StoCellAtor

A stochastic whole cell simulator that uses TASEP (explain) to model heterologous protein expression and the resulting burden (explain) on the bacterial cell

Synthetic biology

- Synthetic biology
  - Input: inducible promoters, e.g. Pbad with Arabinose, Ptet with aTc
  - Output: for example, fluorescent proteins, e.g. GFP, YFP
  - ALU: promoters can be activated/repressed by produced proteins, this gives rise to different combinations of proteins produced depending on input (logic gates): Tasmir et al, Nature 2011; Wang et al, Nature Comms 2011; Siuti et al, Nature Biotech 2012
  - Memory: Permanent genetic memory using orthogonal recombinases: Yang et al, Nature 2014
  - CPU, oscillator: tunable oscillator: Stricker et al, Nature 2008; relaxation oscillator synchronized by intercell signalling: McMillen et al, PNAS 2002; repressilators coupled by quorum sensing: Garcia-Ojalvo et al, PNAS 2004; synchronized quorum of genetic clocks: Danino et al, Nature 2010
- We have all the components to build a cellular computer

Synthetic biology 2

- Other uses:
- Creative new medicines:
  - Invasion of cancer cells by engineered bacteria: Anderson et al, J. Mol. Biol. 2006
  - Synchronized cycles of bacterial lysis (to limit growth) with in-vivo delivery (of medicine e.g. cytotoxic agents against tumours): Din et al, Nature, 2016
- Detection:
  - Programmable probiotics for detection of cancer in urine: Danino et al, Science Trans. Med. 2015
  - Biosensors for arsenic detection: Merulla et al Curr. Opin. Biotechnol. 2013
  - Biosensor for schistosome cercariae to reduce the spread of schistosomiasis: Webb et al, Sci.Rep. 2015

Design Problem

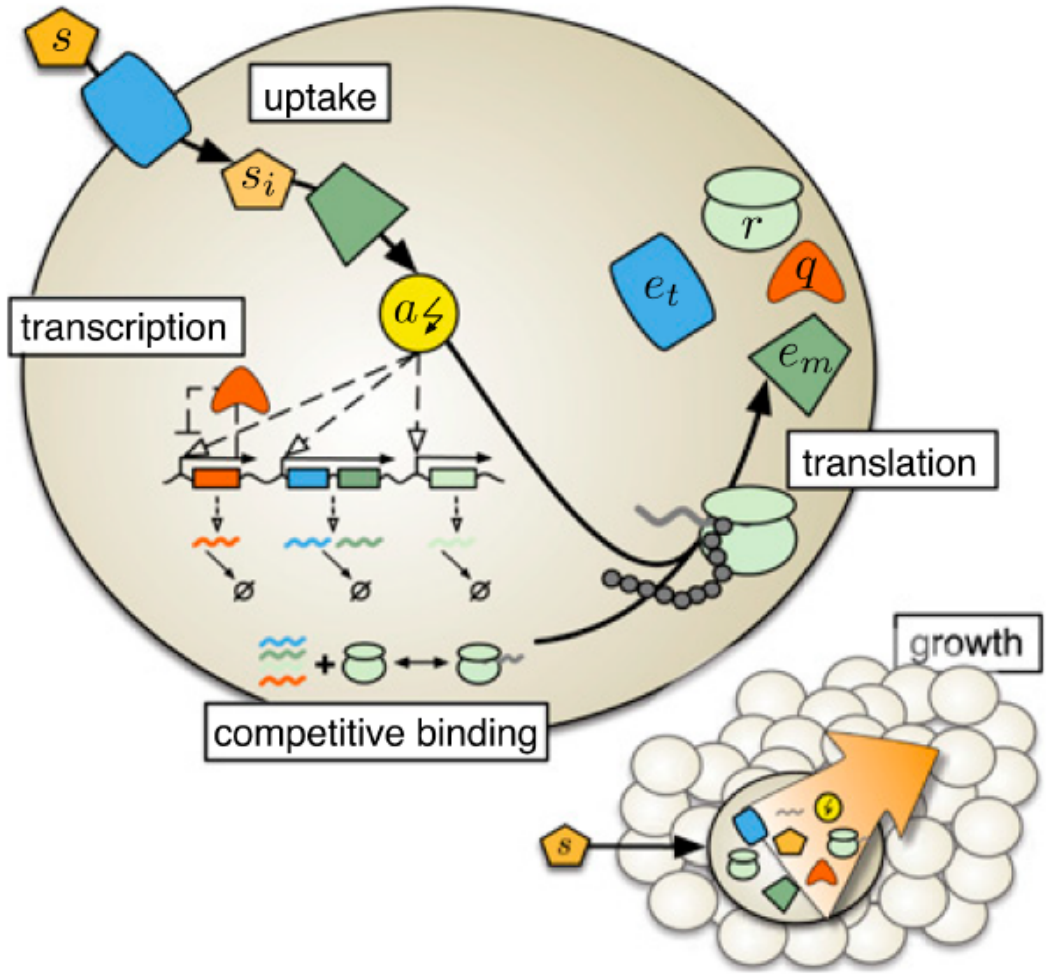
- Elowitz and Liebler (Nature, 2000) came up with a synthetic oscillatory network of transcriptional regulators.
- It worked in theory, the oscillations were not regular and dies out in the experimental results
- Weisse et al (PNAS, 2015) managed to explain why theory does not correspond to practice: isolated repressilator model ignores resource trade-offs:
  - Finite proteome
  - Finite energy
  - Finite ribosomes

