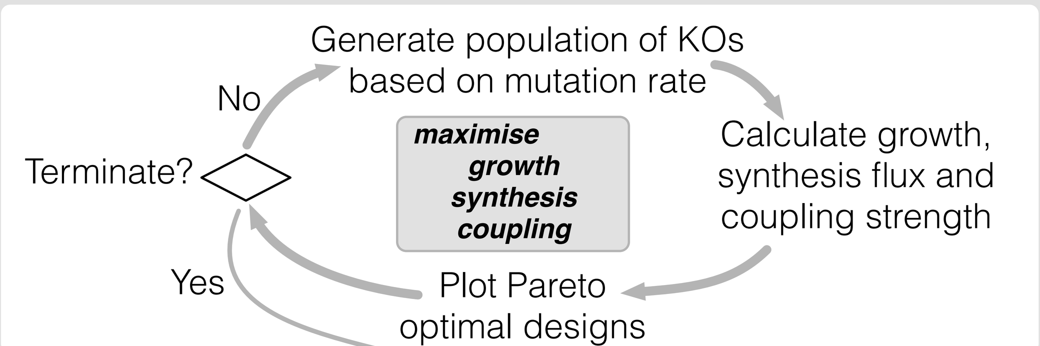
- Population size
- Mutation rate
- Define termination condition: max number of generations or time limit

gcFront toolbox & computational processing

Pre-processing

- Model reduction - delete dead rxns, lump unbranched p/w
- Generate list of candidate KO - excluding those defined in options & where single KOs do not allow growth or synthesis
- Tilt objective vector (to ensure minimum synthesis flux is found)

*Solving **multiobjective optimisation** problem*

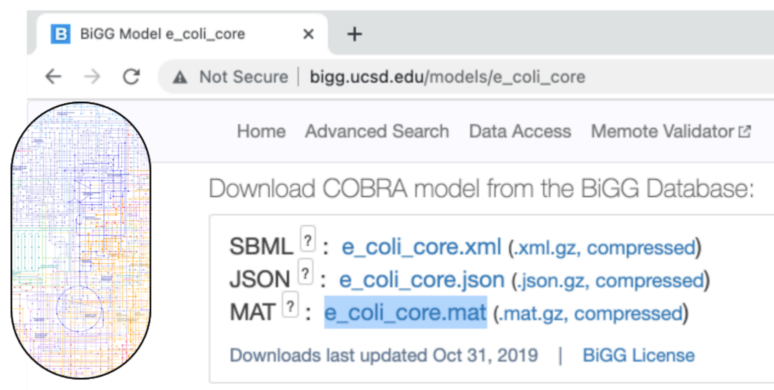


Post-processing

- Remove redundant KOs from Pareto optimal designs
- Calculate Euclidean distance of each design to ideal point
- Save designs and metrics

How to run gcFront

1— Download genome-scale model (e.g. from BiGG database)



