Cell Growth Model - v0.9

Table of Contents

Model Overview

Features

Phenotypes
- Cell partition in 3 biomolecular species classes:
  - Metabolic
  - Ribosome-associated
  - Housekeeping
- Cell mass
  - Phenomenological model/Logistic function shaped
  - Variable mass dependent on intracellular energy levels
- Growth rate
  - Function of protein mass growth rate & cell mass size

Cell processes
- Translation models:
  - Initiation Complex formation
  - Elongation Complex maturation
  - Protein Synthesis
- Elongation complex maturation model:
  - Allows polysomes formation implicitly
  - Maturation dependent on cell energy levels
- Energy production & utilisation models
  - Simple one-step energy production model
  - Energy Utilised by Protein Synthesis & Elongation Complex Maturation
- Ribosome Assembly model
  - Ribosome formation from ribosome-associated proteins
  - No rRNA species involved
  - Allows for realistic numbers of ribosomes in cell

Version Modifications
- Derived from model v0.8
- Different threshold value for energy activation of ec_form & pro_syn process

Introduction

The cell growth model developed here describes gene expression patterns in a bacterial cell during steady state growth when the cell is supplied with various quality food supply nutrients. The developed model aim is to quantitatively describe the biomolecular cell composition and allocation of key cellular growth resources. It particularly focuses on the utilisation of key cellular resources utilised during the various processes of translation. The network of biomolecular species and reactions is modelled using ODE-based mass action & Michaelis-Menten reaction rate laws.

The model is composed of 14 biomolecular species as state variables representing proteins, mRNAs, the ribosome, translation-related complexes and an energy molecule (Figure 1 Panel A). The proteome is partitioned into 3 generalised protein molecules that includes metabolic proteins, ribosome-associated proteins, and housekeeping proteins. In relation to cell growth, the functionality of metabolic proteins is to convert extracellular nutrients to intracellular energy & biological building blocks that the cell utilises to build cell mass, while ribosome-associated proteins serve to form the ribosome complex that is responsible for protein synthesis. The housekeeping proteins are not assigned any particular functionality but are included in the modelled system as their presence as a separate proteome fraction has been demonstrated (Hwa et al., 2010) ad they seem to occupy a substantial proportion of the cell proteome mass. These proteins can be thought of as the proteome sector responsible for colonising/monitoring/responding to the variable extracellular environment or niche habitat but its effect to the cell growth is not direct. The cell model transcriptome is partitioned into mRNA molecules for each of the 3 protein classes. Additionally, the ribosome complex can associate with each of the mRNA molecules to form 3 types of translation initiation complexes (IC) for each of the protein classes. In turn, the ICs can transform into a translation elongation complex (EC) molecule of their respective class. Finally, the cell includes a generalised energy molecule that powers cell processes.

Results & Discussion

The model allows the quantitative description of the biomolecular composition of the cell at steady state as well as to predict the cell growth rate based on capturing the dynamics of the intracellular competition for three valuable growth facilitating resources (ribosomes, energy and proteome space). Some subprocesses in the model have been adopted from work by others (Weiße et al., 2015) that motivated this work, which we combine with new/modified subprocesses to develop a cell growth model with additional features to better address the needs of synthetic biology research and applications.

Ribosomes allocation though polysomes formation model

The ribosome is a key resource for cell growth as it is the cellular machinery that facilitates mass growth through the synthesis of protein molecules. To capture more accurately the allocation of this valuable cellular resource we developed a new description for translation composed of 3 process models: i) IC formation, ii) EC formation ad iii) Protein Synthesis model. The utilisation of ribosomes is determined by the kinetics of these 3 processes.

The IC formation model involves the association of the ribosome complex to an RBS/mRNA molecule. This is modelled by a reversible binding reaction, where the fluxes of ribosome capture are determined by the association/dissociation rate constants of the RBS/mRNA and the molecule abundance levels. Then, the EC formation model involves the irreversible

transformation of an IC molecule to an EC molecule. The reaction velocity is stimulated by the system's energy levels as the process of formation of fully assembled ribosome is known to require energy in the form of GTP. Concurrently with the formation of an EC molecule, the RBS molecule is regenerated and can be used to recruit more free ribosomes. The Protein Synthesis model involves utilisation of ribosomes in the form of an EC performing consecutive transpeptidation events for the production of a protein molecule. The process dynamics (reaction velocity or time to completion) depend on the amino acid size of each protein class and the system's energy levels at that moment in time. Formation of a protein molecule also results in the release of a ribosome back to the communal resource pool. The combined translation description models are shown diagrammatically in Figure 1B.

