Proteome constrained model of *L. lactis*

In this document we show detailed description of proteome constrained model for *Lactococcus lactis* MG1363. In addition to metabolism, the model includes many subprocesses in protein expression: transcription, stable RNA cleavage, mRNA degradation, tRNA modification, rRNA modification, tRNA charging, ribosozzmal assembly, translation, protein maturation, protein assembly, enzyme formation, and protein degradation. Besides, some other reactions were formulated in the model for modelling purpose, including generic RNA renaming, enzyme dilution and RNA dilution reactions. We developed a Matlab package to construct the model for *L. lactis*, which requires the COBRA toolbox (PMID: 30787451). The package as well as the modelling approach are only for *L. lactis* here but expected to be used in future for other prokaryotes with existing genome-scale metabolic models (M models). The procedure is divided into five steps: information collection, reformulation of M model, construction of protein expression model (E model), formulation of other reactions for modelling purpose, and determination of constraints.

Contents

[1 Information collection 2](#_Toc13217541)

[1.1 Coverage of E model 2](#_Toc13217542)

[1.2 RNA 3](#_Toc13217543)

[1.3 Sequence information 3](#_Toc13217544)

[1.4 Transcription unit 3](#_Toc13217545)

[1.5 RNA modification 4](#_Toc13217546)

[1.6 Protein stoichiometry 4](#_Toc13217547)

[1.7 EC number 5](#_Toc13217548)

[1.8 N-terminus prediction 5](#_Toc13217549)

[2 M model reformulation 5](#_Toc13217550)

[2.1 Reconstructing the M model 6](#_Toc13217551)

[2.2 Changing unknown genes 6](#_Toc13217552)

[2.3 Adding dummy GPR (Gene-Protein-Reaction association) 7](#_Toc13217553)

[2.4 Splitting isozymes and reversible reactions 7](#_Toc13217554)

[3 E model construction 9](#_Toc13217555)

[3.1 Transcription 10](#_Toc13217556)

[3.1.1 Transcription initiation 10](#_Toc13217557)

[3.1.2 Transcription elongation and termination 10](#_Toc13217558)

[3.2 Stable RNA cleavage 11](#_Toc13217559)

[3.3 mRNA degradation 12](#_Toc13217560)

[3.4 tRNA modification 13](#_Toc13217561)

[3.5 Generic RNA renaming 14](#_Toc13217562)

[3.6 rRNA modification 15](#_Toc13217563)

[3.7 Ribosomal assembly 15](#_Toc13217564)

[3.8 tRNA charging 18](#_Toc13217565)

[3.9 Translation 19](#_Toc13217566)

[3.9.1 Translation initiation 20](#_Toc13217567)

[3.9.2 Translation elongation 20](#_Toc13217568)

[3.9.3 Translation termination 22](#_Toc13217569)

[3.10 Protein maturation 23](#_Toc13217570)

[3.11 Protein assembly 24](#_Toc13217571)

[3.12 Enzyme formation 25](#_Toc13217572)

[3.13 Protein degradation 25](#_Toc13217573)

[3.14 Enzyme dilution 28](#_Toc13217574)

[3.15 RNA dilution 28](#_Toc13217575)

[4 Formulation of other reactions 29](#_Toc13217576)

[4.1 Adding an unmodeled protein 29](#_Toc13217577)

[4.2 Adding a biomass dilution reaction 29](#_Toc13217578)

[5 Constraints 31](#_Toc13217579)

[5.1 Collection of turnover rates for metabolic reactions 31](#_Toc13217580)

[5.2 Estimation of catalytic rates for protein expression process 31](#_Toc13217581)

[5.3 Constraints 33](#_Toc13217582)

# Information collection

## Coverage of E model

We determined the coverage (the number of proteins and RNA) of E model using two methods. The first method is comparative genomics analysis. We used two published ME models, which are for *Escherichia coli* (PMID: ﻿24084808) and *Thermotoga maritima* (PMID: ﻿22760628), as templates for gene orthology analysis using the EggNOG database (PMID: 26582926) (15-Apr-2015 09:03). Briefly, we imported all the protein genes from E models of *E. coli* and *T. maritima* and then searched in the database the orghologous genes in *L. lactis*. As a result, we collected 108 orthologous genes in *L. lactis* compared to *E. coli* (194) and 112 compared to *T. maritima* (159). By combining these two, we obtained 134 protein genes. The second method is subsystem analysis. We used the SEED subsystem (<http://pubseed.theseed.org/>) to collect protein genes in the protein expression process. We searched for “Lactococcus lactis subsp. cremoris MG1363”, chose the latest genome annotation version (416870.9), exported subsystem information, and selected all the annotated genes in the category “RNA metabolism” and “Protein metabolism” (choosing these two categories is because most of the selected genes based on the first step are located in these two categories). The subsystem analysis resulted in 187 protein genes. Combining the results from two methods, we obtained in total 218 protein genes.

