# An Integrated Model Quantitatively Describing Metabolism, Growth and Cell Cycle in Budding Yeast

Pasquale Palumbo[1,2(✉)], Marco Vanoni[1,3], Federico Papa[1,2], Stefano Busti[1,3], Meike Wortel[4,5], Bas Teusink[4], and Lilia Alberghina[1,3]

[1] SYSBIO Centre for Systems Biology, Milan, Italy
pasquale.palumbo@iasi.cnr.it,
{marco.vanoni,lilia.alberghina}@unimib.it
[2] Institute for System Analysis and Computer Science "Antonio Ruberti" – CNR, Via dei Taurini 19, Rome, Italy
[3] Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza Della Scienza 2, Milan, Italy
[4] Systems Bioinformatics, VU University, De Boelelaan 1087, 1081 HV Amsterdam, The Netherlands
[5] Centre for Ecological and Evolutionary Synthesis (CEES), The Department of Biosciences, University of Oslo, Blindernveien 31, 0371 Oslo, Norway

**Abstract.** Computational models are expected to increase understanding of how complex biological functions arise from the interactions of large numbers of gene products and biologically active low molecular weight molecules. Recent studies underline the need to develop quantitative models of the whole cell in order to tackle this challenge and to accelerate biological discoveries.

In this work we describe three major functions of a yeast cell: Metabolism, Growth and Cycle, through two coarse grain models, MeGro (Metabolism + Growth) and GroCy (Growth + Cycle). GroCy effectively recapitulates major phenotypic properties of cells grown in glucose and ethanol supplement media. MeGro can act as a parameter generator for GroCy. The resulting iMeGroCy integrated model can be used as a scaffold for molecularly detailed models of yeast functions.

**Keywords:** Computational models · Systems biology · Whole cell models

## 1 Introduction

*Saccharomyces cerevisiae* is a major eukaryotic model organism in both fundamental and applied research. Computational approaches are required to analyze, structure and integrate the ever-increasing data sets available for yeast. Ultimately, a dynamic, comprehensive computational model of *S. cerevisiae* should be the ambition: it would, in part, allow further improvement of industrial bioprocesses by extending the understanding presently possible by genome-scale metabolic model [1]. It also would allow translation of the methodologies to human cells, as it previously happened for


M. Pelillo et al. (Eds.): WIVACE 2017, CCIS 830, pp. 165–180, 2018.
https://doi.org/10.1007/978-3-319-78658-2_13

genome sequencing, functional analysis and interactomics, just to name a few fields in which yeast research has recently led the way [2].

The design rules followed in the construction of the pioneering *Mycoplasma* whole cell model [3] were to divide the functionality of the cell into modules, each modeled bottom-up for short enough periods of time to assume module independence. Simple translation of this approach to a eukaryote, even as simple as the unicellular budding yeast, may not be straightforward. In fact, in contrast to *Mycoplasma*, yeast has a compartmentalized cellular organization, a ten-fold larger genome [4], sophisticated nutritionally modulated sensing and differentiation pathways [5] and an asymmetric cell division that results in population heterogeneity in terms of size, age and cellular content of individual cells. Accordingly, the successful building of models of cells more complex than *Mycoplasma* may face significant challenges [6] and originate models that are difficult to structure and parametrize.

To deal with yeast complexity we developed a multi-level approach. In the following, we present an integrated coarse grain model of the basic functions of a yeast cell (metabolism, growth and cycle), investigate how they respond to availability of a major yeast nutrient, glucose, and discuss how the model can be used as a scaffold for molecularly detailed models of yeast functions.

## 2 The Metabolism and Growth Model (MeGro)

The Metabolism and Growth Model (MeGro) connects growth and metabolism in *S. cerevisiae*. Growth rate maximization forms a rational basis for explaining growth strategies (see e.g. [7] and references therein), since a faster growing unicellular microrganism will have higher evolutionary fitness than its competitors, producing more progeny per time unit in a given environment. So we considered a coarse grain representation of a yeast cell that maximizes its specific growth rate by allocating total protein synthesis capacity to different protein pools. MeGro - derived from the generic "self-replicator" model proposed in [7] for unicellular microorganisms - is conceived to highlight the common patterns connecting growth rate-dependent regulation of cell size, ribosomal content and metabolic efficiency in a cell. All the metabolic reaction rates, the kinetic parameters and the stoichiometry of the flux balance constraints in MeGro are suitably tuned for *S. cerevisiae* and only the relevant classes of enzymes and metabolites are considered.

MeGro accounts for five classes of proteins and five kinds of metabolites. The proteins with enzymatic activity (square blocks in the MeGro scheme of Fig. 1) are (i) the hexose transporters, '*hxt*', (ii) the glycolytic enzymes, '*gly*', (iii) the ribosomes, '*rib*', (iv) the respiration and (v) fermentation pathways enzymes, '*resp*' and '*ferm*' respectively. Three kinds of metabolites are involved in metabolic conversions (green ovals in the MeGro scheme of Fig. 1): the (a) extracellular and (b) intracellular glucose, '*glc, ex*', and '*glc,in*' respectively, and (c) pyruvate, '*pyr*'; other two kinds of metabolites are involved in energy production/consumption: (d) *ATP* and (e) *ADP*. In the following we indicate with $c_x$, $x \in \{Prot, Met\}$, $Prot = \{hxt, gly, rib, resp, ferm\}$, $Met = \{glc,ex, glc,in, pyr, ATP, ADP\}$, the protein/metabolite concentrations, (mM), and with $v_x$, $x \in Prot$, the metabolite fluxes, (mM/h) catalyzed by a specific protein $x$.

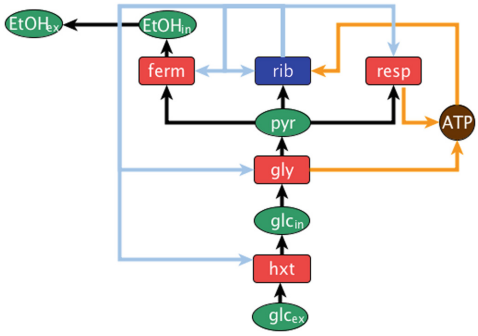



**Fig. 1.** Concept map of MeGro sub-module. (Color figure online)

MeGro captures resource allocation strategies, partitioning the investment into ribosomes in producing the different metabolic proteins. The accumulation of each protein pool is provided by a proper fraction $\alpha_x$ of the net ribosomal flux $v_{rib}$. Thus, in exponential growth conditions we have the following steady-state constraints (see also [7] for analytical details):

$$\lambda c_x - \alpha_x v_{rib} = 0, \qquad x \in Prot \tag{1}$$

where $\lambda$ ($h^{-1}$) is the specific growth rate and $\sum_{x \in Prot} \alpha_x = 1$, with $\alpha_x \geq 0$. Equation (1) refers to a steady state, each protein pool resulting from the balance - not explicitly modeled - of synthesis and degradation. Protein synthesis and degradation are instead explicitly modeled in the GroCy dynamical model, detailed in the next section.

