

## Simulation codes

### 1. Folder “iMeGroCy”

Folder “iMeGroCy” consists of 2 subfolders, “MeGro” and “GroCy”.

#### MeGro

In folder “MeGro” there is the code for implementing the MeGro model. The folder contains the following files in MATLAB language:

- “MeGro.m” – the main program;
- “GrocyParEval.m” – the program for the evaluation of GroCy parameters from the MeGro outputs;
- “MinusLambda.m” and “EtOHconstr.m” – the objective function and, respectively, the nonlinear constraint required by the *fmincon* to solve the optimization problem in the absence of ethanol;
- “MinusLambda\_NoGlucose\_FixHxt.m” and “EtOHconstr\_NoGlucose\_FixHxt.m” – the objective function and, respectively, the nonlinear constraint required by the *fmincon* to solve the optimization problem in the absence of glucose.

In “MeGro.m” there are two main sections: (i) the first (lines 117-199) related to the simulation in pure glucose environment ( $c_{\text{EtOHex}} = 0$ ); (ii) the second (lines 201-289) related to the simulation in pure ethanol environment ( $c_{\text{glcex}} = 0$ ). The user can simultaneously run both sections or only one, by commenting the other one.

In the pure glucose section (i), the user must insert the values of the external glucose environment and of the related fermentative ratio, as entries of the vectors “Data\_c\_glcex” and “Data\_F” (please match the entry ordering of the two vectors).

In the pure ethanol section (ii), the user must insert the values of the external ethanol environment and of the related values for the fermentative ratio and transporter investment, as entries of the vectors “Data\_c\_EtOHex”, “Data\_F” and “Data\_gamma\_hxt”.

The main variables (both inputs and outputs) of the simulation are printed in the folder “Result” as “.txt” files, denoted by the suffix “\_Exper1.txt”, for the variables of section (i), and by “\_Exper2.txt”, for the output of section (ii).

In particular, the user can find the optimal values of the maximal growth rate ( $h^{-1}$ ) and of the ribosome-over-protein ratio in the output files “Lambda\_Exper1.txt” and “Rho\_Exper1.txt”, for section (i), and “Lambda\_Exper2.txt” and “Rho\_Exper2.txt”, for section (ii). The rows of these output files match the ordering of the input vectors introduced by the user (which are also printed as .txt files with straightforward names in the same folder).

Finally, the “GrocyParEval.m” subprogram (run automatically by the main program “MeGro.m” at the end of its implementation) computes the GroCy parameters on the basis of the optimal MeGro outputs, saving the computed quantities as .txt files in the subfolder “iMeGroCyParam” (placed inside the folder “Results”). Each row of the .txt files refers to an external condition set in the two sections of “MeGro.m”, following the section sequence (results of section (ii), if any, are appended to those of Section (i), if any) and the same ordering of the input vectors set by the user. In particular, the user can find:

- The values of the rate constant of protein synthesis  $K_2$ , where column 1 refers to daughter cells while columns from 2 to 7 refer to parent cells of age from 1 to 6, for all the experiments;
- The values of rho for all the experiments.

## GroCy

In folder “Grocy” there are two subfolders: “GroCy\_Chain” and “GroCy\_Population”. Folder “GroCy\_Chain” contains four files in MATLAB language:

- “GC\_Model.m” – the main program;
- “GCMfun.m” – the function providing the ODE equations for the G1 phase, as well as the budded phase before the reset;
- “GCMfun\_reset” – the function providing the ODE equations for G1\* phase;
- “eta\_f.m” – the function of the specific degradation rate for Far1.

The code “GC\_Model.m” can be used to simulate a chain of yeast cells growing in different nutrient environments. In the current version, the simulation reproduces the chain of parent cells from  $P_1$  to  $P_7$ , born from a first daughter cell  $D_1$  (or equivalently the code can be exploited to simulate the chain  $P_{k+1}$ - $P_{k+7}$ , born from a parent cell  $P_k$ ). There are two parts in “GC\_Model.m”: the first part simulates the cell chain growing in 0.05% glucose (at the beginning of the code), while the second part in 2% glucose (after the “break” line).

In order to perform a simulation, the user needs to open the file “GC\_Model.m” using MATLAB and set the parameters of each part (at the beginning of the code, for 0.05% glucose and, soon after, for 2% glucose).

The program displays the time behaviour of the following quantities related to the first cell of the chain: 1) time evolution of protein and ribosome contents (Figures 1 and 5 for 0.05% and 2% glucose, respectively) and 2) time evolution of the molecular players (Figures 2 and 6 for 0.05% and 2% glucose, respectively). Moreover, it shows two plots related to the chain  $P_1$ - $P_7$ , that is 3) time evolution of the protein content from birth (scaled to zero) to division of each cell  $P_1$ - $P_7$  (Figures 3 and 7 for 0.05% and 2% glucose, respectively); 4) behaviour of  $P_s$  as a function of the increasing genealogical age  $k$  (Figures 4 and 8 for 0.05% and 2% glucose, respectively). All pictures are saved as “.fig” and “.png” files.

