Translation and protein decay modules of a benchmark whole-cell model.

Jonathan R. Karr*[1], Pedro Mendes[2], Denis Kazakiewicz[3], and James Yurkovich[4]

[1]Department of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York NY 10029, USA
[2]Manchester Centre for Integrative Systems Biology, University of Manchester, Manchester M1 7DN, U.K
[3]Center for Statistics, Universteit Hasselt, Hasselt BE3500, Belgium
[4]Department of Bioengineering , University of California, San Diego, La Jolla, CA 92093, USA

April, 2016

1 Modules description

1.1 Protein decay

Models decay of each protein species (monomers and complexes) as a single reaction with kinetic rate that is linear in the protein copy number. The degradation of cytosolic-localized proteins is catalyzed by peptidases PepA, PepF, PepX, and Pip. The degradation of membrane-localized proteins is also catalyzed of the GTP-protease FtsH.

1.1.1 Species

- Small molecules
  - Amino acids: Produced by each protein degradation reaction
  - H_2O: Required for hydrolysis of each peptide bond and each GTP
  - GTP: Required to FtsH-catalyzed membrane protein degradation
  - GDP: Produced by hydrolysis of each GTP
  - PI: Produced by hydrolysis of each GTP
  - H^+: Produced by hydrolysis of each GTP
- Proteins
  - Protein monomers: includes both cytosolic and localized species
  - Protein complexes: includes only localized species because protein complexes are assembled in place
- RNA: Degradation of ribonucleoprotein complexes releases rRNA and sRNA subunits
  - rRNA
  - sRNA: scRNA (MG_0001), RNase P (MG_0003)
- Enzymes
  - Peptidases PepA (MG_391_HEXAMER), PepF (MG_183_MONOMER), PepX (MG_324_MONOMER), Pip (MG_020_MONOMER): Catalyze the hydrolysis of each peptide bond. Their individual functions are not known.
  - Protease FtsH: Required, in addition to the proteases, to degrade membrane proteins

1.1.2 Reactions

- Protein monomer degradation
  - Hydrolyzes peptide bonds, producing individual amino acids from proteins
  - Stoichiometry:

$$\text{monomer}_i + (l_i - 1)\ \text{H}_2\text{O} \xrightarrow{\text{Peptidases}} \sum_j a_{ij}\ \text{AA}_i,$$

    where a_{ij} is the composition of amino acid j in protein monomer i and $l_i = \sum_j a_{ij}$ is the length of protein monomer i.

*Contact: karr@mssm.edu

- Catalysis: Peptidases PepA (MG_391_HEXAMER), PepF (MG_183_MONOMER), PepX (MG_324_MONOMER), Pip (MG_020_MONOMER); their individual functions are not known.
- Rate law:

$$r_{\text{decay,protein}} = \frac{k_{\text{cat}} \cdot [\text{PROTEASE}]_1 \cdot ... \cdot [\text{PROTEASE}]_n \cdot [\text{PROTEIN}]_i}{1 + \frac{[\text{PROTEIN}]_i}{k_m}}$$

- Note: different proteases cut at specific cleavage points. Keeping monomer degradation reaction lumped allows to avoid tracking for all possible combinations.

- Protein complex degradation
  - Hydrolyzes all peptide bonds, producing individual amino acids from proteins subunits and releases RNA subunits
  - Stoichiometry:

$$\text{complex}_i + (l_i - n_i + g_i)\ \text{H}_2\text{O} + g_i\ \text{GTP} \xrightarrow{Protease,peptidases} \sum_j a_{ij}\ \text{AA}_i + \sum_j r_{ij}\text{RNA}_i + g_i\ \text{GDP} + g_i\ \text{PI} + g_i\ \text{H}^+$$

$$g_i = \begin{cases} l_i/\lambda & \text{membrane proteins} \\ 0 & \text{otherwise} \end{cases},$$

    where a_{ij} is the composition of amino acid j in protein complex i, $l_i = \sum_j a_{ij}$ is the total number of amino acids in protein complex i, n_i is the number of protein monomer subunits in protein complex i, r_{ij} is the composition of RNA j in protein complex i, $\lambda = 1.8$ is average number of amino acids cleaved by FtsH per GTP, and g_i is the GTP cost of degrading protein complex i.
  - Catalysis: Peptidases PepA (MG_391_HEXAMER), PepF (MG_183_MONOMER), PepX (MG_324_MONOMER), Pip (MG_020_MONOMER); their individual functions are not known. Membrane proteins are also degraded by protease FtsH (MG_457_HEXAMER).
  - Kinetics: First order in the protein concentration, $v = k_i[\text{complex}]_i$ where $k_i = \ln 2/\tau_i$ and τ_i is the half-life of protein complex i

1.2 Translation

Represents the translation of each protein monomer species as single reaction. Note: mRNA species represent transcription units (i.e. one or more protein-coding genes).