The net dynamics of $c_{glc,in}$, $c_{pyr}$, $c_{ATP}$ are determined by the combination of the fluxes of production and consumption:

$$dc_{glc,in}/dt = v_{hxt} - v_{gly}, \tag{2}$$

$$dc_{pyr}/dt = 2v_{gly} - v_{ferm} - v_{resp} - 600v_{rib}, \tag{3}$$

$$dc_{ATP}/dt = 2v_{gly} + 10v_{resp} - 2000v_{rib}, \tag{4}$$

providing steady-state constraints by imposing the derivatives equal to zero.

The total amount of ADP + ATP is constant, according to the following relationship

$$c_{ATP} + c_{ADP} = 1. \tag{5}$$

All protein and metabolite concentrations $c_x$ are such that $c_x \geq 0$. All the metabolic conversions are catalysed by enzymes and the corresponding fluxes are modeled using the Michaelis-Menten formalism:

$$v_{hxt} = \frac{k_{cat,hxt} c_{glc,ex} c_{hxt}}{(c_{glc,ex} + k_{m,hxt})(1 + c_{glc,in}/k_{i,glui})}, \tag{6}$$

$$v_{gly} = \frac{k_{cat,gly} c_{ADP} c_{glc,in} c_{gly}}{(c_{ADP} c_{glc,in} + k_{m,gly} c_{ADP} + k_{m,ADPgly} c_{glc,in} + k_{gly} k_{m,ADPgly})(1 + c_{pyr/k_{i,pyr}})}, \tag{7}$$

$$v_{rib} = \frac{k_{cat,rib} c_{ATP} c_{pyr} c_{rib}}{c_{ATP} c_{pyr} + k_{m,rib} c_{ATP} + k_{m,ATPrib} c_{pyr} + k_{m,rib} k_{m,ATPrib}}, \tag{8}$$

$$v_{resp} = \frac{k_{cat,resp} c_{ADP} c_{pyr} c_{resp}}{c_{ADP} c_{pyr} + k_{m,resp} c_{ADP} + k_{m,ADPresp} c_{pyr} + k_{m,resp} k_{m,ADPresp}}, \tag{9}$$

$$v_{ferm} = \frac{k_{cat,ferm} c_{pyr} c_{ferm}}{c_{pyr} + k_{m,ferm}}. \tag{10}$$

(1) to (10) define the set of algebraic-differential equations of MeGro. The exponential growth rate $\lambda$ is maximized as a function of the external glucose concentration $c_{glc,ex}$ (model input), with the fractions $\alpha_x$ as optimization variables, and subject to exponential growth constraints (1), flux balance constraints (derived from steady-state Eqs. (2–4)), feasible constraints (5) and Michaelis-Menten flux Eqs. (6–10).

The optimal set of $\lambda$ and $\alpha_{hxt}$, $\alpha_{gly}$, $\alpha_{rib}$, $\alpha_{resp}$, $\alpha_{ferm}$ together with the proteins/metabolites concentrations and the protein fluxes provide a first level of MeGro outputs. A second level of cellular outcomes are computed by properly exploiting concentrations and fluxes. These are (i) the fermentative ratio $F$,

$$F = v_{ferm} / (v_{ferm} + v_{resp}), \tag{11}$$

(ii) the ribosome-over-protein ratio $\rho$,

$$\rho = c_{rib} / \left(600 \left(c_{hxt} + c_{gly} + c_{rib} + c_{resp} + c_{ferm}\right)\right), \tag{12}$$

with proteins expressed in terms of number of polymerized amino acids, which explains the division by 600, the average number of polymerized amino acids per protein [8, 9] and (iii) the yield of ethanol $Y_{EtOH/glc}$,

$$Y_{EtOH/glc} = v_{ferm} / v_{hxt}. \tag{13}$$

If we leave $F$ as an optimization variable, the model predicts that the cell behavior is fully respiratory for values of external glucose smaller than a critical value and then switches to purely fermentative for values of external glucose greater than the threshold. This model behavior (respiratory-to-fermentative switch) is independent of the setting of the model parameters (no threshold mechanism is artificially imposed), instead it is an emergent property of MeGro, with the model parameters allowing the tuning of the value of the external glucose threshold. Such a behavior is coherent with experimental results showing ethanol production only when the dilution rate exceeds a certain level, see e.g. [16].

MeGro can treat the fermentative ratio $F$ as an input rather than an output, thus allowing the modeler to compute the optimal growth rate (as well as all the other model outputs) according to different values of $F$. Indeed, by properly exploiting the flux balance constraints (derived from steady state Eqs. (2–4)) and the fermentative ratio Definition (11), we can write $v_{ferm}$ and $v_{hxt}$ in terms of $F$ and of the ribosomal flux $v_{rib}$:

$$v_{ferm} = (1.4 \times 10^3 F)/(1 + 10(1-F))v_{rib}, \tag{14}$$

$$v_{hxt} = 10^3 \times (1 + 3(1-F))/(1 + 10(1-F))v_{rib}, \tag{15}$$

so that, according to the ethanol yield Definition (13) the fermentative ratio $F$ is provided as a function of a given ethanol yield:

$$F = 20Y_{EtOH/glc}/\,(7 + 15Y_{EtOH/glc}). \tag{16}$$

This last equation will be exploited to feed MeGro with the fermentative ratio associated to experimental yield, Fig. 2A.