Next, we manually selected protein genes for and then assigned themto each subprocess or functional machinery in the E model. As a result, we determined 157 protein genes for the protein expression process in *L. lactis*.

## RNA

Besides protein genes, RNA genes are essential in protein expression process as they make mRNA and tRNA as transcription and translation machineries. We collected RNA genes from NCBI database. There are 19 rRNA and 62 tRNA genes in *L. lactis* MG1363, in total 81 RNA genes.

## Sequence information

We downloaded sequence information (chromosome and plasmid sequences) from NCBI database:

Chromosome sequences (<https://www.ncbi.nlm.nih.gov/nuccore/NC_009004.1>),

pNZ712 sequences (<https://www.ncbi.nlm.nih.gov/nuccore/1044892037>),

pLP712 sequences (<https://www.ncbi.nlm.nih.gov/nuccore/NC_019377.1>).

All the base sequences correspond to the base sequences of the DNA coding strand, i.e., same as transcripts produced (with thymine replaced by uracil).

Besides, locus ID names should be collected as there are two types of ID used in NCBI, i.e., old and new ID. We downloaded the ID relationship from BioCyc database (PMID: 29447345).

## Transcription unit

We collected all the transcription units (TUs) for *L. lactis*. The model produces TU as transcription product rather than transcript of a single gene. For chromosome genes, the TUs were predicted and stored in BioCyc, and we downloaded them (Lactococcus lactis, Subspecies cremoris, Strain MG1363, version 21.5). For plasmid genes, we manually grouped neighboring genes in the same strand into one TU.

## RNA modification

We assumed that transcribed tRNA and rRNA should be modified before they participate in protein expression process. We collected a great number of post-transcriptional modifications of RNA, which is a comprehensive dataset for *L. lactis*.

We determined tRNA modifications based on a published paper (PMID: 25040919) and bioinformatics predictions (PMID: 27016142). The published paper experimentally identified 16 different modifications in 40 tRNA in *L. lactis*. Then, we used the bioinformatics analysis provided on the website (<http://genesilico.pl/trnamodpred/>) to predict the remaining modifications which cannot be identified experimentally or detected by standard MS approaches (e.g. pseudouridine). We filtered the resulting predictions using the term “gram positive bacteria”. Subsequently, we determined proteins responsible for the modification events based on the predicted enzymes and genome annotations. For most of the modification events we can find the catalysts, but we cannot find genes responsible for C:ac4C and U:mo5U, so we assumed that these two modification events occur spontaneously. At last, we formulated all the tRNA modification reactions as done for *E. coli* and *T. maritima*.

Due to the fact that there are no organism-specific publications regarding rRNA modification in *L. lactis*, we determined rRNA modification events for *L. lactis* based on the information from *E. coli*. We only modeled modifications on 16S and 23S rRNA as 5S rRNA modifications are infrequent in bacteria. We aligned rRNA in *L. lactis* and *E. coli* using ClustalW2 (PMID: 17846036) and then determined modification types and sites based on the *E. coli* rRNA modification events collected before (PMID: 19282977). As a result, we obtained modification events at 10 (11 in *E. coli*) positions for 16S rRNA and 24 (24 in *E. coli*) positions for 23S rRNA in *L. lactis*. Using the same method as mentioned in tRNA modification determination process, we collected genes encoding catalysts based on KO, COG or KEGG BLAST. All the rRNA modification reactions were formulated as done for *E. coli* and *T. maritima*.

## Protein stoichiometry

For each functional protein, we should determine whether its functional unit is a monomer or oligomer. Firstly, we downloaded all the proteins with their sequences from the RCSB Protein Data Bank (PMID: 30357364) (2018.10.30). Then, we used BLASTP to search the protein with reported protein stoichiometry information that closely related to a protein of interest. For a protein in *L. lactis*, we normally chose the identified protein with the highest bit scores, but we then chose the one with the lowest E values if highest bit scores correspond to two or more proteins. When proteins had the same bit scores and E values, we chose the simplest structure from PDB. Note that we adopted the protein stoichiometry information from the biological assembly rather than the asymmetric unit as the former one is believed to be the functional form of the protein molecular (<http://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/biological-assemblies>). We only considered the proteins with identity >= 30% and coverage >= 80%. When no similar proteins were found, we assumed that the protein of interest was a monomer and marked in the model. This resulted in a comprehensive dataset of protein stoichiometry for *L. lactis*.

