

gcFRONT TUTORIAL

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

Laurence Legon^{1,2}, Christophe Corre², Declan G. Bates^{1,*} and Ahmad A. Mannan^{1,*}

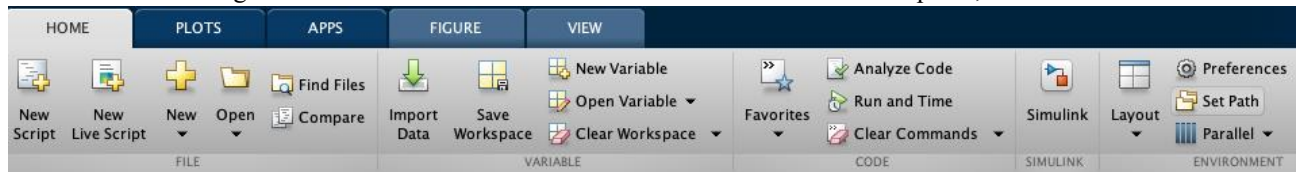
¹ Warwick Integrative Synthetic Biology Centre, School of Engineering, University of Warwick, Coventry, UK

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

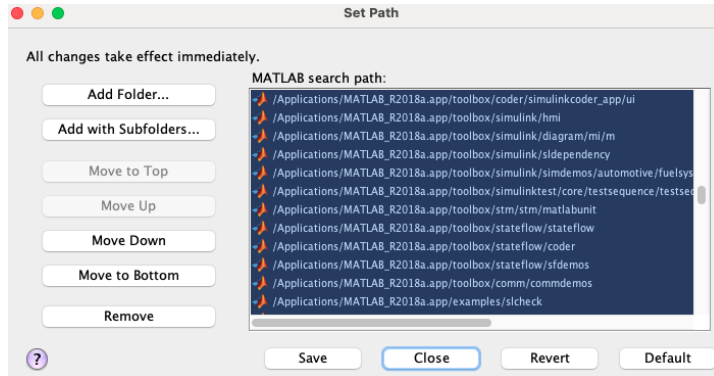
* Corresponding authors: ahmad.mannan@warwick.ac.uk, d.bates@warwick.ac.uk

A. Work-through example - finding gc-design for succinate overproduction in *E. coli* core

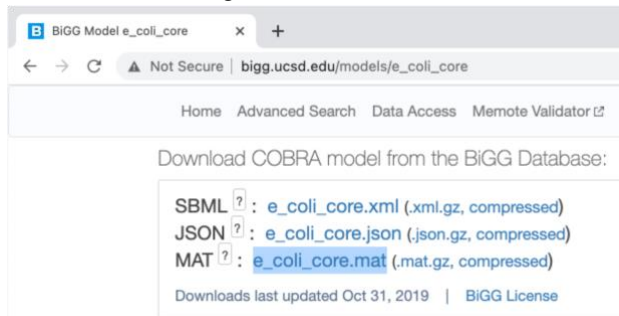
1. First, ensure that all prerequisites (MATLAB Global Optimization toolbox, COBRA toolbox, LP solver) are available or have been downloaded and are working correctly (gcFront should warn you if they are not).
2. Also, ensure that you have added the gcFront_code folder to your MATLAB path. This only needs to be done the first time. To do this go to the home tab and select “Set Path” under the environment panel, like so:



Then click “Add Folder...”, navigate to the gcFront folder, click “Open”, and finally click “Save”:



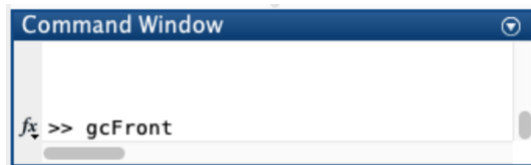
3. Then, download the model of interest. In this example, we will use the *E. coli* core GSM model from the BiGG model database (King *et al.*, 2016):



- Before running gcFront, we should ensure all model reaction constraints are set for the defined media condition, genotype, etc, and save the changes. For this *E. coli* core model, we use the COBRA toolbox to set the maximum glucose uptake rate to 10 mmol/gDW/h by adjusting the lower bound of the glucose exchange reaction to -10 mmol/gDW/h:

```
% setting glucose supply to 10 mmol/gDW/h in E. coli core model
% 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');
```