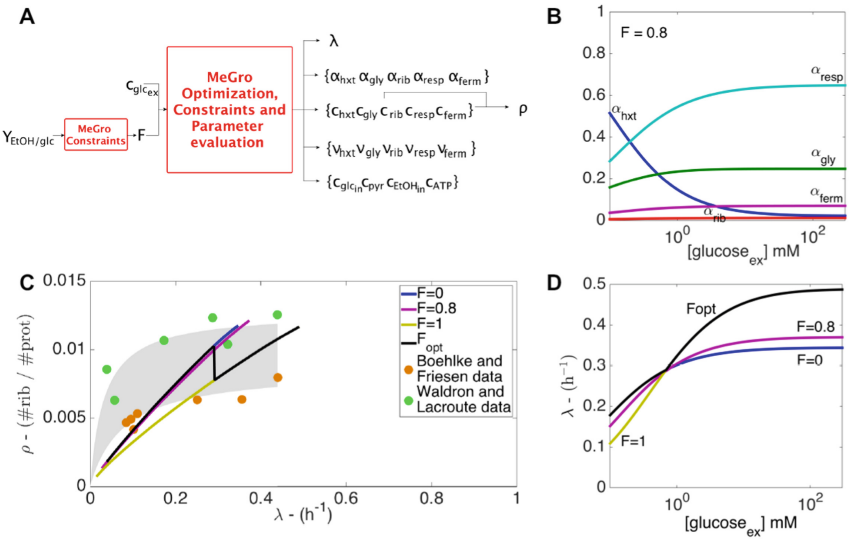



**Fig. 2.** AMeGro outcomes, when both the external glucose concentration and the fermentative ratio (i.e., the yield of ethanol) are exploited as model inputs. **B:** optimal fractions of ribosomal activity ($\alpha_j$) engaged in the synthesis of the corresponding protein modules as functions of the external glucose concentration. **C:** MeGro optimal ribosome-over-protein ratio $\rho$ as a function of MeGro optimal growth rate $\lambda$, with fixed fermentative ratio $F$ (colored curves, $F$ ranging in [0, 1]) and not fixed $F$ (bold black line). MeGro simulations are compared to the experimental data redrawn from [10, 11]: in grey we highlight the region between Michaelis-Menten experimental data best fitting. **D:** MeGro optimal growth rate $\lambda$ as a function of the external glucose concentration $c_{glc,ex}$, with fixed fermentative ratio $F$ (colored curves, $F$ ranging in [0, 1]) and not fixed $F$ (bold black line). (Color figure online)

Figure 2B reports the steady-state protein fluxes for the different protein pools as a function of glucose concentration for a fermentative ratio $F = 0.8$. $\lambda$ increases as a function of external glucose concentration following a saturation kinetics, whose parameters depend on the fermentative ratio $F$ (Fig. 2D). Figure 2C shows the behavior of the ribosome-over-protein ratio $\rho$ as a function of the external glucose concentration, at different fixed values of the fermentative ratio $F$. Model predictions (solid lines) are compared with two sets of experimental data (green and orange circles) from different yeast strains [10, 11], showing overall agreement between model predictions and experimental data. MeGro parameters can be found in Table 1. Most parameters are chosen by following the criteria developed in [12], with minor modifications, mostly related to the use of different units.

**Table 1.** MeGro parameters. Most parameters are chosen by following to the same criteria developed in [12], with minor modifications, mostly related to the use of different units. To determine the parameters $k_{cat,x}$, $x \in \{Prot\}$, we used experimental data on the specific growth rate of yeast cell populations growing in batch cultures at different glucose concentrations and literature data of yeast cells in chemostat. Since the maximal growth rate is reached by fully fermenting cells, we tuned $k_{cat,x}$ (except $k_{cat,resp}$) in order to fit our maximal experimental growth rate of 0.424 $h^{-1}$ at a glucose concentration of 278 mM, obtaining a maximal growth rate of 0.48 $h^{-1}$ for fully fermenting cells at saturating glucose concentrations. In order to tune $k_{cat,resp}$ we consider literature data reporting the growth rate at which the switch from respiration to fermentation occurs. According to such data, *S. cerevisiae* in a chemostat starts to produce ethanol at a dilution rate between 0.25 and 0.28 $h^{-1}$ [13]. Then we set $k_{cat,resp}$ such that an equal growth rate of about 0.28 $h^{-1}$ is achieved either using the fermentation pathway or the respiration pathway.

| Parameter | Meas. unit | Value | Parameter | Meas. unit | Value |
|---|---|---|---|---|---|
| $k_{cat,hxt}$ | $h^{-1}$ | 37492 | $\boldsymbol{k_{m,\ hxt}}$ | mM | 20 |
| $\boldsymbol{k_{cat,gly}}$ | $h^{-1}$ | 4166 | $\boldsymbol{k_{m,\ gly}}$ | mM | 0.2 |
| $\boldsymbol{k_{cat,\ rib}}$ | $h^{-1}$ | 670 | $\boldsymbol{k_{m,\ rib}}$, $\boldsymbol{k_{i,\ pyr}}$, $\boldsymbol{k_{i,\ glui}}$ | mM | 1 |
| $\boldsymbol{k_{cat,\ resp}}$ | $h^{-1}$ | 99 | $\boldsymbol{k_{m,\ resp}}$, $\boldsymbol{k_{m,\ ADPgly}}$, $\boldsymbol{k_{m,\ ADPresp}}$, $\boldsymbol{k_{m,\ ATPrib}}$ | mM | 0.5 |
| $\boldsymbol{k_{cat,\ ferm}}$ | $h^{-1}$ | 6427 | $\boldsymbol{k_{m,\ ferm}}$ | mM | 5 |

## 3 The Growth and Cycle Model (GroCy)

In yeast, the critical cell size required to enter S phase ($P_S$) is modulated by nutrient availability [14]. It remains small and nearly constant when glucose is utilized through respiration. In contrast, $P_S$ and hence average protein content increases as cells shift their metabolism towards fermentation [15]. Cells forced to ferment under slow-growing conditions show the same increase [16].

GroCy is composed by three modules (Fig. 3): (1) a dynamical cell growth model in which a set of ordinary differential equations describes dynamics of synthesis and degradation of ribosomes and proteins; (2) a molecular triggering mechanism that links cell growth and cell cycle. It exploits a set of ordinary differential equations which detail the dynamics of the growth-controlled activator Cdk1Cln3 and of its cognate

inhibitor Far1; (3) a cell cycle module, that consists of three consecutive timers ($T_{1b}$, $T_2$ and $T_B$) that describe the cycle progression after the triggering mechanism activates the first timer $T_{1b}$. The period that leads from the birth of the cell up to the time instant when the molecular machinery triggers the first timer is denoted by $T_{1a}$.

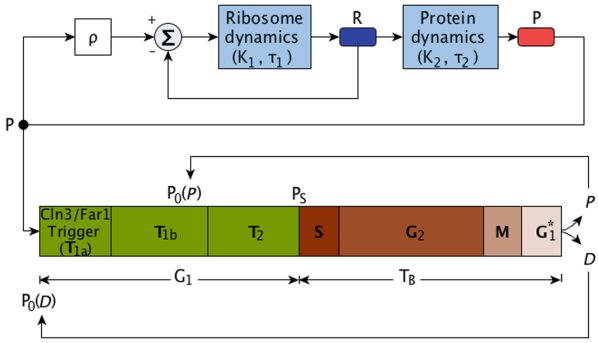



**Fig. 3.** Concept map of GroCy sub-module.

### 3.1 The Growth Module

The growth module deals with the ribosome content $R$, expressed as number of ribosomes per cell (rib), and the protein content $P$, expressed as number of polymerized amino acids per cell (aa), and is taken from [17] (where the reader can find the equations and the details which are below briefly recalled). Both ribosome and protein dynamics are described by the balance between production and degradation rates. Figure 4A shows the time course of the protein content and of the number of ribosomes for two different parameter settings: fast growth (2% glucose, solid line) or slow growth (ethanol, dashed line).