## EC number

We collected EC numbers for proteins in *L. lactis*, which can be used to retrieve kcat values. The EC numbers were downloaded from the Uniprot (PMID: 29425356) and KEGG (PMID: 30321428) database. The EC numbers of plasmid genes were assigned according to protein homology.

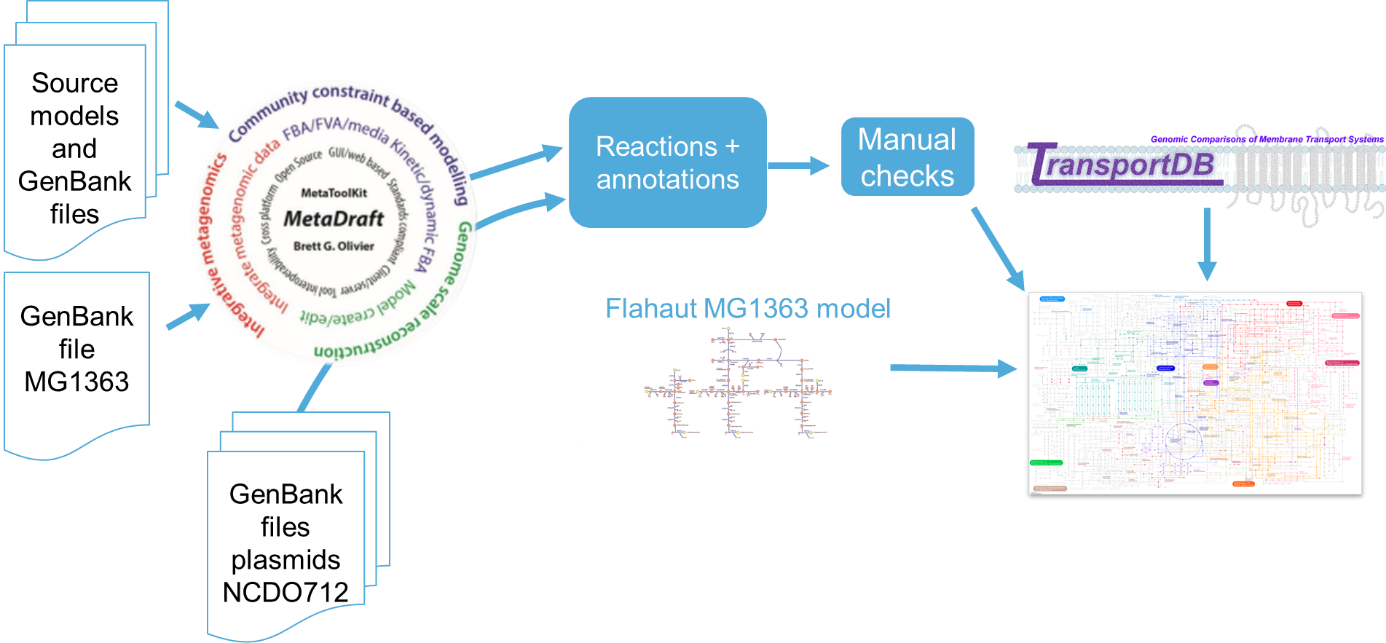
## N-terminus prediction

Due to the fact that the N-terminus is usually cleaved after a peptide has been translated, we predicted for each protein whether or not the N-terminal methionine should be cleaved. This was performed using the TermiNator (<https://bioweb.i2bc.paris-saclay.fr/terminator3/>). Before running the prediction, we selected “prokaryote”, “eubacteria”, “intrinsic, chromosome-encoded gene”, and “no LPR cleavage”. Then we uploaded the sequences of all the proteins and ran the prediction. For the result we only chose the proteins with likelihood > 80%, while for the rest we assumed that there is no N-terminal methionine cleavage.

# M model reformulation

In this step, we updated the existing genome-scale metabolic model of *L. lactis* (Flahaut et al. 2013, PMID:23974365) and then reformulated the reactions to enable integration with E model.