Problem in general

- Production of heterologous protein takes away resources from the innate proteins of the cell (e.g. ribosomal proteins), hence cell growth slows down -> this is referred to as *burden* on the bacterial cell
- As a result slower growth might decrease the heterologous protein yield, even though heterologous protein production rate *per cell* is bigger
- Modelling is needed to understand the amount of burden and give estimates for heterologous protein yield

Weisse model (Figure: Weisse et al., PNAS, 2015)

- Proteome partition
  - Transporter/metabolic enzymes (E)
  - Ribosomal proteins (R)
  - Housekeeping proteins (Q)
  - Heterologous proteins (H)
- Both translation and transcription rates are Hill functions of energy
- Energy is incremented by parameter n_s each time an internalized nutrient is metabolized
- Energy is used by translation
- Polyribosomes are not considered
  - Unable to model ribosomal queues and ribosome sequestration
  - Not realistic



Translation models

- We want model that consider
  - Polyribosomes
  - Translation at the codon level to be able to model slow codons that cause ribosomal queues that reduce growth rate due to the reduction in the pool of actively translating ribosomes
- We have three options
  - TASEP: totally asymmetric exclusion process as known in Physics: stochastic model, exclusion refers to the fact that only one ribosome can be at a particular place at a given time
  - RFM: the mean-field approximation of TASEP: essentially models a 1D constrained random walk of ribosomes on the mRNA; set of ODEs
  - Joaquin's model: deterministic approach: $\delta = v_{el} * \Delta t$

Why we chose TASEP

- RFM: set of ODEs describing the movement of ribosomes on the mRNA, but at the same time, in a whole-cell setting the number of ribosomes and the number of transcripts are changing -> not clear how to extend RFM to a whole-cell model
- Deterministic approach: yields constant ribosome density along the mRNA, whereas TASEP/RFM yields a monotonically decreasing ribosome density along the mRNA, which corresponds to experimental results (Ingolia et al, Science 2009)
- TASEP: can be extended to Whole-Cell modelling framework, but then everything needs to be stochastic

Our contribution

- Transform Weisse et al's deterministic model to a stochastic model so that it can be used with TASEP
- Made TASEP algorithm faster by recursively changing the transition matrix that only considers transitions to *non-restricted states* (explain)
- Hence we created a stochastic whole cell model that considers heterologous protein production with arbitrary transcript length, promoter and RBS strength and individual codon elongation rates
- It tells gives us the average instantaneous growth rate and the mass fraction of different types of proteins in the proteome

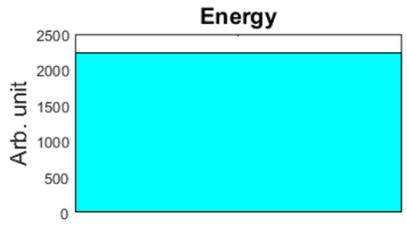
Results – changing nutrient quality

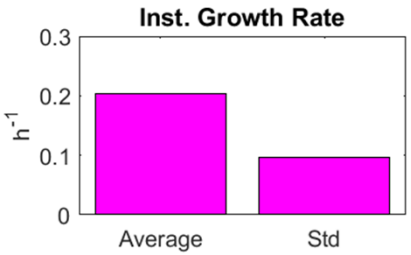
- Internal energy (a.u.) is incremented by parameter n_s (nutrient quality) if the internal nutrient is metabolized
- The model reproduces Monod's growth law (Monod, Ann. Rev. Mic. Biol., 1949): growth rate increases with increased nutrient quality until saturation