The developed description of translation allows for some unique properties for ribosome molecules utilisation. A simulation of translation system with a fixed value for system energy levels is shown in Figure 1C. This plot demonstrates that in the model the number of ECs is not limited by the number of rbs/mRNA molecules in the system. Previous modelling attempts (Macklin et al., 2020; Weiße et al., 2015) lacked a realistic representation of cellular counts for both ECs and mRNAs as these molecules exhibit greatly different molecular counts in the cell environment (BIONUMBERS IDs: 112795 & 101441). This is because of the formation of polysome structures where multiple ECs perform translation elongation concurrently on the same mRNA molecules. Here, the developed model facilitates the implicit emergence of polysomes structures. This is achieved by the release of the rbs/mRNA molecule back to the pool following the formation an EC molecule from a matured IC molecule. The result of this is the formation of multiple ECs molecules from a single mRNA molecule. Nevertheless, this reaction structure creates the potential for unregulated production of ECs that may exit the maximum capacity of that mRNA for ECs. In the physical world the number of ECs in a polysome structure is limited by the ribosome footprint (30 nts or 10 codons). This limit can be accounted in by modelling explicitly the ribosome flow on the mRNA molecules (Shaham & Tuller, 2018) but this methodology is computationally expensive. Instead, here the computationally cheap solution employed to limit excessive ECs formation is making the velocity of both EC formation and Protein Synthesis processes a function of the system energy levels. If the rate constant of EC maturation (kc) is the equal or smaller than the rate constant of transpeptidation, and both of these constants values are modulated by the system energy levels excessive EC formation beyond the capacity of the system (total available ribopositions) is avoided. This is demonstrated inn Figure panels D, E and F where changes in the system parameters (energy, ribosomes numbers and CDS length) never cause the system to be overflooded with EC beyond the available capacity of ribopositions. In addition, the modulating energy functions for the EC formation and Protein Synthesis processes allows to capture another aspect of ribosome utilisation inn the cellular environment where it is observed that in richer nutritional conditions the distance between ribosomes is shorter (Bremer et.al., 2008). For this, the half maximal threshold value for the energy modulation function can be assigned a larger value inn EC formation process rather than in the Protein Synthesis process which allows for a higher ratio of EC per mRNA molecule in the system (Figure 1G). This choice can be justified by the high energetic cost that the cell occurs for the production of Methionine amino acid that is required for the assembly of the elongation competent ribosome complex. In the complete cell growth model, formation of ribosome complex is the result of the ribosome assembly process in which protein molecules of the ribosome-associated class come together to form a free ribosome molecule.

Simple model for cell energy production

The next key cellular resource that is contested by cell processes is the available energy molecules. These energy molecules represent a generalised a coarse-grained metabolic building block (eg. ATP, amino acids, translation factors, tRNAs etc) required to drive the cell's biosynthetic processes (eg. Protein synthesis). The intracellular energy levels are determined by the competing processes of energy production against the energy consuming processes of protein synthesis and EC maturation. For energy production, our cell growth model uses a very simple reaction model that generates intracellular energy molecules from supplied nutrients. This simple reaction model was preferred as opposed to more fine-grained models as synthetic biology applications often deal with cell processes of biosynthesis rather than processes involved in metabolism and also enables us to keep the model complexity low. The energy production reaction model is described using a Michaelis-Menten rate law whose maximum reaction velocity depends on the quantity & quality of the extracellular nutritional environment and on the abundance of protein molecules of the metabolic class (can represent membrane transporters, glycolysis enzymes, TCA enzymes, or whatever is the limiting step in the metabolism). The energy consumption in the cell model is the result of the sum of energy expenditure from EC formation & Protein Synthesis processes. Figure 1H shows the result of simulation of an simplified system that integrates process models for energy production & utilisation at various extracellular amounts of nutrients. The model exhibits the well-known Monod relationship between the concentration of a limiting nutrient and the observed mass growth rate. Also from the figure we can observed that the consumption rate of energy by the EC formation process is very small compared to the consumption rate of the Protein Synthesis process, and that the biomass accumulation rate very closely tracks the energy production rate. In the complete cell model, the 3 cellular coarse grain classes have to compete for the available/limiting energy molecules in the common pool.