## Reconstructing the M model

For the PC-model we need as much gene protein reaction associations (GPR) as possible, because we want to assign a protein cost to each metabolic reaction. Therefore, we updated the existing metabolic model of *L. lactis* in terms of reactions, transport reactions and especially in terms of GPRs. For this we used MetaDraft (version 0.7.2, Brett G Olivier, cbmpy-metadraft, doi:10.5281/zenodo.2398336), a tool that reconstructs genome-scale metabolic networks based on previous manually curated ones by homology between genes. As source models we used the existing *L. lactis* MG1363 model by Flahaut, and the default models in the MetaDraft database in the following importance ranking: the Lactic Acid Bacteria (LAB) models (*L. lactis* model (Goel et al. 2015), *L. Plantarum* model (Teusink et al. 2006)), then the remaining BIGG2 models that were present in the default MetaDraft database (iJO1366, iAF692), iYO844, iHN637, iAF987, iIT341, iYL1228, iJN746, iSB619, iMM904, iS\_1188, STM\_v1\_0). We manually checked all the reactions, whether the reaction and annotated genes were already present in the existing metabolic model and whether the annotated function is seen for lactis in databases. Each newly added reaction was checked within uniport, string gene database and Artemis, to find an indication whether this gene-function combination is present in *L. lactis* cremoris strain MG1363. We then conducted a search for transporter genes in lactis with the TransportDB to improve exiting GPRs, add new GPRs and new reactions.

Furthermore, we converted all metabolite and reaction Ids to be compatible with the BIGG database (PMID: 26476456) for (Miriam) annotations, reaction mapping and visualization purposes. For this we used a newly developed in-house tool (S. Mendoza et al. 2019)

We adjusted the biomass function of the metabolic model for practical reasons: all RNA modification reactions and all tRNA loading reactions were removed from the model, because they will be present in the E-part. We therefore also adjusted the biomass protein reaction such that it includes the single amino acids in the correct ratios instead of the tRNAs loaded with amino acids.

The new M-model and a comparison with the Flahaut model is attached in the supplementary materials [ADD REFERENCE].

## Changing unknown genes

There are some unknown genes in the M model because of missing annotations. The unknown genes cannot be ignored, especially for the case that a complex has several genes but most of them are unknown, removing the unknown genes will thereby decrease the usage of materials and energy for producing the complex. Therefore, we made rules to cope with unknown genes. For the case “(unknown or geneA)”, we removed the unknown gene. For the case “(unknown and geneA)”, we assumed that the complex is made up of two geneA protein, i.e., the unknown one is regarded as geneA.

## Adding dummy GPR (Gene-Protein-Reaction association)

There are a lot of reactions without GPRs due to missing information in genome annotations. We assigned a “dummy” monomer to represent the enzyme for these reactions to eliminate the bias toward using these reactions without energy and material cost. We did not assign the “dummy” GPR for spontaneous reactions, e.g., “L glutamate 5 semialdehyde dehydratase”. The length of the “dummy” enzyme was assumed to be a monomer with 636 amino acids, the composition was assumed to be the same as that in biomass composition, and the kcat of the enzyme was assumed to be the median among other experimentally-determined values.

The number of amino acids of the dummy GPR was determined as follow. Firstly, the median molecular weight of metabolic catalysts was calculated, i.e., 71503 g/mol. Then, this value was divided by the average amino acid molecular weight of 113 g/mol to obtain the number of amino acids.

## Splitting isozymes and reversible reactions

There are several types of GPRs in the M model, e.g., “geneA”, “(geneA and geneB)”, “(geneA or geneB)”, “(geneA and (geneB or geneC))”, “((geneA and geneB) or (geneC and geneD))”, and so on. We split the reaction with isozymes, i.e., with “or” in its GPR, into multiple reactions catalyzed by each isozyme. For instance, a reaction with the GPR “(geneA and (geneB or geneC))” should be split into two reactions: one with the GPR “(geneA and geneB)” and the other with the GPR “(geneA and geneC)”.

We also split reversible reactions into forward and reverse reactions, which was done only for enzymatic reactions.

In the following part, we use acetate kinase reaction as an example to show how a metabolic reaction was split into multiple reactions in the model.