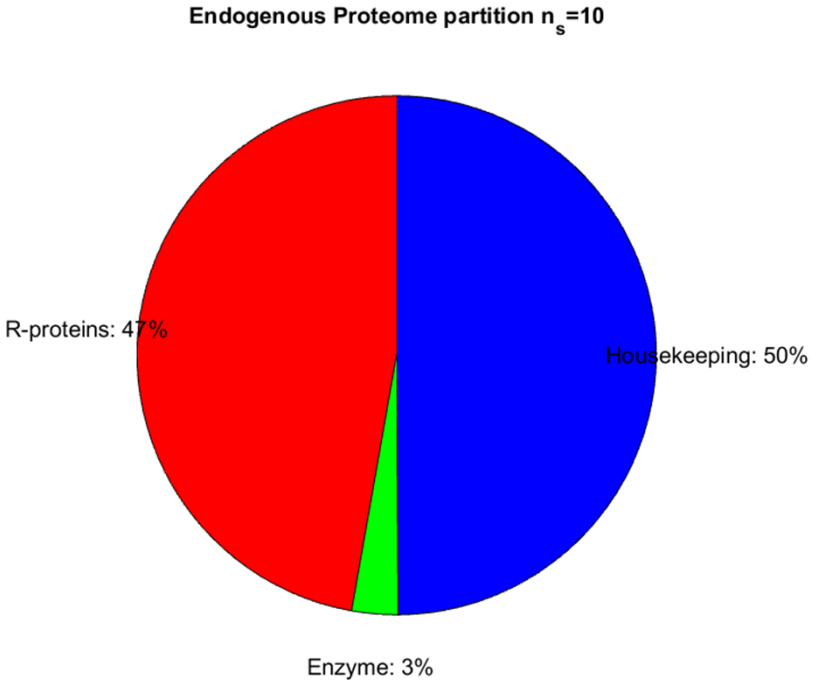
n_s	10	50	100	200	400	500	600
Average Inst. Growth Rate	0.20	0.88	1.70	2.73	4.34	4.77	4.81

- Growth rate saturates around $n_s = 500$
- By changing only n_s, we can achieve experimentally valid growth rates ($2\ h^{-1}$)