2— Import and **modify model** using COBRA toolbox **and …**

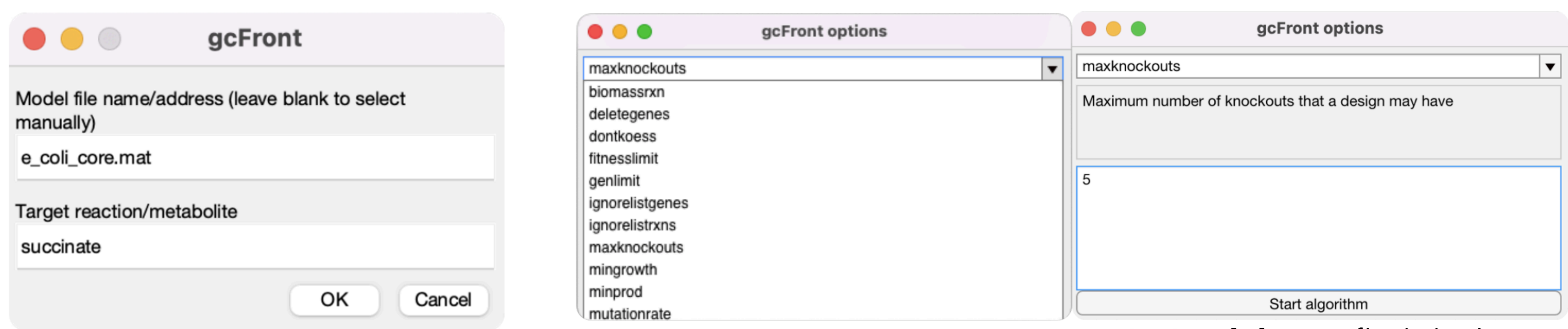
Command Window

```
% setting glucose supply to 10 mmol/gDW/h in E. coli core
% by changing lower bound of exchange reaction
model=readCbModel('e_coli_core.mat');
model=changeRxnBounds(model, "EX_glc__D_e", -10, 'l');
save('e_coli_core.mat','model');

fx >> gcFront
```

… run "gcFront" in MATLAB command window

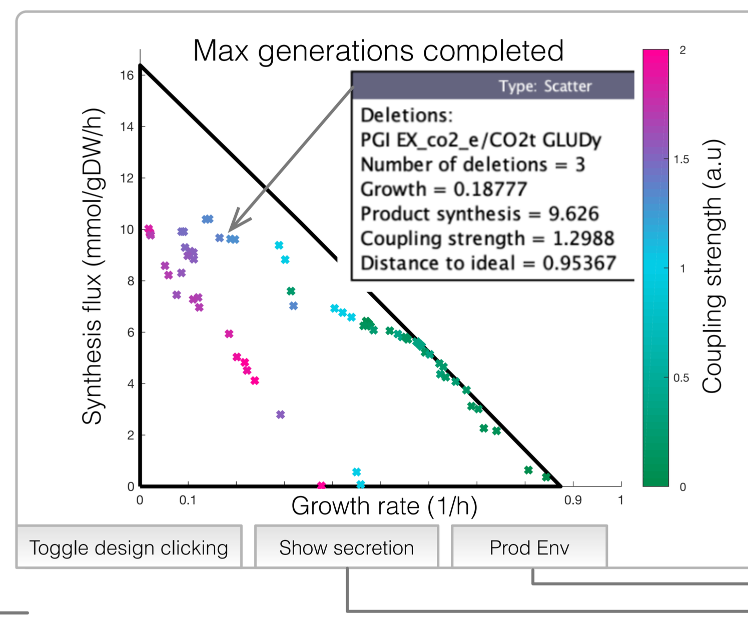
3— Define model, **target product**, and **options**, and …



… start solving to find designs

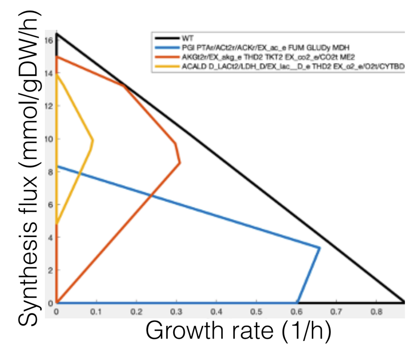
4— Design output and **exploration**

*Interactive **Pareto front** of optimal designs*



***Output** designs and performance measures (optimized objectives) to **Excel** table*

ReactionDeletions	NoOfDels	GrowthRate	ProductFlux	CouplingStrength	DistFromIdeal
"PFL PGI EX_co2_e/CO2t"	3	0.14332	10.406	1.3336	0.9711
"PFL PGI EX_co2_e/CO2t GLUDy"	4	0.13686	10.388	1.3342	0.97779
"PFL D_LACt2/LDH_D/EX_lac__D_e ETOHt2r/ALCD2x/EX_etoh_e THD2 EX_o2_e/O2t/CYTBD"	5	0.017631	10.025	1.8369	1.057
"ACALD D_LACt2/LDH_D/EX_lac__D_e THD2 EX_o2_e/O2t/CYTBD"	4	0.090648	9.9108	1.4837	1.0129
"PFL D_LACt2/LDH_D/EX_lac__D_e ETOHt2r/ALCD2x/EX_etoh_e EX_o2_e/O2t/CYTBD"	4	0.020154	9.8848	1.8488	1.0571