Variable cell size/mass model as a function of produced energy

The last contested cellular resource in the cell growth model is implemented as a competition between protein classes for the finite space of cell mass (=cell size defined as total number of amino acids per cell). In the literature, early experimental observations had informed the formulation of a hypothesis that cell size is coordinated with growth rate independent of the chemical composition of the growth medium, with the cell size to increase exponential with respect to the nutrient imposed growth rate (Neidhardt, 1999). Nevertheless, newer experimental observations support the idea that the cell size is modulated independently from growth rate according as a function of the richness of biochemical building blocks in the cell (Büke et al., 2022). Cell size determination is important for the cell growth model as it determines the phenotype of cell growth rate through its relative relationship to the rate of biosynthesis (synthesis of proteomic mass). Since the cell's total mass can vary significantly (~4-fold) over commonly used experimental conditions, to describe cellular composition and growth rates more accurately, we make use of a variable cell size model for our system. As investigations are ongoing for a mechanistic understanding of cell size regulation, we opted for the use of a phenomenological logistic function for the cell size process that is a function of the intracellular energy levels (Figure 1H). The logistic function can approximately linearly track experimentally derived cell size data when their values fall around the function's midpoint. At the same time, the logistic function has the desirable property of being bounded that helps with simulations during parameter fitting. In the complete cell growth model, the 3 cellular coarse grain classes must compete for the available proteome space, and the

reaction dynamics of their gene expression processes in combination with the effect of the cell division process result in different phenotypes of proteome partition for the bacterial cell.