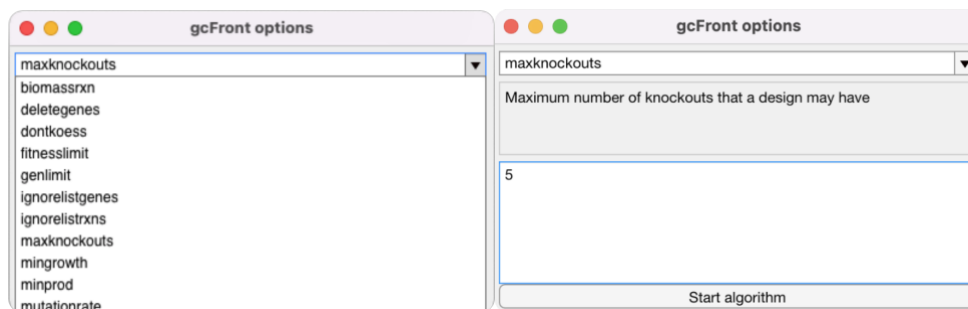
- Type “gcFront” into the command window (if desired, algorithm inputs and parameters can be supplied as arguments to the function as well - see the SUCCINATE_TUTORIAL.m file in the code downloadable from GitHub for an example of how to do this).



- Enter the file path and name of the GSM that needs to be imported (here it is “e_coli_core.mat”), and the reaction or metabolite that you want to couple (here it appears as “succinate” in the GSM):

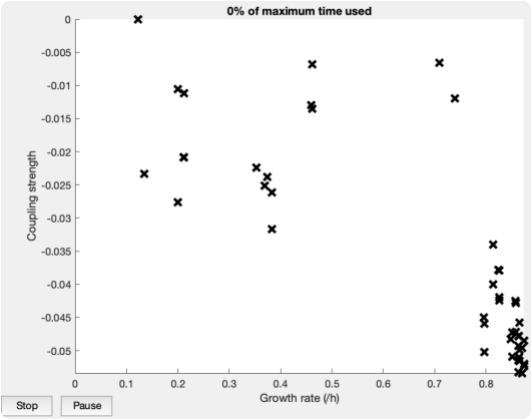


- After clicking OK, another window appears. Here we can redefine optional input parameters with the drop-down menu, defining the new values in the lower text box. Each parameter is assigned a default value, but the default is overridden by any user-defined value. Here, we set the maximum number of KOs to 5 and enable results of all Pareto optimal designs and their performance measures to be outputted into a table in a csv file by setting “saveresults” to true or 1:

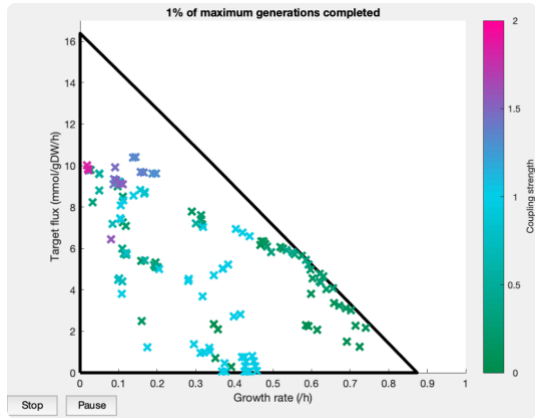


- After clicking on “Start algorithm”, gcFront will then begin - it iteratively solves the multi-objective optimization problem, updating a plot of all the designs of the current generation on a 2D plot of the objective values (i.e. target flux vs growth rate, with designs (dots) coloured by coupling strength).

At first, it may plot the growth rates and coupling strengths for the non-gc-designs identified, as shown here:



As soon as gc-designs are found, the plot is updated to display growth rates (x-axis), product synthesis flux (y-axis), and coupling strengths (dot colouring), of all designs of the current generation, together with the production envelope of the parent genotype (model with no deletions). Users can end the optimization early by pressing the “Stop” button as shown here:

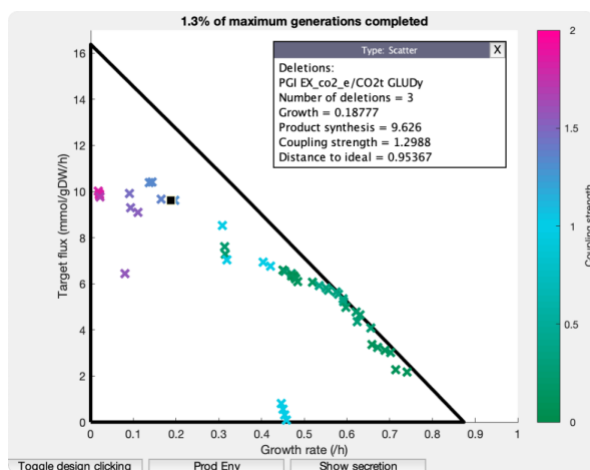


9. Once the algorithm is stopped or reaches a termination condition, only the Pareto front of the final population of designs are displayed on the plot and in the command window. Designs will also be saved as a .csv file in the current folder since the “saveresults” option is enabled by default.