In the M model, it is a reversible reaction catalyzed by two isozymes:

|  |
| --- |
| * **Reaction ID:** R\_ACKr |
| * **Reaction Formula:** M\_atp\_c + M\_ac\_c <-> M\_adp\_c + M\_actp\_c |
| * **Catalyst:** (llmg\_2288 or llmg\_2289) |

Subsequently, we adjusted the reaction ID by adding “M” to mark it as metabolic reaction, and then split it into forward and reverse reactions catalyzed by different isozymes. As a result, we obtained four unique reactions below:

|  |
| --- |
| * **Reaction ID:** R\_M\_ACKr\_1\_fwd |
| * **Reaction Formula:** M\_atp\_c + M\_ac\_c -> M\_adp\_c + M\_actp\_c |
| * **Catalyst:** M\_ACKr\_1\_fwd\_Enzyme\_c |

|  |
| --- |
| * **Reaction ID:** R\_M\_ACKr\_1\_rvs |
| * **Reaction Formula:** M\_adp\_c + M\_actp\_c -> M\_atp\_c + M\_ac\_c |
| * **Catalyst:** M\_ACKr\_1\_rvs\_Enzyme\_c |

|  |
| --- |
| * **Reaction ID:** R\_M\_ACKr\_2\_fwd |
| * **Reaction Formula:** M\_atp\_c + M\_ac\_c -> M\_adp\_c + M\_actp\_c |
| * **Catalyst:** M\_ACKr\_2\_fwd\_Enzyme\_c |

|  |
| --- |
| * **Reaction ID:** R\_M\_ACKr\_2\_rvs |
| * **Reaction Formula:** M\_adp\_c + M\_actp\_c -> M\_atp\_c + M\_ac\_c |
| * **Catalyst:** M\_ACKr\_2\_rvs\_Enzyme\_c |

In the reaction ID we used the number “1” and “2” (and “3” or more for the case that has three or more isozymes) to distinguish reactions catalyzed by different isozymes and “fwd” and “rvs” to represent forward and reverse direction, respectively. It should be noted that we used new term, e.g., “M\_ACKr\_1\_fwd\_Enzyme\_c”, which is derived from the reaction ID, to represent catalyst rather than the original gene name in the M model. This enables each enzymatic reaction to have its own and unique catalyst ID, and thereby makes it possible to perform coupling constraints using inequalities. Therefore, we had to formulate extra reaction representing the formation of catalyst for each reaction. We moved such reaction in a new subprocess named “Enzyme formation”. For the four reactions above, we formulated reactions for catalyst formation:

|  |
| --- |
| * **Reaction ID:** R\_M\_ACKr\_1\_fwd\_Enzyme |
| * **Reaction Formula:** llmg\_2288\_2mer\_c -> M\_ACKr\_1\_fwd\_Enzyme\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_M\_ACKr\_1\_rvs\_Enzyme |
| * **Reaction Formula:** llmg\_2288\_2mer\_c -> M\_ACKr\_1\_rvs\_Enzyme\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_M\_ACKr\_2\_fwd\_Enzyme |
| * **Reaction Formula:** llmg\_2289\_2mer\_c -> M\_ACKr\_2\_fwd\_Enzyme\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_M\_ACKr\_2\_rvs\_Enzyme |
| * **Reaction Formula:** llmg\_2289\_2mer\_c -> M\_ACKr\_2\_rvs\_Enzyme\_c |
| * **Catalyst:** - |

In addition, we made dilution reactions for all the catalysts in the model, which belong to a subprocess named “Enzyme dilution”. For the example here, we also had four dilution reactions:

|  |
| --- |
| * **Reaction ID:** R\_dilution\_M\_ACKr\_1\_fwd\_Enzyme |
| * **Reaction Formula:** M\_ACKr\_1\_fwd\_Enzyme\_c -> |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_dilution\_M\_ACKr\_1\_rvs\_Enzyme |
| * **Reaction Formula:** M\_ACKr\_1\_rvs\_Enzyme\_c -> |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_ dilution\_M\_ACKr\_2\_fwd\_Enzyme |
| * **Reaction Formula:** M\_ACKr\_2\_fwd\_Enzyme\_c -> |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_ dilution\_M\_ACKr\_2\_rvs\_Enzyme |
| * **Reaction Formula:** M\_ACKr\_2\_rvs\_Enzyme\_c -> |
| * **Catalyst:** - |

# E model construction

E model accounts for synthesis of proteins used in both M model and E model itself. Therefore, we constructed the protein expression process, including transcription, stable RNA cleavage, mRNA degradation, rRNA modification, tRNA modification, ribosomal assembly, tRNA charging, translation, protein maturation, and protein assembly. Besides, we formulated generic RNA renaming reactions for modelling purpose.