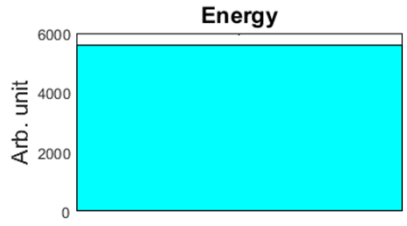
Proteome partition as a function of ns

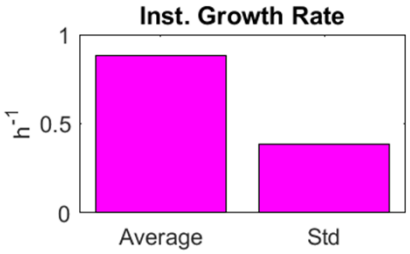


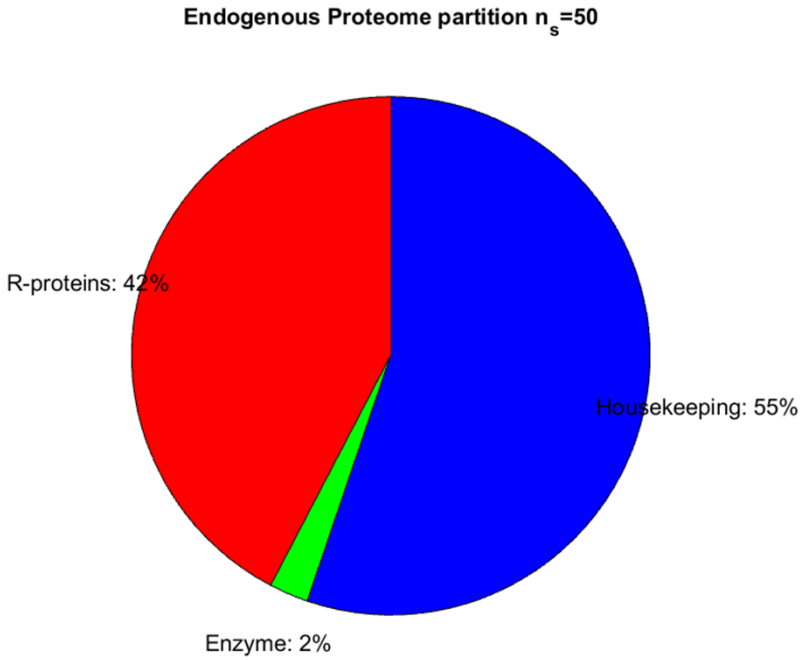




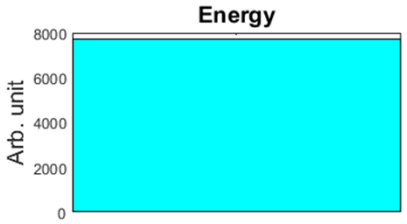
Proteome partition as a function of ns

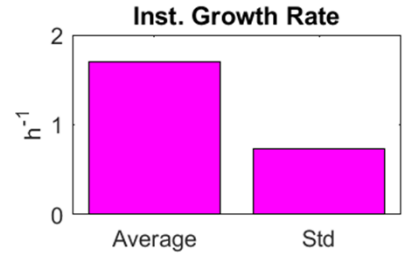


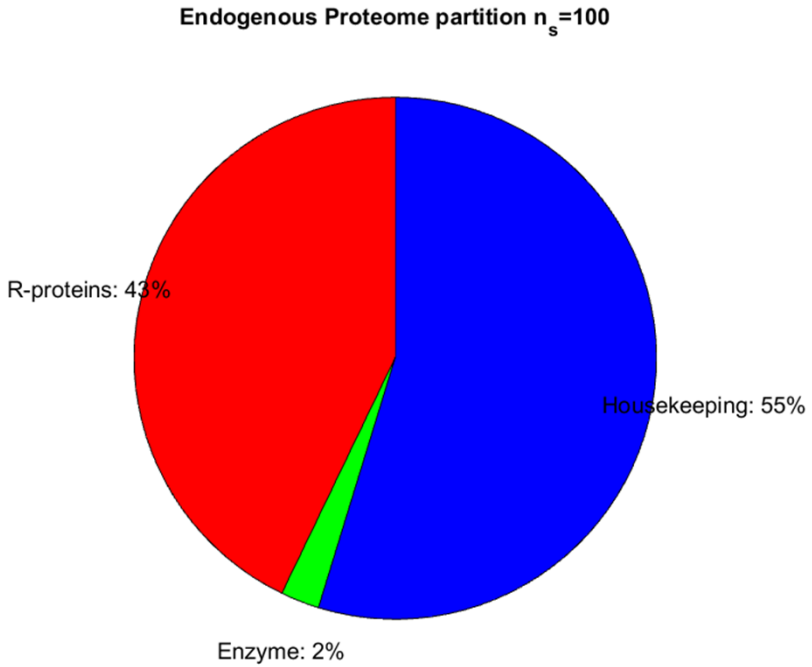




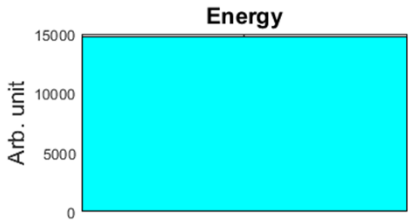
Proteome partition as a function of ns

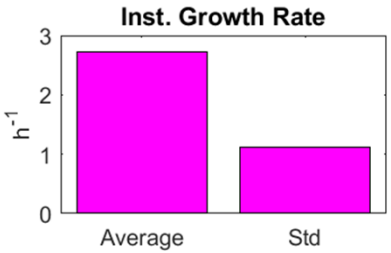


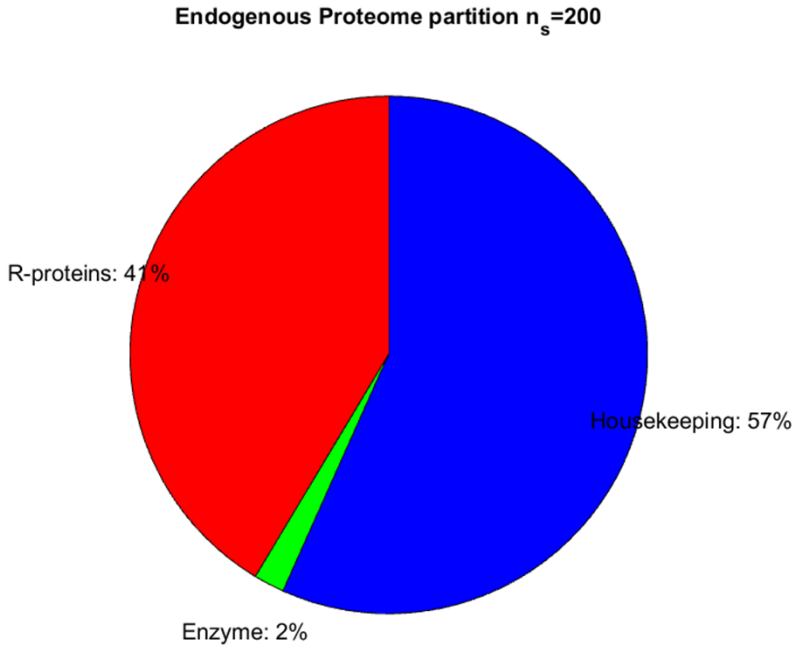




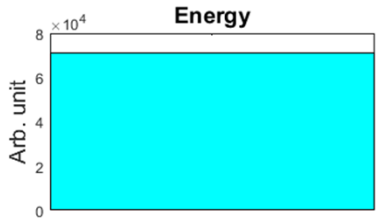