For each steady-state growth condition, the target ribosome/protein ratio $\rho$ is an output of MeGro that can be directly fed into GroCy, providing the link between the two models. According to the model, when the ratio $R/P$ is greater than $\rho$, then there is no ribosome production; otherwise, the ribosome production rate is proportional to the (positive) difference $\rho P - R$. Denoting with $K_2$, $\tau_2$, the average translational efficiency and the protein degradation time constant, respectively, it can be shown that, provided the exponential growth condition is satisfied, $\rho K_2 - 1/\tau_2 > 0$, the ratio $R/P$ asymptotically converges to the value of parameter $\rho$. The exponential growth condition ensures that both ribosomes and proteins grow according to the same exponential law, with an exponential growth rate $\lambda$ ($\text{min}^{-1}$) given by: $\lambda = \rho K_2 - 1/\tau_2$.

$\lambda$ is not hard-wired in the model, but rather it is linked to the macromolecular composition and biosynthetic activity of the cells, a connection whose detection is made possible by the appropriate choice of the measurement units for ribosome and protein content, synthesis and degradation.

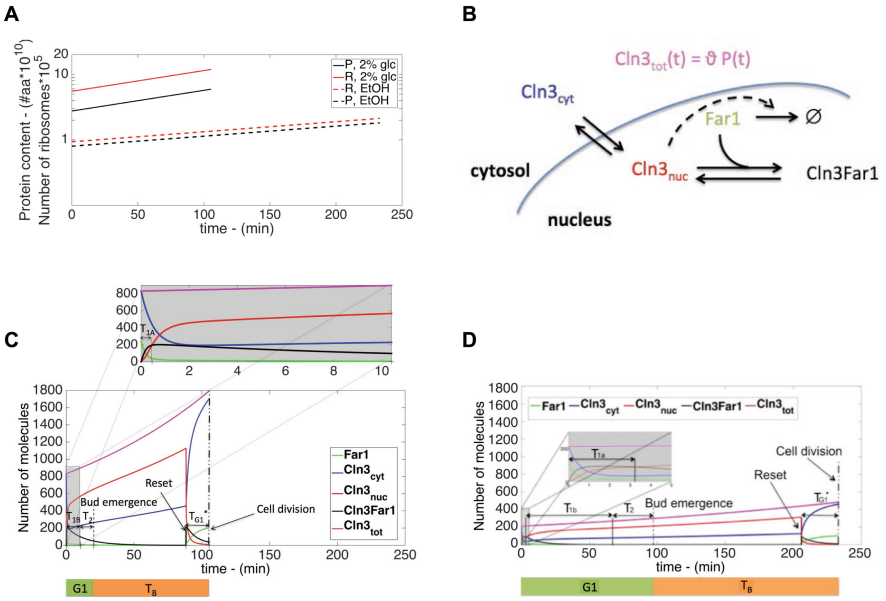



**Fig. 4.** ATime course of the protein content P and of the number of ribosomes R for fast (2% glucose, solid line) and slow growth conditions (ethanol, dashed line). **B:** schematic view of the interactions among the molecular players involved in the GroCy molecular triggering mechanism. **C-D:** time evolution of the players of the molecular triggering module, for fast (2% glucose, panel C) and slow (ethanol, panel D) growth conditions.

### 3.2 The Molecular Trigger Linking Growth to Cell Cycle Initiation

In budding yeast the entrance into S phase and budding starts when cells reach a critical cell size, thus connecting growth and cycle [18, 19]. Cln3 is an activator of S phase entrance, whose amount is proportional to the overall protein content, Eq. (20), therefore linking the growth and cycle modules. Despite some discordant results discussed in [20], we take that Cln3 accumulation is constant during G1. Cln3 production takes place in the cytoplasm. Cytoplasmic Cln3 is defined straightforwardly by Eq. (21). Nuclear volume is a constant fraction of total cell volume throughout the cycle [21].

The cyclin-dependent kinase inhibitor Far1 is involved in the cell size control mechanism in cycling cells by inhibiting Cln3 in early G1 [22, 23]. After mitosis, newly synthesized Far1 is endowed to each nucleus [24]. Ensuing Cln3 nuclear transport and accumulation allows overcoming of Far1 inhibition, which is made irreversible by Far1 degradation primed by the rising Cln3 activity [25, 26].

Here we use a simplified version of the equations used in [27] to model the molecular interplay between Cln3 and Far1, which we call the molecular triggering mechanism. Cdk1 – present in excess over its regulatory subunits – is implied, but not explicitly modeled (Fig. 4B). Cln3 transport in the nucleus and Cln3/Far1 interaction follow mass action kinetics. Far1 degradation is governed by rate $\eta$ ($\text{min}^{-1}$) modeled

according to a Hill function (Eq. (21)), that increases from 0 to a high level $\eta^*$ as soon as the free nuclear Cln3 exceeds Cln3Far1. $nF$ is the Hill coefficient, modeling the steepness of the Hill function. Equation (19) accounts for reversible nucleo-cytoplasmic transport of Cln3 and interaction with Far1.

$$dCln3Far1/dt = (k_{on}/V_{nuc})Clin3_{nuc}Far1 - k_{off}Clin3Far1, \tag{17}$$

$$dFar1/dt = -(k_{on}/V_{nuc})Cln3_{nuc}Far1 + k_{off}Cln3Far1 - \eta(Cln3_{nuc}/Cln3Far1)Far1, \tag{18}$$

$$dCln3_{nuc}/dt = -(k_{on}/V_{nuc})Cln3_{nuc}Far1 + k_{off}Cln3Far1 + k_{cn}Cln3_{cyt} - k_{nc}Cln3_{nuc}, \tag{19}$$

$$Cln3_{cyt} = Cln3_{tot} - (Cln3_{nuc} + Cln3Far1), Cln3_{tot} = \theta P, \tag{20}$$

$$V_{nuc} = hV_{cell}, V_{cell} = P/H, \eta(x) = \eta^* x^{nF}/\left(1 + x^{nF}\right). \tag{21}$$

When 80% of the budded period has elapsed, a RESET function takes place and the $G1^*$ phase begins. RESET denotes the time instant when nuclear division (but not cell division) occurs: i.e. during the $G1^*$ phase each cell has two G1 nuclei and an undivided cytoplasm. At RESET the nuclear players show a discontinuity because they no more represent the whole (and unique) nuclear content and also because Far1 has been reset to a higher value. The RESET function includes instantaneous synthesis of Far1 and equal partition of Cln3Far1, Far1, $Cln3_{nuc}$ in two nuclei whose volume is half of the original volume before RESET: $V_{nuc} = hV_{cell}/2$. Far1 degradation is inhibited ($\eta = 0$ in Eq. (18)), and Cln3 diffusion from the cytoplasm into the nucleus is strongly reduced ($k_{cn}$ in Eq. (19) reduces of 5 orders of magnitudes during the $G_1^*$).