Integrated cell model & properties

In addition to the aforementioned subcellular processes, our cell model incorporates a transcription model that produces an mRNAs/RBSs as a function of the intracellular energy levels, a degradation model that removes mRNAs/RBSs from the system by a first order kinetics reaction process, and a process for ribosome complex assembly from protein molecules of the ribosome-associated class. In addition, the model is able to derive the instantaneous cell growth rate from a phenomenological model that normalises the total protein synthesis in the system by the derived value of the cell size. The integration of all these described submodels and their application to each model functional class result in a cell growth model that describes a bacterial cell's proteome partition, its cell size(mass) and cell growth rate from genome encoded gene expression dynamics.
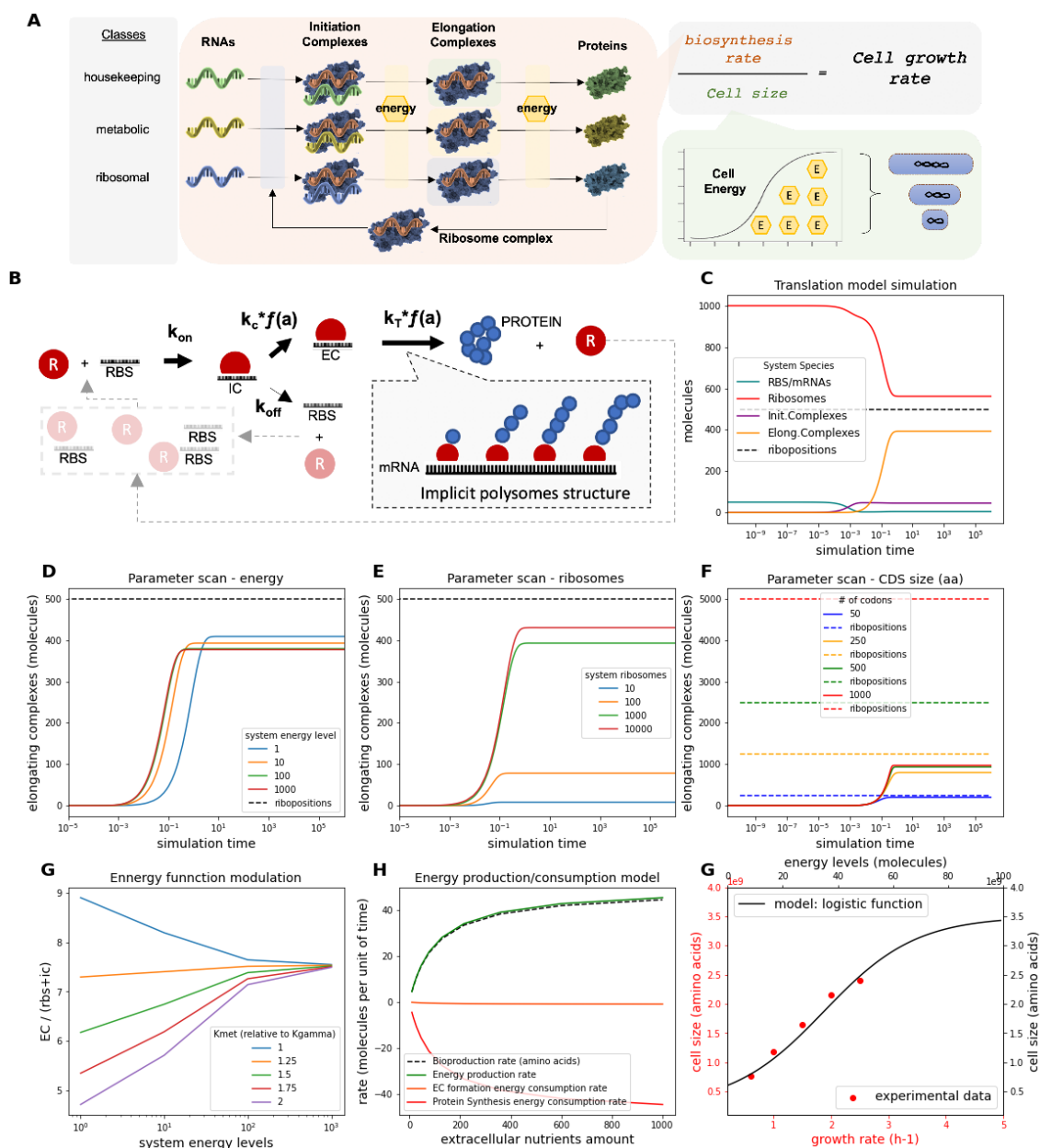

Figure 1

Methodology

Cell Processes

Transcription

Assumptions
- Transcription is an energy dependent process, but it is not made to consume any energy as its consumption rate is small compared to the translation processes (Weiße et al., 2015)   - The threshold value for half-maximal activation of transcription occurs at higher intracellular energy levels for ribosomal protein gene expression, as it has been demonstrated that intracellular effector molecules inhibit transcription of ribosome-associated genes at poor intracellular metabolic environment (Barker et al., 2001; Lemke et al., 2011).

Reaction	
Forward rate Reverse rate	wx*mod_fcn_a $\rightarrow$ m_x
Energy modulating function	$$mod\_fnc\_a = \left(\frac{a}{theta + a}\right)$$ theta for hsk & met = thetax theta for rib = thetar

Species	
m_x	RBS of class x = {rib, met, hsk}
a	Intracellular energy molecule

Parameters	Description	Value	Units
wx	transcription rate		molecules/minutes

Modifications to basic Model

Assumptions
- For hsk class: - For rbs/mRNA of housekeeping class, the transcription rate is inhibited by a modulation function based on the number of hsk proteins in the cell, as previous research suggests the existence of an autoregulation mechanism so that their relative proteome abundance is maintained across different growth conditions (Hwa et al., 2010)

Modulating function (hsk class)	$$mod\_fnc\_tx\_hsk = \left(\frac{1}{1 + \left(\frac{p\_hsk}{Kq}\right)^{nq}}\right)$$

Parameters/Species	Description	Value	Units
p_hsk	housekeeping proteins		molecules
Kq	repression threshold		molecules
nq	hill coefficient		-