Proteome partition as a function of ns

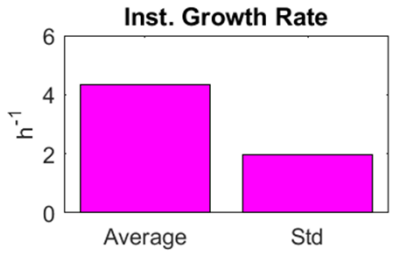


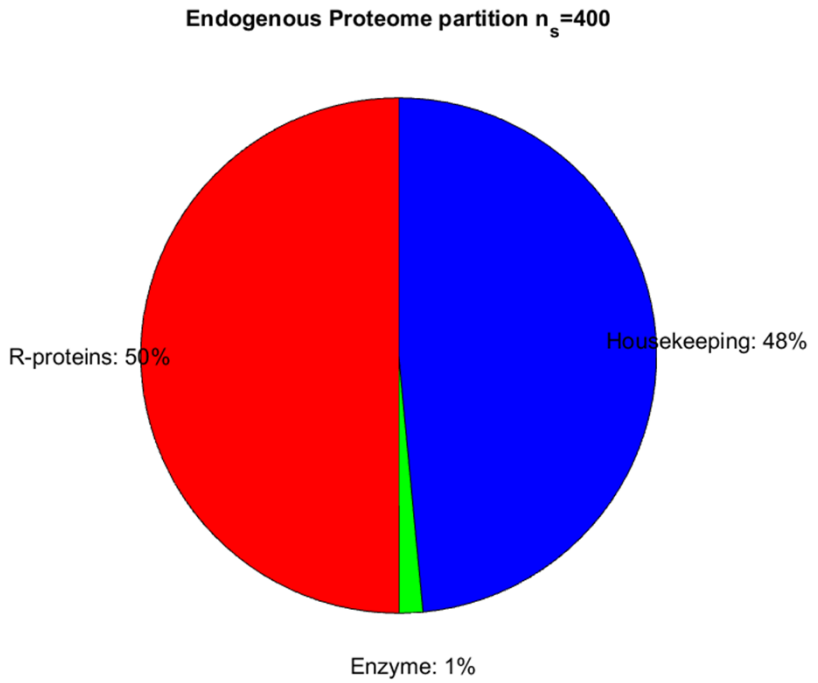




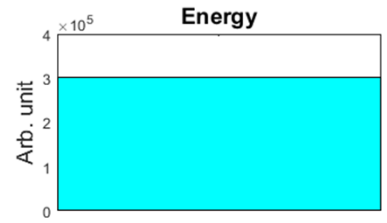
Proteome partition as a function of ns

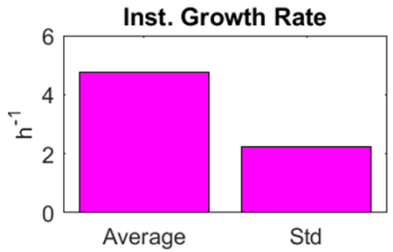


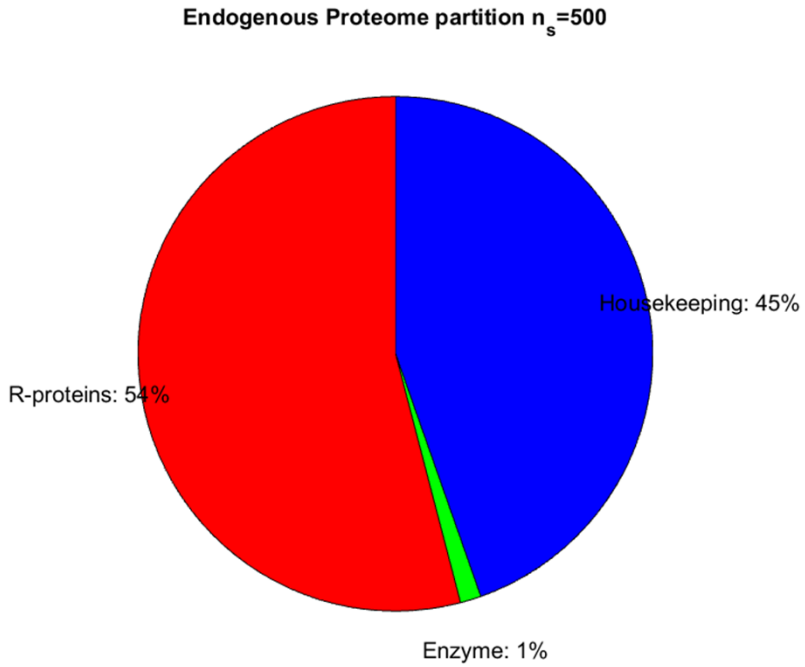




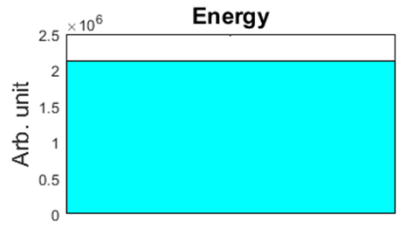
Proteome partition as a function of ns

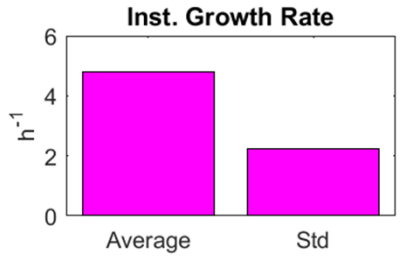


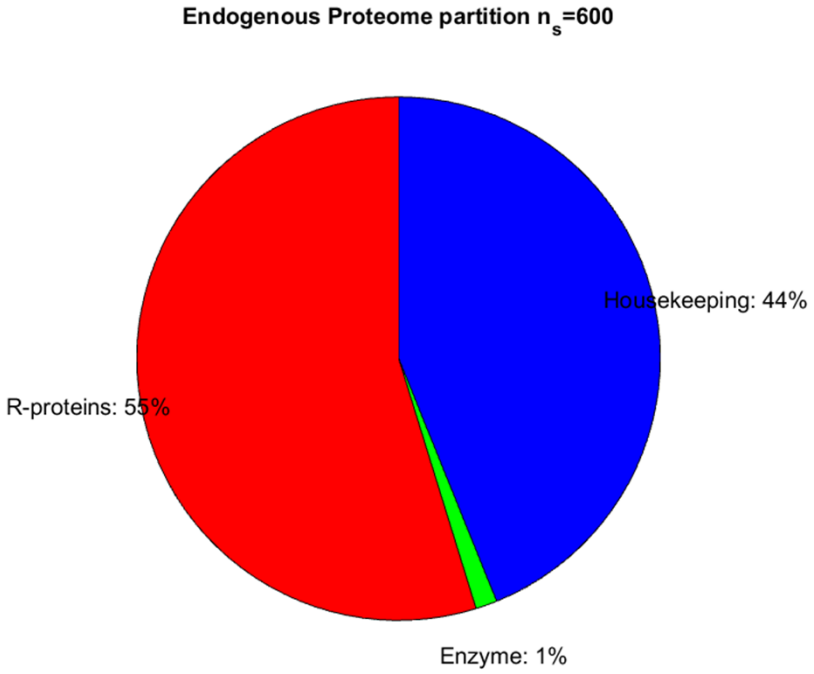




Proteome partition as a function of ns





