Translation

Model implementation. Translation is the process by which the coding sequences of mRNA transcripts are translated by 70S ribosomes into polypeptides that then fold into proteins. This process accounts for more than two thirds of an *E. coli* cell's ATP consumption during rapid growth [3] and the majority of macromolecular mass accumulation. In the *E. coli* model translation occurs through the action of two processes in the model: `PolypeptideInitiation` and `PolypeptideElongation`.

`PolypeptideInitiation` models the complementation of 30S and 50S ribosomal subunits into 70S ribosomes on mRNA transcripts. Full 70S ribosomes are formed on mRNA transcripts by sampling a multinomial distribution with probabilistic weights calculated from the abundance of mRNA transcripts, and each transcript's translational efficiency (See Algorithm 1). Translational efficiencies were calculated from ribosomal profiling data [2].

`PolypeptideElongation` models the polymerization of amino acids into polypeptides by ribosomes using an mRNA transcript as a template, and the termination of elongation once a ribosome has reached the end of an mRNA transcript. This process is implemented assuming that tRNA charging by synthetases, ternary complex formation (GTP : EF-Tu : charged-tRNA), and ternary complex diffusion to elongating ribosomes are not rate limiting for polypeptide polymerization (unless a flag is set to use tRNA charging as described below in optional feature). Given this assumption this process directly polymerizes amino acids based on the codon sequence of the mRNA transcript. Polymerization occurs across all ribosomes simultaneously and resources are allocated to maximize the progress of all ribosomes up to the limit of the expected ribosome elongation rate in a medium, available amino acids, and available transcripts (see Algorithm 2).

Optional features: tRNA charging. tRNA charging can be enabled for a simulation and and is used in `PolypeptideElongation` to capture a more mechanistic view of translation. The rate of amino acid incorporation becomes a function of the state of the cell including the codon sequence of mRNAs being translated as well as amino acid, tRNA, synthetase and ribosome concentrations. With the assumption that charging happens sufficiently fast ($k_{cat} \approx 100\ s^{-1}$ vs $\sim$1 s time step) and the state of the cell does not significantly change between time steps, the ratio of uncharged to charged tRNA can be adjusted until rates of tRNA charging ($v_{charging}$) and ribosome elongation ($v_{elongation}$) reach a steady state during each time step. This is shown with ODEs for each tRNA species, i, shown below:

$$\frac{d[tRNA_{charged,i}]}{dt} = v_{charging,i} - v_{elongation,i}$$

$$\frac{d[tRNA_{uncharged,i}]}{dt} = -\frac{d[tRNA_{charged,i}]}{dt}$$

Currently, the rates of charging and elongation are defined as below with the same constants for all species but could be altered to capture specific parameters for each species and additional concentrations (eg ATP).

$$v_{charging,i} = k_S \cdot [synthetase_i] \cdot \frac{\frac{[tRNA_{uncharged,i}]}{K_{M,tf}} \cdot \frac{[AA_i]}{\cdot K_{M,aa}}}{1 + \frac{[tRNA_{uncharged,i}]}{K_{M,tf}} + \frac{[AA_i]}{K_{M,aa}} + \frac{[tRNA_{uncharged,i}]}{K_{M,tf}} \cdot \frac{[AA_i]}{\cdot K_{M,aa}}}$$

$$v_{elongation,i} = f_i \cdot \frac{v_{rib} \cdot [ribosome]}{1 + \sum_i (f_i \cdot (\frac{k_{ta}}{[tRNA_{charged,i}]} + \frac{k_{tf}}{[tRNA_{uncharged,i}]} + \frac{[tRNA_{uncharged,i}]}{[tRNA_{charged,i}]} \cdot \frac{k_{ta}}{k_{tf}}))}$$

Where k_S is the synthetase charging rate, $K_{M,tf}$ is the Michaelis constant for free tRNA binding synthetases, $K_{M,aa}$ is the Michaelis constant for amino acids binding synthetases, f_i is the fraction of codon i to total codons to be elongated, v_{rib} is the max ribosome elongation rate, k_{ta} is the dissociation constant of charged tRNA to ribosomes and k_{tf} is the dissociation constant of uncharged tRNA to ribosomes.

With tRNA charging, translation will be limited by the calculated elongation rate ($v_{elongation}$) instead of the supply of amino acids to `PolypeptideElongation`. With a variable amount of amino acids being used at each time step, the concentration of each amino acid species, i, in the cell can vary as shown below based on the rate of supply of amino acids ($v_{supply,i}$ which includes both synthesis and uptake and is calculated for each condition) and the rate of charging ($v_{charging,i}$) as determined above:

$$\frac{d[AA_i]}{dt} = v_{supply,i} - v_{charging,i}$$

This change in amino acid concentrations will update the homeostatic target in `Metabolism`.