Fluxes
d(m_rib)/dt   = +wr*mod_fcn_a
d(m_met)/dt   = +wr* mod_fcn_a* mod_fnc_txhsk
d(m_hsk)/dt   = +wr*mod_fcn_a

mRNA degradation

Assumptions			
- Only the free form of the mRNA is degraded (RBS) - Degradation rate is the same for all mRNA classes			
Reaction			
Forward rate	$$m\_x \xrightarrow{dm} null$$		
Reverse rate			
Species			
m_x	RBS of class x = {rib, met, hsk}		
Parameters	Description	Value	Units
dm	mRNA degradation rate		1/minute
Fluxes			
d(m_rib)/dt = -dm* m_rib d(m_met)/dt = -dm* m_met d(m_hsk)/dt = -dm* m_hsk			

Translation Initiation Complex formation

Assumptions
• The association rate constant between ribosomes and rbs_mRNAs is variable, while all classes exhibit the same dissociation kinetics. This simplification does not represent reality as experimental observation show both constant values can vary within an interval of 3 orders of magnitude (Gualerzi & Pon, 2015) but its implementation was necessary to reduce the search space and the number of dimensions in parameter fitting.

Reaction	
Forward rate	$$m\_x + ribo \xrightleftharpoons[ku]{kb\_x} ic\_x$$
Reverse rate	

Species	
m_x	RBS of class x = {rib, met, hsk}
ribo	Ribosomes
ic_x	Translation Initiation complex of class x = {rib, met, hsk}

Parameters	Description	Value	Units
kb_x	RBS/ribosome association rate		1/(molecules*minutes)
ku	Initiation Complex dissociation rate		1/minute

Modifications to basic Model

Assumptions
• For rib class: ○ The association rate of rbs_mRNA with ribosome for the formation of ic_rib is inhibited by a modulating function based on the amount of free ribosomal protein molecules in the cell, as negative feedback regulation at translation initiation stage is well established for the control of ribosomal protein expression (Nomura et al., 1980)

Modulating function (rib class)	$$mod\_fnc = \left(\frac{1}{Krepr + p\_rib} \right)$$

Parameters/Species	Description	Value	Units
p_rib	ribosomal proteins		molecules
Krepr	repression threshold		molecules

Fluxes

d(ic_rib)/dt   = +kb_rib *ribo *r_rib * (1 / (1 + (*p_rib* / Krepr)))
               -ku*ic_rib
 	
d(ic_met)/dt   = +kb_met *ribo * r_cat
                -ku*ic_cat
 	
d(ic_hsk)/dt   = +kb_others*ribo *r_hsk
                -ku*ic_hsk

Translation Elongation Complex formation

Assumptions
- The maturation rate of IC to EC is energy dependent with an activating modulating function that depends on the cell energy levels. The modulating function half-activation threshold is set at the same value as that of translation elongation process. This parameter value can be independently modified but only towards a higher value to prevent ribosome overcrowding on the mRNA molecule.   1 energy molecule is consumed per maturation event a GTP molecule is utilised by IF2 in the pathway for 70SIC formation (Gualerzi & Pon, 2015).

Reaction	
Forward rate	$kc * mod\_fcn\_gamma$ $ic\_x + a \rightarrow ec\_x + m\_x$
Reverse rate	
Energy modulating function	$$mod\_fnc\_gamma = \left(\frac{a}{Kgamma + a} \right)$$

Species	
ic_x	initiation complex of class x = {rib, met, hsk}
ec_x	elongation complex of class x = {rib, met, hsk}
m_x	RBS of class x = {rib, met, hsk}
a	Intracellular energy molecule

Parameters	Description	Value	Units
kc	rate constant of IC maturation		1/minutes
Kgamma	Threshold of half-maximal transpeptidation rate (based on energy levels)		molecules