Figure 4C–D shows the time course for the different molecular players throughout the whole cycle of an average size cell, growing in fast conditions (2% glucose, panel C) or in slow conditions (ethanol, panel D): when - very early after division - free nuclear Cln3 overcomes its inhibited form Cln3Far1 the first of the three consecutive Timers related to the cell cycle module is triggered. The time period spanning from the birth of the cell up to the aforementioned time instant is named $T_{1a}$. The kinetic parameters of the molecular trigger do not vary in different nutrient environments, except for the total amount of Far1, known to diminish in poor media [22], and for the parameters $H$, $\theta$ assumed to decrease in case of poor growth (see Table 2).

### 3.3 The Cell Cycle Module

In *S. cerevisiae* cell mass at division is unequally partitioned [19] between a larger, old parent cell (P) and a smaller, newly synthesized daughter cell (D). The degree of asymmetry of cell division in *S. cerevisiae* is modulated by nutrients: poor media – such as ethanol - yield a high level of asymmetry with large parent cells and very small daughter cells, whereas in rich media - such as glucose - parents and daughters at division are very close in size (reviewed in [15]). Since cells have to grow to a critical

**Table 2.** GroCy parameters.

| Parameters | Meas. unit | glc 2% | Ethanol | Parameters | Meas. unit | glc 2% | Ethanol |
|---|---|---|---|---|---|---|---|
| $\rho$ | rib/aa | 2.02e−5 | 1.18e−5 | $\tau_{2s}$, $s \geq 1$ | min | 1500 | 3000 |
| $P(0)$ | aa | 2.76e10 | 0.8e10 | $k_{on}$ | $(\text{molec/L})^{-1}/\text{min}$ | 1.63e−15 | 1.63e−15 |
| $R(0)$ | rib | 5.57e5 | 0.94e5 | | | | |
| $Far1(0)$ | molec | 240 | 110 | $k_{off}$ | $\text{min}^{-1}$ | 25 | 25 |
| $Cln3_{nuc}(0)$ | molec | 0 | 0 | $h$ | – | 0.07 | 0.07 |
| $Cln3Far1(0)$ | molec | 0 | 0 | $H$ | aa/L | 7.09e23 | 6.18e23 |
| $Far1_{reset}$ | molec | 240 | 110 | $n_F$ | – | 10 | 10 |
| $K_1$ | $\text{min}^{-1}$ | 1 | 0.6 | $\bar{\eta}$ | $\text{min}^{-1}$ | 1 | 1 |
| $\tau_1$ | min | 4000 | 2000 | $\Theta$ | molec/aa | 3.02e−8 | 2.66e−8 |
| $K_2$ | aa/rib/min | 380 | 316.66 | $k_{cn}$ | $\text{min}^{-1}$ | 1.5 | 1.5 |
| $K_2^1$ | aa/rib/min | 342 | 285.46 | $k_{nc}$ | $\text{min}^{-1}$ | 0.6 | 0.6 |
| $K_2^2$ | aa/rib/min | 178 | 149.34 | $k_{cn,reset}$ | $\text{min}^{-1}$ | 5e−4 | 5e−4 |
| $K_2^3$ | aa/rib/min | 69 | 58.59 | $T_{1b,min}$ | min | 1 | 8 |
| $K_2^4$ | aa/rib/min | 51 | 43.47 | $W_0$ | min | 1503 | 7045 |
| $K_2^5$ | aa/rib/min | 42 | 35.91 | $W_1$ | min | 62.1 | 306 |
| $K_2^s$, $s > 5$ | aa/rib/min | 35 | 29.86 | $T_2$ | min | 10 | 30 |
| $\tau_2$ | min | 3000 | 6000 | $T_B$ | min | 85 | 136 |

cell size before entering S phase and budding, small daughter cells have a longer cycle time than the corresponding parent cells, most notably in poor media. This difference in cycle time between daughter and parent cells is due to differences in the G1 phase, whilst the budded period $T_B$ has essentially the same length in both parents and daughter cells [18]. Differences in growth rate have marginal effects on the length of $T_B$ and dramatic effects on the length of G1 (reviewed in [15]).

As explained, $T_{1a}$ is the period from the birth of a cell till the time instant when free nuclear Cln3 exceeds its inhibited form Cln3Far1; the rest of the cycle is modeled by the sequence of three consecutive timers $T_{1b}$, $T_2$, and $T_B$. The sum of the period $T_{1a}$ + timer $T_{1b}$ corresponds to timer $T_1$ in [18]. The G1 phase is given by $T_1 + T_2$. Timer $T_B$ encompasses the budded phase.

The first timer $T_{1b}$ starts when free Cln3 exceeds its inhibited form Cln3Far1. The length of $T_{1b}$ is related to the size of the cell, so that larger cells have smaller $T_{1b}$ periods, and vice versa. More in details, $T_{1b}$ length is set according to the equation

$$T_{1b} = \max\left\{T_{1b,min}, W_0 - W_1 \ln(P_{T1a})\right\}, \tag{22}$$

with $P_{T1a}$ denoting the size of the cell at the end of $T_{1a}$. Notice that $P_{T1a}$ plays an active role in the setting of $T_{1b}$ only for cells small enough, i.e. only when:

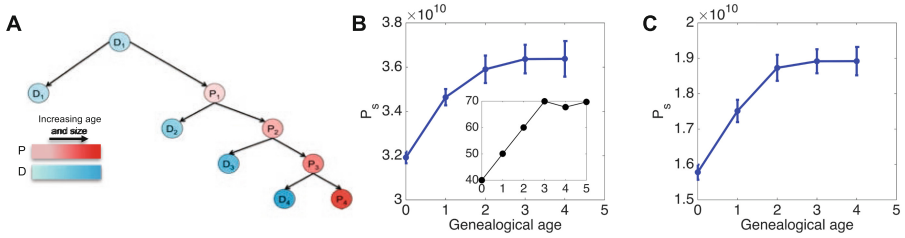
$$W_0 - W_1 \ln(P_{T1a}) > T_{1b,min} \rightarrow P_{T1a} < \exp\left\{\left(W_0 - T_{1b,min}\right)/W_1\right\}. \tag{23}$$

This happens, for instance, with most of daughter cells. In parent cells $P_{T1a}$ is, usually, greater than the upper bound in inequality (23), so that their $T_{1b}$ length is fixed to $T_{1b,\ min}$ and does not depend on the size.