Inside the folder “GroCy\_Population” there is an executable file written in java language that can perform simulations of a whole yeast population. In order to perform such a simulation, the user needs to set the final size of the population (i.e. the final number of cells), by choosing the value of the variable “size” reported in “info\_simulation.txt”. A deterministic or stochastic simulation can be performed setting, respectively, “none” or “gaussian” in the field “rand” of “info\_simulation.txt” (selecting “gaussian”, the parameters are sampled from a lognormal distribution). The user can also change the values of the model parameters reported in “parameters.txt” contained in the folder “config”. There is a perfect match between the parameter names used in the code and the names reported in the manuscript (except for “avgT1”, “k3”, “k4” that correspond, respectively, to  $T_{1b,min}$ ,  $K_{cn}$ ,  $K_{nc}$  in Palumbo et al, 2018). In the present version, the parameters in “parameters.txt” are tuned for 2% glucose. The user can also select the type of the first cell (setting the variable “cellType” in “parameters.txt” to 0, for a parent, or to 1, for a daughter) and its genealogical age (setting the variable “cellAge” in “parameters.txt” to the desired age).

In order to run the simulation, the user needs to open the Java command window, browse to the folder “GroCy\_Population” containing the file “population.jar” and launch the

executable file by typing the command “java -jar population.jar info\_simulation.txt”. It is possible to follow the progress of the simulation time and of the increasing number of cells of the population from the command window itself.

The execution of the code will create, inside the folder “data”, a new folder (whose name is given by the starting time of the simulation) containing different outputs in “.csv” format.

The main outputs saved into “data” folder are:

- “pt\_daughter\_XXXX.csv”, “pt\_parent\_XXXX.csv” (where “\_XXXX” denotes the final time of the simulation, when the population size overcomes the chosen final size), reporting in the first column protein contents of daughter and parent subpopulations at time XXXX. The second and third columns of the mentioned files report the genealogical age of each cell (the genealogical age for a daughter is simply the mother’s age after the division) and the information on its budding state (1: budded, 0: not budded). The last column shows the initial protein content of the cells.
- “rt\_daughter\_XXXX.csv”, “rt\_parent\_XXXX.csv” reporting ribosome contents for daughters and parents in the first column. Columns 2 and 3 report the same quantities described for “pt\_daughter\_XXXX.csv”, “pt\_parent\_XXXX.csv”.
- “ratio.csv”, showing the time behaviour (from the beginning of the simulation until time “XXXX”) of meaningful quantities. In particular, col1: time; col2: number of cells; col3: fraction of cells in  $T_B$ ; col4: fraction of cells in  $T_{G1}^*$ ; col5: fraction of cells in  $T_2$ ; col6: fraction of cells in  $T_{1b}$ ; col7: fraction of cells in  $T_{1a}$ ; col8: total ribosome content; col9: total protein content; col10: ribosome-over-protein ratio; col11: number of daughters; col12: number of budded daughters; col13: fraction of budded daughters (col12/col11); col14: number of parents; col15: number of budded parents; col16: fraction of budded parents (col15/col14).
- The statistics of the population is reported in two files: (i) “verbose\_XXXX”, collecting the population features at time “\_XXXX”, obtained by freezing the entire population when the chosen final size is reached; (ii) “verbose\_YYYY\_2”, collecting the same features at time “\_YYYY”, i.e. when all the cells (initially frozen at time “\_XXXX”) have completed their cycle. Note that periods  $T_0$ ,  $T_1$ ,  $T_{01}$  of files “verbose\_XXXX” and “verbose\_YYYY\_2” correspond, respectively, to  $T_{1a}$ ,  $T_{1b}$ ,  $T_1$  in the paper.

## 2. Folder “Hy-iMeGroCy”

*GroCy + G1-S transition molecular model*

The folder “G1\_S\_GroCy\_Population” contains six files in MATLAB format:

- “G1\_S\_GroCy\_Pop.m” – the main program;
- “Integration\_T1.m” – the ODE equations of the molecular details for the G1/S transition ( $T_1$  plus  $T_2$  periods);
- “Integration\_TB.m” – the ODE equations for proteins and ribosomes during the budded phase;
- “Generation\_Cl.m” – the function generating the initial conditions for each newborn cell;
- “parameters.m” – the file containing the model parameters;
- “Prob\_matrices.m” – the file for generating the probability matrices of the multisite phosphorylation.

The code "G1\_S\_GroCy\_Pop.m" can be used to simulate a single cell, setting "Ncfinali = 1", or a population of cells, setting "Ncfinali" to a value higher than 1, in different nutrient environments (by suitably tuning the model parameters in "parameters.m"). In the present format, the parameters in "parameters.m" are tuned for 2% glucose. The user can also select the type of the first cell (parent, setting the first entry of "PDET\_0" to 0, or daughter, setting the first entry of "PDET\_0" to 1) and its genealogical age (setting the second entry of "PDET\_0" to the desired age).

The final statistics of the population can be found in the matrices "Period" and "ProteinContent". Each matrix has three columns: the first one refers to the mean values of periods and protein contents of the whole population, while the second and third columns refer to parent and daughter subpopulations, respectively. The rows of "Period" report the mean values of: 1)  $T_1$ , 2)  $T_2$ , 3)  $T_{G1}$  ( $T_1+T_2$ ), 4)  $T_B$ , 5) length of  $G1^*$ , 6) length of the cell cycle ( $T_{cd}$ ). The rows of "ProteinContent" report the mean values of: 1)  $P_0$ , 2)  $P_s$ , 3) protein content at division ( $P_{cd}$ ).

Protein and ribosome contents of the population, at the final time of the simulation, are saved in vectors "ProtDistr" and "RibDistr".

## References

Palumbo, P., Vanoni, M., Papa, F., Busti, S., Wortel, M., Teusink, B., and Alberghina, L. (2018). An integrated model quantitatively describing metabolism, growth and cell cycle in budding yeast (Communications in Computer and Information Science 830), pp. 165-180.