Fluxes

d(ic_rib)/dt   = -ic_rib * kc * mod_fcn_gamma
 	
 d(ic_met)/dt   = -ic_met * kc * mod_fcn_gamma
 	
 d(ic_hsk)/dt   = -ic_hsk * kc * mod_fcn_gamma
 	
 d(a)/dt   = -ic_rib   * kc * mod_fcn_gamma
                  -ic_met * kc *mod_fcn_gamma
                  -ic_hsk * kc *mod_fcn_gamma
 	
 d(ec_rib)/dt   = +ic_rib * kc * mod_fcn_gamma
 	
 d(ec_met)/dt   = +ic_met * kc *mod_fcn_gamma

```
d(ec_hsk)/dt   = +ic_hsk * kc * mod_fcn_gamma


d(m_rib)/dt    = +ic_rib * kc *mod_fcn_gamma

d(m_met)/dt   = +ic_met * kc * mod_fcn_gamma

d(m_hsk)/dt   = +ic_hsk * kc *mod_fcn_gamma
```

Translation Protein Synthesis

Assumptions
• Simplified translation elongation model developed in (Weiße et al., 2015). In short, the net rate is dependent on the amino acid length of the protein molecule and is controlled by a modulating function that is activated by increased cellular energy levels.

Reaction	
Forward rate Reverse rate	$ec\_x + a*lenX \quad \rightarrow \quad \overset{gmax * mod\_fcn\_gamma}{p\_x + ribo}$
Energy modulating function	$mod\_fnc\_gamma = \left(\dfrac{a}{Kgamma + a}\right)$

Species	
ec_x	elongation complex of class x = {rib, met, hsk}
a	Intracellular energy molecule
p_x	protein of class x = {rib, met, hsk}
ribo	ribosome complex

Parameters	Description	Value	Units
lenX {lenO,lenR,lenC}	Protein length {p_hsk, p_rib, p_met}		amino acids
Kgamma	Threshold of half-maximal transpeptidation rate (based on energy levels)		molecules

Fluxes
d(ec_rib)/dt = -ec_rib * gmax / lenR *mod_fcn_gamma
d(ec_met)/dt = -ec_met * gmax / lenC *mod_fcn_gamma
d(ec_hsk)/dt = -ec_hsk * gmax / lenO *mod_fcn_gamma
d(a)/dt = -ec_rib * gmax *mod_fcn_gamma -ec_met * gmax *mod_fcn_gamma -ec_hsk * gmax *mod_fcn_gamma
d(p_rib)/dt = + ec_rib * gmax /lenR *mod_fcn_gamma
d(p_met)/dt = + ec_met * gmax /lenC *mod_fcn_gamma
d(p_hsk)/dt = +ec_hsk * gmax /lenO *mod_fcn_gamma

$$d(ribo/dt) = d(p\_rib)/dt + d(p\_met)/dt + d(p\_hsk)/dt$$

Ribosome Assembly

Assumptions
• Ribosomal proteins instantaneously form sets of ribosomal proteins of size equal to the total amino acid mass of an assembled ribosome • Ribosome Assembly is a single association event where a ribosomal proteins set forms an assembled ribosome

Reaction	
Forward rate	c
Reverse rate	

Species	
p_rib	Free ribosomal proteins
p_rib_set	p_rib_set = ribosomal proteins / (lenRibo/lenR)
ribo	Ribosomes

Parameters	Description	Value	Units
lenRibo	Size of ribosome with all its ribosomal proteins		aa
lenR	Size of ribosomal protein		aa
k_form	Rate constant for ribosome assembly		molecules/minute

Fluxes
d(rib)/dt = - k_form * [p_rib_set] * (lenRibo/lenR) d(Ribo)/dt = + k_form * [p_rib_set]

Energy production

Assumptions
- Energy production is limited by a single bottleneck enzyme in the metabolic pathway and is modelled based on Michaelis-Menten reaction rate model   - Efficiency of energy generation is dictate by the quality of the media (ns). The ns variable linearly scales the effectiveness by which energy molecules are produced by the metabolic sector. This simplification has also been used by others (Weiße et al., 2015) and is necessary as the true relationship between media composition/quality and energy molecules production would be very hard to calculate   - The production rate flux is determine by protein molecules that are able to carry out a series of chemical reactions that facilitate the conversion of an extracellular nutrient molecule to an intracellular energy molecule. Thus, the Vmax, Km, and ns values are not meant to represent any "real" values but adjustable quantities that facilitates the devotion of amino acid mass to the metabolic proteome sector for the needs of resources allocations