## Transcription

Transcription is the first step of protein expression, which, using DNA as a template, consumes nucleotides and energy to produce transcription units. The model only considers the TUs containing the protein and RNA genes involved in metabolism and gene expression. We divided transcription process into two steps, i.e., 1) transcription initiation, and 2) transcription elongation and termination.

### Transcription initiation

Transcription initiation is catalyzed by RNA polymerase together with one or more sigma factors. In *L. lactis*, the core RNA polymerase consists of two alpha subunits (llmg\_2354), one beta subunit (llmg\_1982), one beta prime subunit (llmg\_1981), one omega subunit (llmg\_2154), and one delta subunit (llmg\_0608). Besides, there is only one sigma factor in *L. lactis* according to genome annotation, i.e., RNA polymerase sigma factor RpoD (llmg\_0521). We formulated a reaction representing the formation of RNA polymerase with sigma factor:

|  |
| --- |
| * **Reaction ID:** R\_RNAP\_sf\_Enzyme |
| * **Reaction Formula:** 2 llmg\_2354\_2mer\_c + llmg\_1982\_Monomer\_c + llmg\_1981\_assumed\_Monomer\_c + llmg\_2154\_assumed\_Monomer\_c + llmg\_0608\_assumed\_Monomer\_c + llmg\_0521\_Monomer\_c -> RNAP\_sf\_Enzyme\_c |
| * **Catalyst:** - |

We assumed that the first 16 nucleotides were transcribed during the transcription initiation step, and 15 ppi were released. We use the TU “TU1G1R-100” as an example to show the formula of transcription initiation below:

|  |
| --- |
| * **Reaction ID:** R\_TU1G1R\_100\_Initiation |
| * **Reaction Formula:** 6 M\_atp\_c + M\_gtp\_c + 7 M\_utp\_c + 2 M\_ctp\_c -> 15 M\_ppi\_c + TU1G1R\_100\_Initiated\_c |
| * **Catalyst:** RNAP\_sf\_Enzyme\_c |

### Transcription elongation and termination

We lumped transcription elongation and termination to decrease the total number of reactions. From the KEGG database (Transcription machinery [BR: llm03021]), we collected transcription elongation factor NusA (llmg\_1796), transcription termination protein NusB (llmg\_1878), transcription antitermination protein NusG (llmg\_2388), and transcription elongation factor GreA (llmg\_0610). Besides, we manually added transcription-repair coupling factor, Mfd (llmg\_0013). It should be noted that we cannot find Rho termination factor in *L. lactis* genome, so all the transcription reactions are Rho independent. We formulated the formation of transcription complex, which consists of all the factors mentioned above:

|  |
| --- |
| * **Reaction ID:** R\_Transcription\_Complex\_Enzyme |
| * **Reaction Formula:** llmg\_1796\_2mer\_c + llmg\_1878\_assumed\_Monomer\_c + llmg\_2388\_Monomer\_c + llmg\_0610\_Monomer\_c + llmg\_0013\_Monomer\_c -> Transcription\_Complex\_Enzyme\_c |
| * **Catalyst:** - |

Then catalyzed by the transcription complex, the initiated TU will be elongated and terminated in this step (still using “TU1G1R\_100” as an example):

|  |
| --- |
| * **Reaction ID:** R\_TU1G1R\_100 |
| * **Reaction Formula:** 304 M\_atp\_c + 215 M\_gtp\_c + 276 M\_utp\_c + 191 M\_ctp\_c + TU1G1R\_100\_Initiated\_c -> 986 M\_ppi\_c + TU1G1R\_100\_c |
| * **Catalyst:** Transcription\_Complex\_Enzyme\_c |

## Stable RNA cleavage

There are two types of TUs produced by transcription process, one type is the TU exclusively containing protein genes while the other is the TU containing RNA genes. The former will be translated directly, but the latter should be cleaved before entering other subprocesses as RNA genes will not be translated. Therefore, we added stable RNA cleavage process in the model for the TUs containing RNA genes (28 TUs in the model).

