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`.

Initiation. `PolypeptideInitiation` models the complementation of 30S and 50S ribosomal subunits into 70S ribosomes on mRNA transcripts. This process is in many ways analogous to the `TranscriptInitiation` process - the number of initiation events per transcript is determined in a probabilistic manner and dependent on the number of free ribosomal subunits, each mRNA transcript's translation efficiency, and the counts of each type of transcript. The total number of ribosomes to activate in each time step is determined such that the average fraction of actively translating ribosomes matches experimental values. This is done by assuming a steady state concentration of active ribosomes (and therefore a constant active fraction), similar to what was done for RNA polymerases in `TranscriptInitiation`:

$$\frac{dR_{70S}}{dt} = p_{act} \cdot \min(R_{30S}, R_{50S}) - r \cdot R_{70S} = 0 \tag{1}$$

$$p_{act} = \frac{r \cdot R_{70S}}{\min(R_{30S}, R_{50S})} \tag{2}$$

where R_{70S} is the concentration of active 70S ribosomes, R_{50S} and R_{30S} are the concentrations of free 50S and 30S ribosomal subunits, respectively, p_{act} is the activation probability and r is the expected termination rate for active 70S ribosomes. Defining the active fraction as $f_{act} = \frac{R_{70S}}{R_{70S}+\min(R_{30S},R_{50S})}$, p_{act} can be defined in terms of the desired

active fraction:

$$p_{act} = \frac{r \cdot f_{act}}{1 - f_{act}} \tag{3}$$

This activation probability is then used to determine how many 70S ribosomes will be formed and initiated. These newly initiated 70S ribosomes are distributed to mRNA transcripts based on their translation probabilities, which is computed by normalizing the product of the translational efficiency of each transcript and the counts of each transcript.

$$p_i = \frac{c_{mRNA,i} \cdot t_i}{\sum_{j=1}^{n_{gene}} c_{mRNA,j} \cdot t_j} \tag{4}$$

Here, p_i is the translation probability of the transcript of gene i, $c_{mRNA,i}$ is the count of the transcript of gene i, and t_i is the translation efficiency of the transcript of gene i. The translational efficiencies of each transcript were calculated from ribosomal profiling data [2]. For transcripts whose translation efficiencies were not given in this dataset, the average of the existing efficiency values was used as the translation efficiency. Full 70S ribosomes are formed on mRNA transcripts by sampling a multinomial distribution with the p_i's calculated above as the probabilistic weights (See Algorithm 1).

This process is implemented similarly in the *M. genitalium* model with a few key differences. Unlike the *M. genitalium* model, the *E. coli* model is not yet gene complete, and hence does not check for initiation factors. A major advancement 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 efficiencies were not taken into account.

Elongation. `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 within the limits of the maximum ribosome elongation rate, available amino acids and GTP, and the length of the transcript (see Algorithm 2).

Under our simulation conditions, we empirically observe that the rate of translation elongation is always limited by the supply of amino acids, not by the elongation rates of ribosomes or the supply of GTP. Thus, the rate at which amino acids are supplied to translation largely determines the synthesis rates of proteins, which in turn is highly correlated with the growth rate of the cell. We have therefore added an option to add Gaussian noise to this supply rate of amino acids to translation, in cases where heterogeneity in growth rates between individual simulations would be a desired outcome.

Unlike the *M. genitalium* model, this process in the *E. coli* model does not account for elongation factors. 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 into elongating polypeptides. This avoids computational issues with the simulation time step, tRNA pool size, and tRNA overexpression that were present in the *M. genitalium* model. There is no implementation of ribosome stalling or tmRNAs at this point. The polymerization resource allocation algorithm is the same as in *M. genitalium*.

tRNA charging. tRNA charging is used in `PolypeptideElongation` to capture a more mechanistic view of translation but can be optionally disabled. 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 ~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 produced and used at each time step, the concentration of each amino acid species, i, in the cell can vary as shown below, which will update the homeostatic target in `Metabolism`:

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

where $f_{supply,i}$ is a function dependent on internal amino acid concentration and the presence of external amino acids, $v_{supply,i}$ is the rate of supply of amino acids, which includes both synthesis and uptake and is calculated for each condition based on the expected doubling time, and $v_{charging,i}$ is the rate of charging as determined above. $f_{supply,i}$ is defined as:

$$f_{supply,i} = f_{base\_synthesis,i} + f_{inhibited\_synthesis,i} + f_{import,i} - f_{export,i}$$

where

$$\begin{aligned}
f_{base_synthesis,i} &= c_{1,i} \\
f_{inhibited_synthesis,i} &= \frac{1}{1 + \frac{[AA_i]}{K_{I,i}}} \\
f_{import,i} &= \begin{cases} c_{2,i} & \text{if } AA_i \text{ in environment} \\ 0 & \text{otherwise} \end{cases} \\
f_{export,i} &= \frac{[AA_i]}{K_{M,i} + [AA_i]}
\end{aligned}$$

$c_{1,i}$, $c_{2,i}$, $K_{I,i}$, and $K_{M,i}$ can be determined by defining parameters f_I and f_M and constraints below that represent the fraction of contributions to the supply rate at the expected amino acid concentrations when the cell is in the presence of amino acids in the environment and when amino acids are not present:

$$\begin{aligned}
\text{when } [AA_i] = [AA_{i,basal}]: \quad & f_{inhibited_synthesis,i} = f_I \\
& f_{supply,i,basal} = 1 \\
\text{when } [AA_i] = [AA_{i,amino_acid}]: \quad & f_{export,i} = f_M \\
& f_{supply,i,amino_acid} = 1
\end{aligned}$$

Solving shows that the parameters are defined as:

$$\begin{aligned}
K_{I,i} &= \frac{f_I \cdot [AA_{i,basal}]}{1 - f_I} \\
K_{M,i} &= \left(\frac{1}{f_M} - 1\right) \cdot [AA_{i,amino_acid}] \\
c_{1,i} &= 1 - \left(f_I - \frac{[AA_{i,basal}]}{K_{M,i} + [AA_{i,basal}]}\right) \\
c_{2,i} &= 1 - \left(c_{1,i} + \frac{1}{1 + \frac{[AA_{i,amino_acid}]}{K_{I,i}}} - f_M\right)
\end{aligned}$$

Algorithms

Algorithm 1: Ribosome initiation on mRNA transcripts

Input : f_{act} fraction of ribosomes that are active

Input : r expected termination rate for active ribosomes

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_{act}) of free ribosomal subunits binding to a transcript.

$$p_{act} = \frac{r \cdot f_{act}}{1 - f_{act}}$$

2. Calculate the number of 70S ribosomes that will be formed and initiated ($c_{70S,b}$).

$$c_{70S,b} = p_{act} \cdot \texttt{min}(c_{30S}, c_{50S})$$

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

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

4. Sample multinomial distribution $c_{70S,b}$ times weighted by p_i to determine which transcripts receive a ribosome and initiate ($n_{init,i}$).

$$n_{init,i} = \texttt{multinomial}(c_{70S,b}, p_i)$$

5. 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: 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

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 data

Parameter	Symbol	Units	Value	Reference
Active fraction of ribosomes	f_{act}	-	0.8	[1]
Translational efficiency[(1)]	t_i	ribosomes/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 1: Table of parameters for translation process.
[(1)]Non-measured translational efficiencies were estimated to be equal to the average translational efficiency (1.11 ribosomes/mRNA).

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
wcEcoli/reconstruction/ecoli/flat	proteins.tsv	raw data
wcEcoli/reconstruction/ecoli/flat	translationEfficiency.tsv	raw data
wcEcoli/validation/ecoli/flat	schmidt2015_javier_table.tsv	validation data

Table 2: Table of files for translation.

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.