*Output **prodn envelopes***

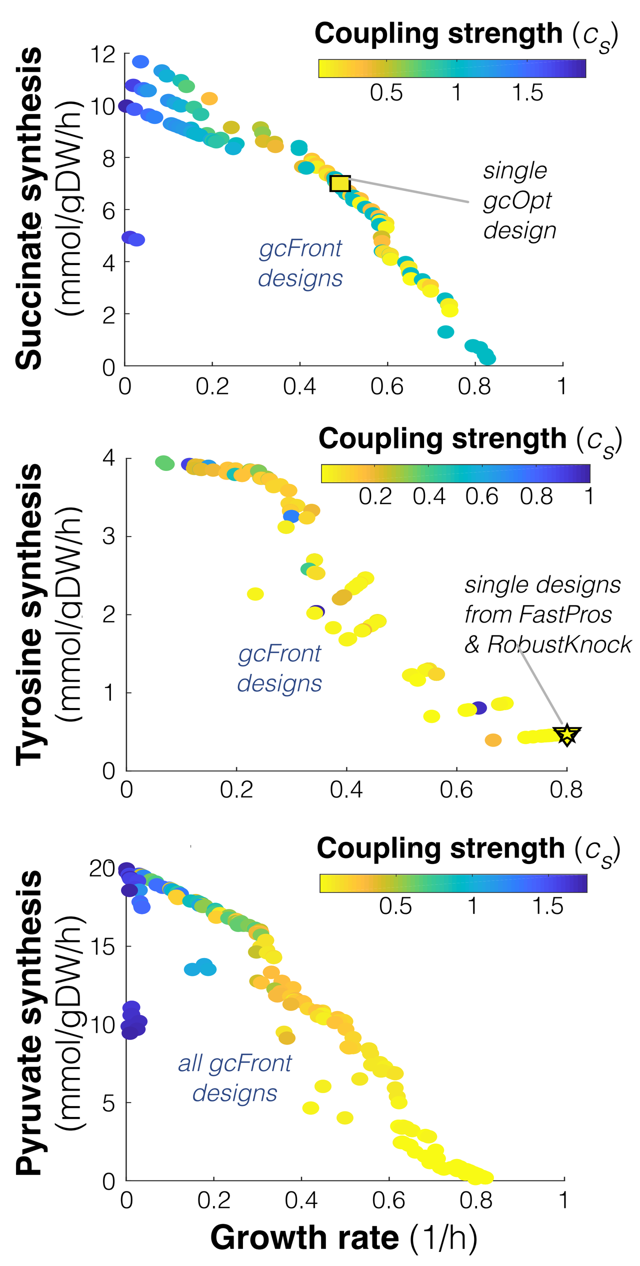
Deletions: PGI EX_co2_e/CO2t GLUDy

Metabolite	MinFlux	MaxFlux	Reaction	ReactionFormula
"H+"	32.324	32.324	'EX_h_e'	'h_e <=> '
"H2O H2O"	10.994	10.994	'EX_h2o_e'	'h2o_e <=> '
"Succinate"	9.626	9.6261	'EX_succ_e'	'succ_e -> '
"Formate"	5.1059	5.106	'EX_for_e'	'for_e -> '
"Acetate"	4.1995	4.1997	'EX_ac_e'	'ac_e -> '
"Pyruvate"	0	0.00010937	'EX_pyr_e'	'pyr_e -> '
"D-Lactate"	0	4.1015e-05	'EX_lac__D_e'	'lac__D_e -> '
"Acetaldehyde"	0	3.6458e-05	'EX_acald_e'	'acald_e -> '
"2-Oxoglutarate"	0	2.9829e-05	'EX_akg_e'	'akg_e -> '
"Ethanol"	0	2.524e-05	'EX_etoh_e'	'etoh_e -> '