The length of timer $T_2$ does not depend on protein content [18]. At the end of timer $T_2$, the critical protein size $P$ is estimated. The budded period $T_B$, includes the S, G2, M and G1* phases. G1* has been modeled as the last 20% period of $T_B$ phase. The end of the timer results in cell division. Like timer $T_2$, timer $T_B$ length does not depend on protein content, (no difference between daughters and parents). Part of the GroCy parameters are influenced by - and vary according to - the nutrient environment.

### 3.4 Genealogical Age Heterogeneity and Pedigree Simulations

When a yeast cell buds, a chitin ring, called bud scar, is formed at the bud isthmus remaining on the Parent after bud separation [15]. The genealogical age 'k' of a parent cell is the same as the number of bud scars 's', that can be visually counted, since each new bud starts at a new site. A cell without bud scars (s = 0) is a Daughter cell and it has not yet completed a cycle. We denote by "$D_k$" a Daughter of genealogical age 'k' (Fig. 5A). Each $D_k$ (k > 1) is born from a $P_{k-1}$ Parent. $D_1$ are born from any $D_k$.




**D**

| | Parent and daughter cell generation times | | | | | | | | Parent and daughter cell frequencies | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $T_{P1}$ | $T_{P2}$ | $T_{P3}$ | $T_{P4}$ | $T_{D1}$ | $T_{D2}$ | $T_{D3}$ | $T_{D4}$ | $F_{P1}$ | $F_{P2}$ | $F_{P3}$ | $F_{P4}$ | $F_{D1}$ | $F_{D2}$ | $F_{D3}$ | $F_{D4}$ |
| 2% glc | 98.0 | 95.6 | 93.9 | 96.2 | 106.9 | 102.6 | 101.9 | 104.0 | 0.25 | 0.13 | 0.07 | 0.04 | 0.28 | 0.13 | 0.07 | 0.03 |
| EtOH | 175.4 | 174.2 | 178.1 | 171.9 | 188.5 | 179.3 | 177.9 | 181.4 | 0.24 | 0.12 | 0.06 | 0.04 | 0.29 | 0.14 | 0.07 | 0.04 |

**E**

| | T | | $T_B$ | | $T_P$ | | $T_D$ | | $2^{-T_P/T} + 2^{-T_D/T}$ | | <P> | | $<P>_{2\%}/<P>_{EtOH}$ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | exp | comp | exp | comp | exp | comp | exp | comp | exp | comp | exp | comp | exp | comp |
| 2% glc | 97 | 100.9 | 81 | 84.9 | 90 | 96.6 | 104 | 105.0 | 1.003 | 1.001 | 508 | 4.37e10 | 2.52 | 2.47 |
| EtOH | 195 | 170.3 | 136 | 136.4 | 177 | 175.1 | 208 | 184.3 | 1.009 | 0.963 | 201 | 1.77e10 | | |

**Fig. 5.** AChain of cells $P_1$−$P_4$, $D_1$−$D_4$. **B, C:** Computed *Ps* (number of polymerized amino acid) for Parents of increasing genealogical age (until age 4), for fast (2% glucose, panel B) and slow (ethanol, panel C) growth conditions. Panel B reports the experimentally determined volume at bud initiation for Parents of increasing genealogical age (redrawn from [28]). **D:** Generation times, $T_{Pk}$, $T_{Dk}$, for Parents and Daughters of increasing genealogical ages obtained by simulating 20 different chains of cells $P_1$, …, $P_4$ and $D_1$, …, $D_4$. Frequencies of each sub-population ($F_{Pk}$, $F_{Dk}$) have been obtained using eqs. (A12, 13) in [30]. **E:** experimental and simulated values for relevant population parameters. $T_D$ and $T_P$ have been calculated from data in panel D, using eqs. (A4, 5) in [30].

Each Parent increases in size before starting to bud [28]. At division, it receives the mass it had at budding, while the mass synthesized during the budding phase goes to the newborn daughter. Hence, cell mass at budding (in Parents) and cell size at birth (in Daughters) increase with genealogical age. A reduced increase in parent cell size at budding with increasing genealogical age has been reported [15, 28] (see inset in Fig Fig. 5B) and explained by mechanical stress of the cell wall, which increases with cell size [29]. Both $K_2$ and $\tau_2$ in the growth module of GroCy (rate of protein synthesis and time constant of protein degradation respectively) decrease in value during the pre-budded period (G1 phase), according to the parent genealogical age, returning to their nominal values at the onset of the budding phase (end of Timer $T_2$), so that the parent cell $P_k$ grows again with the steady-state exponential rate given by $\lambda = \rho K_2 - 1/\tau_2$. Daughter cells (of any genealogical age) are not affected by such a mechanical stress. The behavior of $P_S$ qualitatively recapitulates experimental data (see Fig. 5B, C).

GroCy may be used to replicate small pedigree populations in different nutritional conditions, by suitably setting its parameters. We simulated 20 chains of cells $D_1$, …, $D_4$, $P_1$, …, $P_4$ (Fig. 5A), starting from 20 different initial cells, for two distinct growth conditions: 2% glucose and ethanol. The timers $T_{1b}$, $T_2$, $T_B$ and the initial protein content $P(0)$ of each cell have been allowed to vary, with log-normal distribution, with a 5% CV over their average values. In order to estimate the average cell cycle length for both subpopulations of parents and daughters ($T_P$, $T_D$), we need to estimate the fractions of parents and daughters from the aforementioned "chain-cells" simulation. To this end we adopt the population modeling approximation described by Eq. (A1) given in [29] that provides the critical size of a parent $P_k$ as a function of the critical size of a daughter and of the pair of parameters $a$, $Q < 1$. In exponential growth, the cell cycle length of parents and daughters of any genealogical age can be computed by means of Eqs. (A4, A5) of the same paper, where the parameter $\alpha$ denotes the exponential growth rate. Since these lengths are provided by the "chain-cells" simulation, we exploit the mentioned equations to infer the information on the population growth rate $\alpha$ and to estimate the values of ($a$, $Q$) that best fit these data. Parameters ($a$, $Q$), as well as the growth rate $\alpha$, are finally exploited to derive the fractions of cells ($F_{P1}$, …, $F_{P4}$, $F_{D1}$, …, $F_{D4}$) by way of the age distribution function [29]. The inferred structure population (Fig. 5D) allows to compute $T_P$ and $T_D$, and the average protein content for the whole population, $< P >$. Relative protein content, mass duplication times ($T$), $T_P$ and $T_D$ of yeasts growing on different media are very similar to experimental values. The relationship $2^{-TD/T} + 2^{-TP/T}$ - that links together $T$, $T_P$ and $T_D$ - yields a number very close to the theoretical value of 1 [15, 19], confirming that the simulated parameters capture the structure of yeasts growing on different carbon sources.