Algorithms

Algorithm 1: Algorithm for ribosome initiation on mRNA transcripts

Input : t_i translational efficiency of each mRNA transcript where $i = 1$ **to** n_{gene}
Input : $c_{mRNA,i}$ count of each mRNA transcript where $i = 1$ **to** n_{gene}
Input : c_{30S} count of free 30S ribosomal subunit
Input : c_{50S} count of free 50S ribosome subunit
Input : `multinomial()` function that draws samples from a multinomial distribution

1. Calculate probability (p_i) of forming a ribosome on each mRNA transcript weighted by the count and translational efficiency of the transcript.

$$p_i = \frac{c_{mRNA,i} \cdot t_i}{\sum_{i=1}^{n_{gene}} c_{mRNA,i} \cdot t_i}$$

2. Calculate maximal number of ribosomes that could be formed.

$$r_{max} = \texttt{min}(c_{30S}, c_{50S})$$

3 Sample multinomial distribution r_{max} times weighted by p_i to determine which transcripts receive a ribosome and initiate ($n_{init,i}$).

$$n_{init,i} = \texttt{multinomial}(r_{max}, p_i)$$

4 Assign $n_{init,i}$ ribosomes to mRNA transcript i. Decrement 30S and 50S counts.

$$c_{30S} = c_{30S} - \sum_{i=1}^{n_{gene}} n_{init,i}$$

$$c_{50S} = c_{50S} - \sum_{i=1}^{n_{gene}} n_{init,i}$$

Result: 70S ribosomes are formed from free 30S and 50S subunits on mRNA transcripts scaled by the count of the mRNA transcript and the transcript's translational efficiency.

Algorithm 2: Algorithm for peptide chain elongation and termination

Input : $e_{expected}$ expected elongation rate of ribosome ($e_{expected} < e_{max}$)
Input : p_i position of ribosome on mRNA transcript $i = 1$ **to** $n_{ribosome}$
Input : δt length of current time step
Input : c_{GTP} counts of GTP molecules
Input : L_j length of each mRNA $j = 1$ **to** n_{gene} for each coding gene.

`/* Elongate polypeptides up to limits of sequence, amino acids, or energy */`

for *each ribosome i on mRNA transcript j* **do**

1. Based on ribosome position p_i on mRNA transcript and expected elongation rate $e_{expected}$ determine "stop condition" position (t_i) for ribosome assuming no amino acid limitation. Stop condition is either maximal elongation rate scaled by the time step or the full length of sequence (i.e. the ribosome will terminate in this time step).

$$t_i = \texttt{min}(p_i + e_{expected} \cdot \delta t,\ L_j)$$

2. Derive sequence between ribosome position (p_i) and stop condition (t_i).

3. Based on derived sequence calculate the number of amino acids required to polymerize sequence $c^{req}_{aa,i}$ and number of GTP molecules required c^{req}_{GTP}.

4. Elongate up to limits:

if $\texttt{all}(c^{req}_{aa,k} < c_{aa,k})$ ***and*** $c^{req}_{GTP} < c_{GTP}$ **then**

Update the position of each ribosome to stop position

$$p_i = t_i$$

else

4a. Attempt to elongate all polypeptide fragments.

4b. Update position of each ribosome to maximal position given the limitation of $c_{aa,k}$ and c_{GTP}.

5. Update counts of $c_{aa,k}$ and c_{GTP} to reflect polymerization usage.

`/* Terminate ribosomes that have reached the end of their mRNA transcript */`

for *each ribosome i on transcript j* **do**

if $p_i == L_j$ **then**

1. Increment count of protein that corresponds to elongating polypeptide that has terminated.

2. Dissociate ribosome and increment 30S and 50S counts.

Result: Each ribosome is elongated up to the limit of available mRNA sequence, expected elongation rate, amino acid, or GTP limitation. Ribosomes that reach the end of their transcripts are terminated and released.

Associated files

wcEcoli Path	File	Type
wcEcoli/models/ecoli/processes	polypeptide_initiation.py	process
wcEcoli/models/ecoli/processes	polypeptide_elongation.py	process
wcEcoli/reconstruction/ecoli/dataclasses/process	translation.py	data

Table 1: Table of files for translation.

Difference from *M. genitalium* model. The `PolypeptideInitiation` process is implemented similarly in the *M. genitalium* with a few key differences. As the model of *E. coli* is not yet gene complete, the checks for initiation factors are not present. A major advance over the *M. genitalium* model is that the probability of ribosome initiation on a transcript is now proportional to the product of the mRNA count and its translational efficiency. In the *M. genitalium* model translational efficiency was not taken into account.

The `PolypeptideElongation` algorithm is implemented similarly to the *M. genitalium* model but again because the *E. coli* model is not yet gene complete, elongation factors are not accounted for. Additionally, tRNAs and their synthetases are not accounted for explicitly (unless a flag is set to use tRNA charging). Instead, the model directly polymerizes amino acids. This avoids computational issues with the simulation time step, tRNA pool size, and tRNA over expression that were present in the *M. genitalium* model. There is no implementation of ribosome stalling or tmRNAs. The polymerization resource allocation algorithm is the same as in *M. genitalium*.

Associated data

Parameter	Symbol	Units	Value	Reference
Translational efficiency[(1)]	t_i	RIB/mRNA	[0, 5.11]	[2]
Ribosome elongation rate	e	aa/s	18 (growth-dependent)	[1]
Protein counts (validation data)	$c_{protein}$	protein counts	[0, 250000]	[4]

Table 2: Table of parameters for translation process.
[(1)]Non-measured translational efficiencies were estimated by the average translational efficiency (1.11 RIB/mRNA).

References

[1] Hans Bremer and Patrick P Dennis. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus*, 3(1), 2008.

[2] Gene-Wei Li, David Burkhardt, Carol Gross, and Jonathan S Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635, 2014.

[3] J B Russell and G M Cook. Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiological reviews*, 59(1):48–62, March 1995.

[4] Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. The quantitative and condition-dependent escherichia coli proteome. *Nature biotechnology*, 34(1):104–110, 2016.