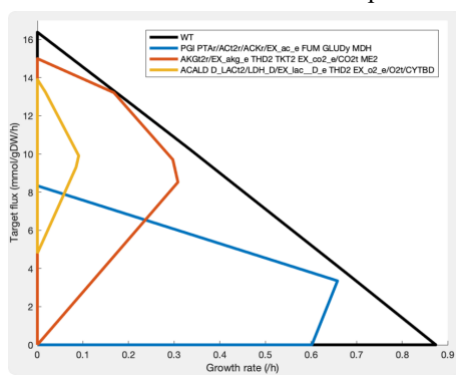
A slash in the name of a deletion indicates that any one of the reactions can be KO'd to achieve the specified performance (so in the example below, the performance of the design on the first line could be achieved with a KO of the reactions PFL, PGI and either EX_co2_e or CO2t).

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 ET0Ht2r/ALCD2x/EX_etoh_e THD2 EX_o2_e/02t/CYTBD"	5	0.017631	10.025	1.8369	1.057
"ACALD D_LACT2/LDH_D/EX_lac__D_e THD2 EX_o2_e/02t/CYTBD"	4	0.090648	9.9108	1.4837	1.0129
"PFL D_LACT2/LDH_D/EX_lac__D_e ET0Ht2r/ALCD2x/EX_etoh_e EX_o2_e/02t/CYTBD"	4	0.020154	9.8848	1.8488	1.0571

An interactive plot of the Pareto front of the discovered designs will also be displayed. Users can click on designs to display information about them:



Also, users can use the “Prod Env” button to display the production envelope of the selected design, and use the “Show secretion” button to show predicted metabolite secretion/uptake for the selected design:



Deletions: PGI EX_co2_e/CO2t GLUDy

Metabolite	MinFlux	MaxFlux	Reaction	ReactionFormula
"H ⁺ "	32.324	32.324	'EX_h_e'	'h_e <=> '
"H ₂ O"	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 -> '
"L-Glutamate"	0	1.6406e-05	'EX_glu_L_e'	'glu_L_e -> '
"Phosphate"	-0.69075	-0.69075	'EX_pi_e'	'pi_e <=> '
"Ammonium"	-1.0239	-1.0239	'EX_nh4_e'	'nh4_e <=> '
"O ₂ O ₂ "	-7.1489	-7.1489	'EX_o2_e'	'o2_e <=> '
"D-Glucose"	-10	-10	'EX_glc_D_e'	'glc_D_e <=> '

- Users must then choose between these designs to find ones that are best suited to their requirements. Alternatively, they can develop bespoke criteria or measures to determine designs that balance between the objectives as they deem suitable. More details of this in the next subsection (B).

B. Examples of ways to harness the results to select gc-designs to take to the lab

gcFront produces a number of designs. To select between these, users must make decisions about the trade-offs between growth, product synthesis, and coupling strength that they are willing to accept. A high growth rate results in faster synthesis processes and raises productivity, high synthesis rates raise both yield and productivity, and a stronger coupling strength can encourage stronger selective pressure for product synthesis and robustness of synthesis to suboptimal growth rates. After choosing a design with favourable parameters, users should assess the biological feasibility of this design if this was not already done at an earlier stage. If toxic metabolite production was not constrained in the model, then users should check what byproducts their coupled designs are expected to result by clicking on the “Show secretion” button. If substantial toxic byproduct secretion is predicted, the user should then check whether their design is still functional after constraining the secretion of the toxic byproduct. In addition, if gcFront was not used to search for gene knockouts, then users must identify genes affiliated with their reactions to convert their reaction knockouts into a strategy that can be experimentally implemented. Furthermore, if essential gene knockouts were not already excluded from consideration, then biological knowledge should be used to assess the feasibility of the identified gene knockouts, as some genes that are not essential *in silico* cannot be knocked out experimentally.

References

King,Z.A. *et al.* (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.*, **44**, D515–D522.