## 4 Conclusions

The growth activity combined to the other two main cellular activities of metabolism and cycle (MeGro and GroCy, respectively) define the modular building blocks constituting the coarse grain backbone of a modular, hierarchical and *i*ntegrated *Me*tabolism, *Gro*wth and *Cy*cle model: iMeGroCy. The light green box in Fig. 6 reports a functional scheme highlighting the general procedure that allows to inter-connect

MeGro and GroCy as a function of the nutritional input (i.e., the glucose concentration). The current version of MeGro does not allow carbon sources other than glucose. MeGro responds to the external glucose $c_{glc,ex}$ and $Y_{EtOH/glc}$ coming from experimental data in order to set the steady-state exponential growth rate $\lambda$ and ribosome-over-protein ratio $\rho$ as outputs of an optimization algorithm aiming at maximizing the growth rate. The MeGro outputs $\lambda$ and $\rho$ enter GroCy as inputs, allowing to set the ribosome and protein dynamics parameters, constituting the growth module. The exponential growth relationship $\lambda = \rho K_2 - 1/\tau_2$ is used to constrain the GroCy parameters $K_2$ and $\tau_2$ to the MeGro outputs ($\lambda$, $\rho$). $K_1$ and $\tau_1$ have been fixed accordingly to [18]. So, for cells growing on glucose-containing media, MeGro acts as a parameter generator for GroCy.

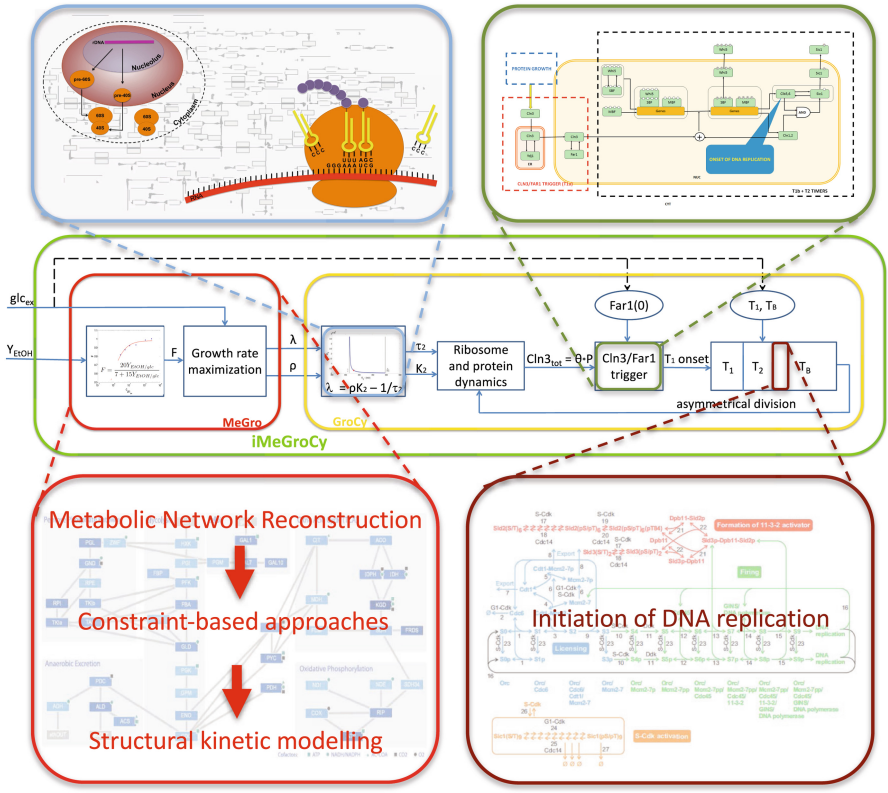



**Fig. 6.** Scheme of iMeGroCy (light green box). The scheme depicts the interconnection of the two main sub-blocks (MeGro, red block, and the GroCy, yellow block). The figure also shows iMeGroCy could host molecular blow-ups (plug-ins) of yeast functions. (Color figure online)

iMeGroCy differs from previous cell cycle models that either relied on defined molecular networks [31] - encompassing 27 components out of the much larger identified number [14, 32] or - when of low granularity [33] - did not show the same

degree of modularity offered by our approach. Other models concentrated on specific cell cycle phases and could be used in conjunction with iMeGroCy, whose modular and hierarchical nature allows it to act as a scaffold for the construction of a whole cell model for *S. cerevisiae* (Fig. 6). For instance, MeGro could be substituted by a genome-wide model [1], appropriately modified to include connections with cell growth and regulation by nutrients, the $G_1$ timers could be substituted by a recently described $G_1$/S module [20], entrance into S phase by a model of the onset of DNA synthesis [34], the budded phase by a wave of cyclins [35].

Adding the modules incrementally, the ability of iMeGroCy to fit experimental data could be monitored at any step. Top-down definition of the molecular modules would allow coherent expansion of iMeGroCy, favoring collaboration within the yeast community, since such an ambitious large-scale project will require a new type of collaborative effort [36].

## References

1. Sánchez, B.J., Nielsen, J.: Genome scale models of yeast: towards standardized evaluation and consistent omic integration. Integr. Biol. (Camb) **7**, 846–858 (2015)
2. Botstein, D., Fink, G.R.: Yeast: an experimental organism for 21st Century biology. Genetics **189**, 695–704 (2011)
3. Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Assad-Garcia, N., Glass, J.I., Covert, M.W.: A whole-cell computational model predicts phenotype from genotype. Cell **150**, 389–401 (2012)
4. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G.: Life with 6000 genes. Science **274**(546), 563–567 (1996)
5. Conrad, M., Schothorst, J., Kankipati, H.N., Van Zeebroeck, G., Rubio-Texeira, M., Thevelein, J.M.: Nutrient sensing and signaling in the yeast Saccharomyces cerevisiae. FEMS Microbiol. Rev. **38**, 254–299 (2014)
6. Macklin, D.N., Ruggero, N.A., Covert, M.W.: The future of whole-cell modeling. Curr. Opin. Biotechnol. **28**, 111–115 (2014)
7. Molenaar, D., van Berlo, R., de Ridder, D., Teusink, B.: Shifts in growth strategies reflect tradeoffs in cellular economics. Mol. Syst. Biol. **5**, 323 (2009)
8. von der Haar, T.: A quantitative estimation of the global translational activity in logarithmically growing yeast cells. BMC Syst. Biol. **2**, 87 (2008)
9. Waldron, C., Jund, R., Lacroute, F.: The elongation rate of proteins of different molecular weight classes in yeast. FEBS Lett. **46**, 11–16 (1974)
10. Boehlke, K.W., Friesen, J.D.: Cellular content of ribonucleic acid and protein in Saccharomyces cerevisiae as a function of exponential growth rate: calculation of the apparent peptide chain elongation rate. J. Bacteriol. **121**, 429–433 (1975)
11. Waldron, C., Lacroute, F.: Effect of growth rate on the amounts of ribosomal and transfer ribonucleic acids in yeast. J. Bacteriol. **122**, 855–865 (1975)
12. Wortel, M.T., Bosdriesz, E., Teusink, B., Bruggeman, F.J.: Evolutionary pressures on microbial metabolic strategies in the chemostat. Sci. Rep. **6**, 29503 (2016)
13. Van Hoek, P., Van Dijken, J.P., Pronk, J.T.: Effect of specific growth rate on fermentative capacity of baker's yeast. Appl. Environ. Microbiol. **64**, 4226–4233 (1998)