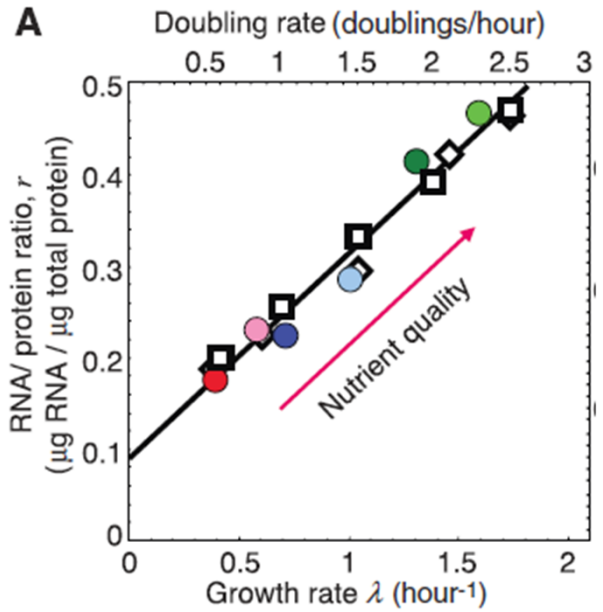


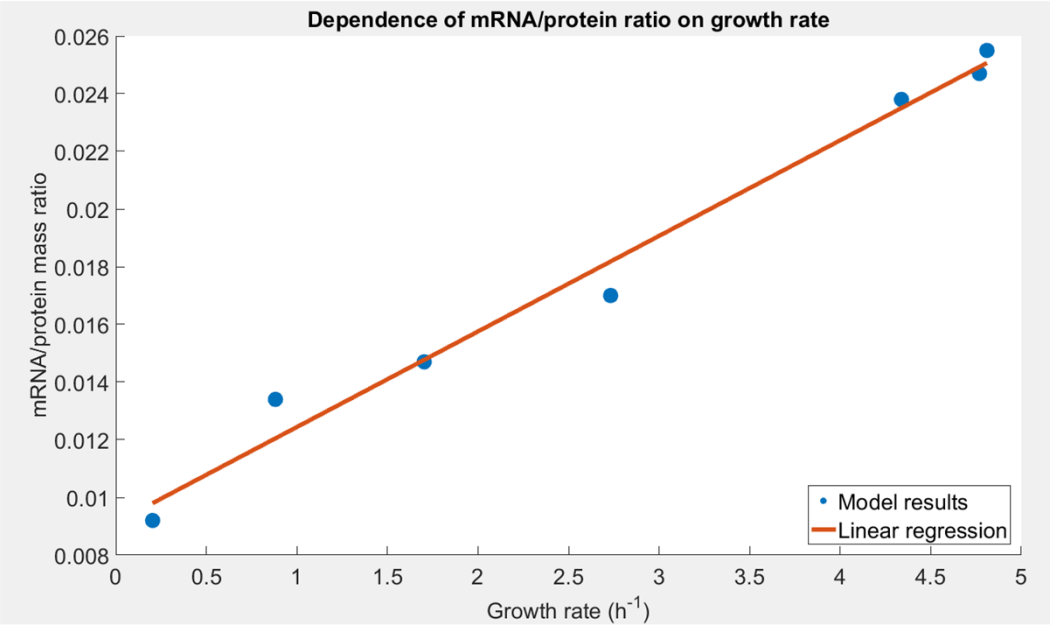
Observations - changing nutrient quality

- With nutrient quality proteome fraction of enzymes go down
- At the baseline (n_S = 10) ribosomal protein fraction (47%) is quite in balance with housekeeping fraction (50%)
- Then as nutrient quality increases housekeeping becomes more dominant (57% vs. 41%) up to n_S = 200; this is because even though the percentage increase is smaller on the Hill curve for housekeeping (closer to half-saturating constant), the maximum rate of transcription for housekeeping (850) is much bigger than the maximum rate for ribosomal proteins (27)
- But the Hill curve for housekeeping starts saturating earlier, so at large ns, R-protein fraction increases and takes over housekeeping protein fraction (55% vs. 44%)

So wait..where is the linear relationship?

- Scott et al. (Science, 2010) showed that RNA/protein ratio increases roughly linearly with growth rate (left, adopted from Scott et al.)
- We reproduce this result with our model! (right)