Ribonuclease is responsible for stable RNA cleavage and there are many types of ribonucleases in *L. lactis* according to genome annotation and the KEGG database. They are RNase III (llmg\_1753), mini-RNase III (llmg\_2039), RNase P (llmg\_0142), RNase HII (llmg\_1176), RNase HIII (llmg\_2549), RNase M5 (llmg\_1883), RNase Z (llmg\_0412), RNase Z (llmg\_0604), RNase J1 (llmg\_0302), RNase BN (llmg\_0578), RNase J2 (llmg\_0876), RNase R (llmg\_1304), RNase R (llmg\_1586). However, we only adopted RNase III (llmg\_1753) in the model for catalyzing RNA cleavage reactions as we cannot find any supporting evidence. So, we formulated the formation of RNase III:

|  |
| --- |
| * **Reaction ID:** R\_RNase\_Enzyme |
| * **Reaction Formula:** llmg\_1753\_2mer\_c -> RNase\_Enzyme\_c |
| * **Catalyst:** - |

For the case that a TU contains only one RNA gene, we formulated a spontaneous reaction:

|  |
| --- |
| * **Reaction ID:** R\_TU1G1R\_1030\_cleavage |
| * **Reaction Formula:** TU1G1R\_1030\_c -> llmg\_tRNA\_36\_unmodified\_c |
| * **Catalyst:** - |

For the case that a TU contains more than one RNA gene, RNase III is used. It should be noted that one H2O should be used per time of cleavage for one TU, and one proton is produced. Besides, sub-TU could be produced in this process, which is the sequence between two RNA genes in the TU. The sub-TU can be either degraded directly or translated if it contains protein genes. Below is an example:

|  |
| --- |
| * **Reaction ID:** R\_TU1G1R\_3106\_cleavage |
| * **Reaction Formula:** 10 M\_h2o\_c + TU1G1R\_3106\_c -> 10 M\_h\_c + llmg\_rRNA\_60a\_unmodified\_c + llmg\_rRNA\_7\_unmodified\_c + llmg\_rRNA\_7a\_unmodified\_c + llmg\_tRNA\_62\_unmodified\_c + llmg\_tRNA\_61\_unmodified\_c + llmg\_tRNA\_60\_unmodified\_c + TU1G1R\_3106\_sub\_1\_c + TU1G1R\_3106\_sub\_2\_c + TU1G1R\_3106\_sub\_3\_c + TU1G1R\_3106\_sub\_4\_c + TU1G1R\_3106\_sub\_5\_c |
| * **Catalyst:** RNase\_Enzyme\_c |

## mRNA degradation

We formulated mRNA degradation reactions for all the TUs produced in transcription process and sub-TUs produced in stable RNA cleavage process. The catalysts for this process are degradosome and oligoribonuclease. In *L. lactis*, the degradosome seems to be RNA degradosome type D according to the KEGG database, which is similar to *Bacillus subtilis*. The type D degradosome is made up of RNaseY (llmg\_2156), CshA (llmg\_0369), RNaseJ1 (llmg\_0302), RNaseJ2 (llmg\_0876), Enolase (llmg\_0617), PNPase (llmg\_2044), and PfkA (llmg\_1118). Thereby, we formulated the formation of the mRNA degradation complex by integrating degradosome and oligoribonuclease Orn (llmg\_1825):

|  |
| --- |
| * **Reaction ID:** R\_mRNA\_Degradation\_Complex\_Enzyme |
| * **Reaction Formula:** llmg\_0617\_Monomer\_c + llmg\_1118\_4mer\_c + llmg\_0302\_4mer\_c + llmg\_0369\_assumed\_Monomer\_c + llmg\_0876\_4mer\_c + llmg\_1825\_Monomer\_c + llmg\_2044\_3mer\_c + llmg\_2156\_assumed\_Monomer\_c -> mRNA\_Degradation\_Complex\_Enzyme\_c |
| * **Catalyst:** - |

The mRNA degradation process is energy-consuming, and we assumed that one ATP needed for every four nucleotides degraded as with the modelling for *E. coli* and *T. maritima*. Nucleotide within TU was hydrolysed into nucleoside monophosphate except the one at the first position, which was returned as nucleoside triphosphate. We assumed that the degradation of one molecule of TU or sub-TU was catalysed by only one molecule of mRNA degradation complex regardless of the number of nucleotides. Below is an example:

|  |
| --- |
| * **Reaction ID:** R\_TU1G1R\_1034\_degradation |
| * **Reaction Formula:** 324 M\_atp\_c + 1619 M\_h2o\_c + TU1G1R\_1034\_c -> 418 M\_amp\_c + 1619 M\_h\_c + M\_gtp\_c + 324 M\_pi\_c + 324 M\_adp\_c + 385 M\_ump\_c + 213 M\_cmp\_c + 279 M\_gmp\_c |
| * **Catalyst:** mRNA\_Degradation\_Complex\_Enzyme\_c |