14. Alberghina, L., Mavelli, G., Drovandi, G., Palumbo, P., Pessina, S., Tripodi, F., Coccetti, P., Vanoni, M.: Cell growth and cell cycle in Saccharomyces cerevisiae: basic regulatory design and protein-protein interaction network. Biotechnol. Adv. **30**, 52–72 (2012)
15. Porro, D., Vai, M., Vanoni, M., Alberghina, L., Hatzis, C.: Analysis and modeling of growing budding yeast populations at the single cell level. Cytom Part J. Int. Soc. Anal. Cytol. **75**, 114–120 (2009)
16. Porro, D., Brambilla, L., Alberghina, L.: Glucose metabolism and cell size in continuous cultures of Saccharomyces cerevisiae. FEMS Microbiol. Lett. **229**, 165–171 (2003)
17. Alberghina, L., Mariani, L., Martegani, E.: Cell cycle modelling. Biosystems **19**, 23–44 (1986)
18. Di Talia, S., Skotheim, J.M., Bean, J.M., Siggia, E.D., Cross, F.R.: The effects of molecular noise and size control on variability in the budding yeast cell cycle. Nature **448**, 947–951 (2007)
19. Hartwell, L.H., Unger, M.W.: Unequal division in Saccharomyces cerevisiae and its implications for the control of cell division. J. Cell Biol. **75**, 422–435 (1977)
20. Palumbo, P., Vanoni, M., Cusimano, V., Busti, S., Marano, F., Manes, C., Alberghina, L.: Whi5 phosphorylation embedded in the $G_1$/S network dynamically controls critical cell size and cell fate. Nat. Commun. **7**, ncomms11372 (2016)
21. Jorgensen, P., Edgington, N.P., Schneider, B.L., Rupes, I., Tyers, M., Futcher, B.: The size of the nucleus increases as yeast cells grow. Mol. Biol. Cell **18**, 3523–3532 (2007)
22. Alberghina, L., Rossi, R.L., Querin, L., Wanke, V., Vanoni, M.: A cell sizer network involving Cln3 and Far1 controls entrance into S phase in the mitotic cycle of budding yeast. J. Cell Biol. **167**, 433–443 (2004)
23. Fu, X., Ng, C., Feng, D., Liang, C.: Cdc48p is required for the cell cycle commitment point at Start via degradation of the G1-CDK inhibitor Far1p. J. Cell Biol. **163**, 21 (2003)
24. McKinney, J.D., Chang, F., Heintz, N., Cross, F.R.: Negative regulation of FAR1 at the Start of the yeast cell cycle. Genes Dev. **7**, 833–843 (1993)
25. Chang, F., Herskowitz, I.: Phosphorylation of FAR1 in response to alpha-factor: a possible requirement for cell-cycle arrest. Mol. Biol. Cell **3**, 445–450 (1992)
26. Peter, M., Gartner, A., Horecka, J., Ammerer, G., Herskowitz, I.: FAR1 links the signal transduction pathway to the cell cycle machinery in yeast. Cell **73**, 747–760 (1993)
27. Barberis, M., Klipp, E., Vanoni, M., Alberghina, L.: Cell size at S Phase Initiation: An Emergent Property of the G1/S Network. PLoS Comput. Biol. **3**, e64 (2007)
28. Johnston, G.C., Ehrhardt, C.W., Lorincz, A., Carter, B.L.: Regulation of cell size in the yeast Saccharomyces cerevisiae. J. Bacteriol. **137**, 1–5 (1979)
29. Alberghina, L., Vai, M., Vanoni, M.: Probing control mechanisms of cell cycle and ageing in budding yeast. Curr. Genomics **5**, 615–627 (2004)
30. Vanoni, M., Vai, M., Popolo, L., Alberghina, L.: Structural heterogeneity in populations of the budding yeast Saccharomyces cerevisiae. J. Bacteriol. **156**, 1282–1291 (1983)
31. Chen, K.C., Calzone, L., Csikasz-Nagy, A., Cross, F.R., Novak, B., Tyson, J.J.: Integrative analysis of cell cycle control in budding yeast. Mol. Biol. Cell **15**, 3841–3862 (2004)
32. Kaizu, K., Ghosh, S., Matsuoka, Y., Moriya, H., Shimizu-Yoshida, Y., Kitano, H.: A comprehensive molecular interaction map of the budding yeast cell cycle. Mol. Syst. Biol. **6**, 415 (2010)
33. Spiesser, T.W., Müller, C., Schreiber, G., Krantz, M., Klipp, E.: Size homeostasis can be intrinsic to growing cell populations and explained without size sensing or Signal. FEBS J. **279**, 4213–4230 (2012)

34. Brümmer, A., Salazar, C., Zinzalla, V., Alberghina, L., Höfer, T.: Mathematical modelling of DNA replication reveals a trade-off between coherence of origin activation and robustness against rereplication. PLoS Comput. Biol. **6**, e1000783 (2010)
35. Barberis, M., Linke, C., Adrover, M.À., González-Novo, A., Lehrach, H., Krobitsch, S., Posas, F., Klipp, E.: Sic1 plays a role in timing and oscillatory behaviour of B-type cyclins. Biotechnol. Adv. **30**, 108–130 (2012)
36. Swierstra, T., Vermeulen, N., Braeckman, J., van Driel, R.: Rethinking the life sciences. To better serve society, biomedical research has to regain its trust and get organized to tackle larger projects. EMBO Rep. **14**, 310–314 (2013)