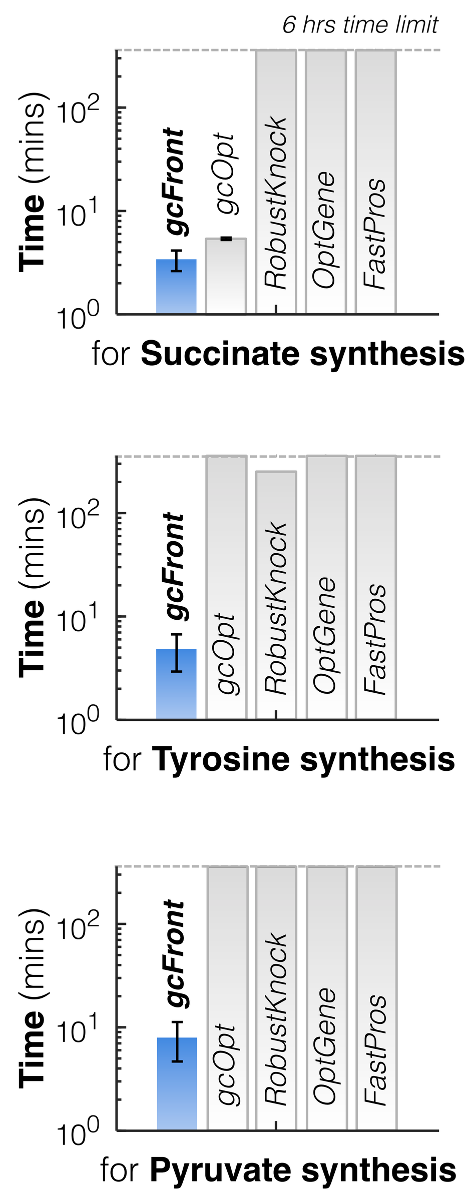
*Also output other **metabolites predicted to be secreted***

E.g. performance — finds many designs & faster

Pareto front of gc-designs



Time to first gc-design



Impact - gcFront in the creation pipeline

Design - **gcFront** suggests KO sets for growth-coupled synthesis

Build and evolve - implement KOs in lab and adopt adaptive lab evolution to converge to predicted optimal perf. (e.g. Tokuyama, et al., 2018)

Test - measure growth and synthesis to assess closeness to optimal predicted.

Take-home messages

— We want to **create cell factories** with **fast growth**, **high synthesis** and **strong coupling**.

— **gcFront** solve this multiobjective optimization problem, **finds many competing designs**, in **reasonable time** - rather than 1 design after hrs.

— **gcFront** is **user friendly**, should be **widely applicable**, and **freely available** at GitHub link:

https://github.com/lLegon/gcFront/tree/V1.0.

Key references

von Kamp, A. and Klamt, S. (2017) Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. Nat. Commun., 8, 15956.

Tokuyama, K. et al. (2018) Application of adaptive laboratory evolution to overcome a flux limitation in an *E. coli* production strain. Biotechnol. Bioeng., 115, 1542–1551.

Heirendt,L. et al. (2019) Creation and analysis of COBRA Toolbox v.3.0. Nat. Protoc., 14, 639–702.

Acknowledgements

We would like to thank UK funding — BBSRC (grant BB/M017982/1) and EPSRC (grant EP/L016494/1).

Author Contact Info
please click icons

ahmad.mannan@warwick.ac.uk

ResearchGate Profile

Google Scholar Profile

LinkedIn Profile

@AhmadMannan17