Reaction	
Forward rate	ns* ((Vmax *s0)/(Km + s0)) p_ met → a + p_met
Reverse rate	

Species	
a	Intracellular energy
p_met	Proteins metabolic class

Parameters	Description	Value	Units
ns	Nutrient quality		No units
Vmax	Catalytic rate of nutrient utilisation		1/minute
s0	Extracellular concentration of nutrients		molecules
Km	Half-maximal threshold of nutrient levels utilisation		molecules

Fluxes
d(a)/dt   = +p_met * ns* ((Vmax *s0)/(Km + s0))

Energy consumption

Assumptions
• One energy molecule is consumed by each Elongation Complex maturation event • One energy molecule is consumed by each transpeptidation event in protein synthesis, where the consumption of each protein synthesis event is determined by the aa length of the synthesized protein

Process 1	
Forward rate Reverse rate	kc * mod_fcn_gamma ic_x + a → ec_x + m_x
Energy modulating function	$$\text{mod\_fnc\_gamma} = \left(\frac{a}{Kgamma + a}\right)$$

Species	
ic_x	initiation complex of class x = {rib, met, hsk}
ec_x	elongation complex of class x = {rib, met, hsk}
m_x	RBS of class x = {rib, met, hsk}
a	Intracellular energy molecule

Parameters	Description	Value	Units
kc	rate constant of EC formation from IC		1/minute
Kgamma	transpeptidation elongation rate threshold for translation based on energy levels		molecules

Process 2	
Forward rate Reverse rate	gmax * mod_fcn_gamma ec_x + a*lenX → p_x
Energy modulating function	$$\text{mod\_fnc\_gamma} = \left(\frac{a}{Kgamma + a}\right)$$

Species	
ec_x	elongation complex of class x = {rib, met, hsk}
a	Intracellular energy mollecule
p_x	protein of class x = {rib, met, hsk}

Parameters	Description	Value	Units
lenX {lenO,lenR,lenC}	Protein length {p_hsk, p_rib, p_met}		1/minute
Kgamma	Threshold of half-maximal transpeptidation rate (based on energy levels)		molecules
Fluxes			

tinitrate   = - (ic_rib + ic_met + ic_hsk) * kc * mod_fcn_gamma

ttrate   = - (ec_rib + ec_met + ec_hsk) * gmax * mod_fcn_gamma

d(a)/dt   = - tinitrate - ttrate

Phenotypes
Cell size

Assumptions
- Phenomenological model based on a logistic function   - Cell size is increases with increasing availability of intracellular energy (metabolic precursors)   - Cell size is equal to an intracellular amino acid count, while other biomolecules are not considered (ex. Lipids, carbohydrates, DNA, RNA etc). This is because the model here aims to describe the competition between various cellular processes for proteome space and protein based biomolecules.   - facilitates the dilution of intracellular species and redistribution of cellular content according to the reaction fluxes.

Phenotype Equation	
cell size	minimal_mass + (max_inf / (1 + exp(- inlation_gradient * (a – mid_inf))))

Species	
a	Intracellular energy molecule

Parameters	Description	Value	Units
min_mass	Minimal cell mass that is allowed for the cell to assume even in the absence of intracellular energy		amino acids
max_inf	Maximum inflation of cell size upon high levels of cell energy		amino acids
mid_inf	Energy levels inflection point of half maximal cell size inflation		molecules

Growth rate

Assumptions
• Growth of the system is defined as the number of transpeptidation (amino acid incorporation in protein molecules) events per unit of time. • Normalised to 1 cell unit set by cell mass variable. • facilitates the dilution of intracellular species and redistribution of cellular content according to the reaction fluxes.

Phenotype Equation	
Cell growth rate	lam = ttrate / cell size where: ttrate = (ec_rib + ec_met + ec_hsk) * gmax * mod_fcn_gamma

Species	
ec_x	Elongation complexes of class x = {rib, met, hsk}

Parameters	Description	Value	Units
Kgamma	Threshold of half-maximal transpeptidation rate (based on energy levels)		molecules