1.2.1 Species

- Small molecules
  - H_2O
  - H^+
  - GTP: 1 GTP hydrolyzed per translation initiation, 2 GTP hydrolyzed per elongation reaction, 2 GTP hydrolyzed per translation termination
  - GDP
  - PI
- $tRNA^{AA}$
- Protein monomers
- mRNA: Note, mRNA species represent transcription units (i.e. transcription units can include one or more ORFs)
- Enzymes
  - 70S ribosome (RIBOSOME_70S)
  - Translation initiation factor IF-1 (MG_173_MONOMER)
  - Translation initiation factor IF-2 (MG_142_MONOMER)
  - Translation initiation factor IF-3 (MG_196_MONOMER)
  - Translation elongation factor G (MG_089_DIMER)
  - Translation elongation factor P (MG_026_MONOMER)
  - Translation elongation factor Tu (MG_451_DIMER)
  - Translation elongation factor Ts (MG_433_DIMER)
  - Peptide chain release factor (MG_258_MONOMER)
  - Ribosome recycling factor (MG_435_MONOMER)

1.2.2 Reactions

- Single lumped reaction for each the translation of each protein monomer i that is an ORF on mRNA k

Table 1. SBO terms used in the Translation and in the Protein Decay modules

Type in WholeCell	complex (macromolecule)	Type in SBGN PD
metabolite	SBO:0000247	simple chemical
rna	SBO:0000354	nucleic acid feature (=n.a.f)
rna-aa	SBO:0000253	complex (simple chemical + n.a.f)
protein-complex	SBO:0000253	complex (macromolecule)
protein-monomer	SBO:0000245	macromolecule

- Stoichiometry:

$$\sum_j t_{ij}\ \mathrm{tRNA}_j^{\mathrm{AA}}\ +\ (2l_i+3)\ \mathrm{GTP}\ +\ (2l_i+2)\ \mathrm{H_2O} \xrightarrow{Ribosome, factors, mRNA_k} \mathrm{monomer}_i + \sum_j t_{ij}\ \mathrm{tRNA}_j + (2l_i+3)\ \mathrm{GDP} + (2l_i+3)\ \mathrm{PI} + (l_i+3)\ \mathrm{H}^+$$

  where t_{ij} is the tRNA composition of protein-coding gene i and $l_i = \sum_j t_{ij}$ is the length of protein monomer i
- Catalysis: 70S ribosome and initiation, elongation, and termination factors
- Kinetics: There are many entities contributing to the rate of translation reaction. Instead of listing all of them simplified rate law was used. It is based on the assumption that the rate is limited by a contributor which is present in the lowest amount (in comparison to other contributors) So the minimum function was used.

$$r_{\mathrm{translation}} = \frac{k_{\mathrm{cat}} \cdot \min(\mathrm{RNA}, \mathrm{RIBOSOME}, \mathrm{FACTORS}) \cdot \min(\mathrm{tRNA}_1^{\mathrm{AA}}, \ldots, \mathrm{tRNA}_j^{\mathrm{AA}}) \cdot [\mathrm{GTP}]^n}{(1 + \frac{\min(\mathrm{tRNA}_1^{\mathrm{AA}}, \ldots, \mathrm{tRNA}_j^{\mathrm{AA}})}{k_{m,i}}) \cdot (1 + \frac{[\mathrm{GTP}]^n}{k_{m,\mathrm{GTP}}})}$$

- Note: Previous attempt to construct Translation reactions describing the per-amino acid elongation resulted in very large SBML model (more than 1 Gb) which was impossible to simulate because of it's size. More information on that attempt can be found in `https://github.com/dagwa/wholecell-translation/blob/master/README_Joe.txt`

2 Implementation

Python scripts were used to generate SBML xml file (level 3 version 1) by parsing the data from Excel files (generated in Matlab). ProteinDecay.xls file is the input for protein-decay.py. As an output, protein-decay.py generates Decay_lvl3_v1.xml. Similarly, protein-decay.py generates Decay_lvl3_v1.xml (and takes Translation.xls). Python scripts are of the same basic structure. xlrd library is used to parse data from the excel tables, libSBML library is used to generate SBML models. SBO terms were then added to the SBML models by java script (Vasundra Toure) (Table 1).

The main hurdle in the programming translation module was the absence of the min() function in SBML. The workaround was constructed by Martin Scharm. It was based on the following min() function definition $min(a,b) = \frac{a+b-|a-b|}{2}$.