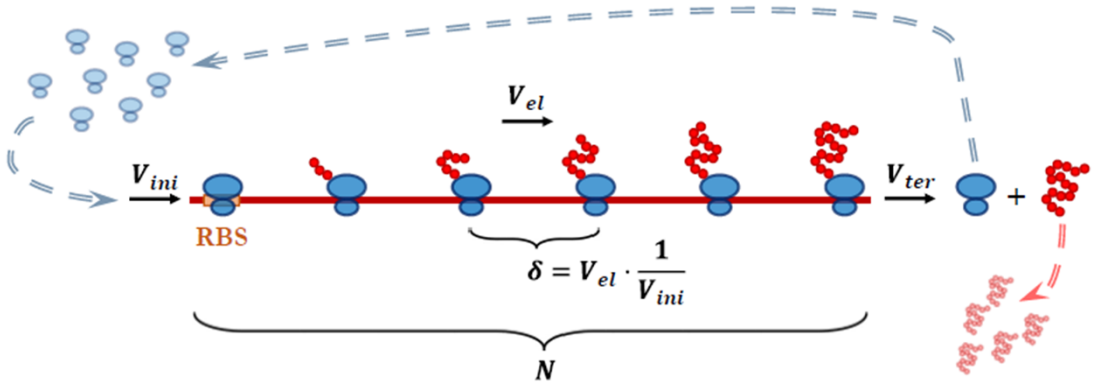




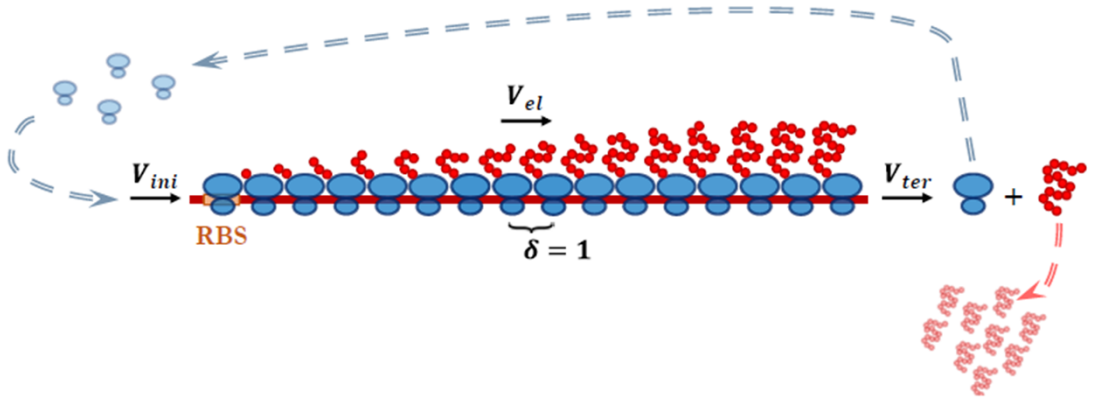
Introduction of heterologous proteins

- In general: higher the heterologous protein production rate, the lower growth rate is, since translational resources are taken away from endogenous proteins. This is referred to as burden!
- We ran simulations with changing the maximal transcription rate (parameter a from 0.1 to 10) and the initiation rate constant (parameter b from 0.1 to 10) for the heterologous protein
- We did this for three cases:
  - No slow codons
  - 10^{th} ribosome site has slow codons (out of 30)
  - 21^{st} ribosome site has slow codons (out of 30)
- With two relative slow codon rates: 0.03 and 0.005

Introduction of heterologous protein with slow codon (graph from thesis of Joaquin)



(a) *Low translation initiation rate:* $V_{ini} < \nu_{el}$



(b) *High translation initiation rate:* $V_{ini} \geq \nu_{el}$

Slow codon limiting translation phenomenon

- The higher the initiation rate compared to the limiting elongation rate, the more likely it is that queues form
- Queues are very inefficient: Without gaining bigger production rate for heterologous proteins, growth rate is decreases since ribosomes in the queues are not actively translating
- Hence better practice to make promoter of heterologous transcript stronger if higher heterologous production rate is required
- It is worse to have slow codons at the 3’ end than at the 5’ end because it allows for longer queue formation
- All these rules above are reflected in the simulation results!

StoCellAtor gives insight into the cell

- Not only do the results correspond clearly to the expectations, StoCellAtor gives insight into transient and stationary proteome partition, internal energy and growth rate of the cell
- It also outputs other state variables of the cell, for example
  - Number of internalized nutrients
  - Transcript levels
  - Which ribosome sites are occupied by the ribosomes
- Let's watch how the cell changes when a heterologous transcript with 10^{th} slow ribosome site (0.005 rel. elongation rate) is introduced (a=1 and b=1): https://www.youtube.com/watch?v=EPMutCumilk&feature=youtu.be

Video observations

- Growth rate decreases
- Energy increases: we simulated a case with slow codon limiting translation phenomenon: ribosomes are queuing, they cannot translate, energy usage is reduced -> energy grows
- Innate proteome is “squeezed” by the heterologous protein in the pie chart
- Sanity check: Simulation clearly converges

Summary of model justification

- Reproduces Monod's growth law
- Reproduces linear relationship discovered by Scott el al.
- Gives experimentally valid growth rate without parameter tuning
  - (doubling time = 20 mins corresponding to 2.4 h^{-1} growth rate); makes sense, since we used values from bionumbers and some arbitrary values for less important and biologically unknown parameters -> robust model
- Effects of burden are clearly visible in the results
- Effects of ribosome queues are clearly observed in the heterologous case with slow codons