## tRNA modification

RNA modification is a process that some bases of an RNA molecule are changed. We modelled this process by changing one base at a time. It should be noted that there are three cases in the model, i.e., 1) modification without any metabolites consumed or produced, 2) modification without any catalysts, and 3) modification with metabolite and catalyst involved. We show the examples below for the three types:

1. modification without any metabolites consumed or produced

|  |
| --- |
| * **Reaction ID:** R\_tRNA\_modification\_llmg\_tRNA\_27\_modified\_1 |
| * **Reaction Formula:** llmg\_tRNA\_27\_unmodified\_c -> llmg\_tRNA\_27\_modified\_1\_c |
| * **Catalyst:** RNAmod\_llmg\_0395\_Enzyme\_c |

1. modification without any catalysts

|  |
| --- |
| * **Reaction ID:** R\_tRNA\_modification\_llmg\_tRNA\_27\_modified\_3 |
| * **Reaction Formula:** M\_accoa\_c + llmg\_tRNA\_27\_modified\_2\_c -> M\_coa\_c + llmg\_tRNA\_27\_modified\_3\_c |
| * **Catalyst:** - |

1. modification with metabolite and catalyst involved

|  |
| --- |
| * **Reaction ID:** R\_tRNA\_modification\_llmg\_tRNA\_27\_modified\_4 |
| * **Reaction Formula:** M\_amet\_c + llmg\_tRNA\_27\_modified\_3\_c -> M\_h\_c + M\_ahcys\_c + llmg\_tRNA\_27\_modified\_4\_c |
| * **Catalyst:** RNAmod\_llmg\_2424\_Enzyme\_c |

The formation of catalyst was also formulated in this stage and stored in “Enzyme formation” subprocess. There are in total 35 catalysts responsible for both tRNA and rRNA modification in *L. lactis*.

## Generic RNA renaming

We formulated generic RNA renaming reactions for tRNA and rRNA to save the number of reactions. This was done for tRNA after having been modified as different tRNA even sharing the same codon may show different sequences and thereby different modification events, but for rRNA before being modified as the same type of rRNA have identical sequences and thereby the same modification events. We show the renaming of tRNA-Val-TAC, which can be made by three genes in *L. lactis*, as an example below:

|  |
| --- |
| * **Reaction ID:** R\_Generic\_RNA\_llmg\_tRNA\_19\_tRNA\_Val\_GUA\_to\_tRNA\_Val\_GUA |
| * **Reaction Formula:** llmg\_tRNA\_19\_tRNA\_Val\_GUA\_c -> tRNA\_Val\_GUA\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_Generic\_RNA\_llmg\_tRNA\_47\_tRNA\_Val\_GUA\_to\_tRNA\_Val\_GUA |
| * **Reaction Formula:** llmg\_tRNA\_47\_tRNA\_Val\_GUA\_c -> tRNA\_Val\_GUA\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_Generic\_RNA\_llmg\_tRNA\_58\_tRNA\_Val\_GUA\_to\_tRNA\_Val\_GUA |
| * **Reaction Formula:** llmg\_tRNA\_58\_tRNA\_Val\_GUA\_c -> tRNA\_Val\_GUA\_c |
| * **Catalyst:** - |

We show the renaming of 16S-rRNA, which can be made by six genes in *L. lactis*, as an example below:

|  |
| --- |
| * **Reaction ID:** R\_Generic\_RNA\_llmg\_rRNA\_1\_unmodified\_to\_rRNA\_16S\_unmodified |
| * **Reaction Formula:** llmg\_rRNA\_1\_unmodified\_c -> rRNA\_16S\_unmodified\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_Generic\_RNA\_llmg\_rRNA\_4a\_unmodified\_to\_rRNA\_16S\_unmodified |
| * **Reaction Formula:** llmg\_rRNA\_4a\_unmodified\_c -> rRNA\_16S\_unmodified\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_Generic\_RNA\_llmg\_rRNA\_5a\_unmodified\_to\_rRNA\_16S\_unmodified |
| * **Reaction Formula:** llmg\_rRNA\_5a\_unmodified\_c -> rRNA\_16S\_unmodified\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_Generic\_RNA\_llmg\_rRNA\_6b\_unmodified\_to\_rRNA\_16S\_unmodified |
| * **Reaction Formula:** llmg\_rRNA\_6b\_unmodified\_c -> rRNA\_16S\_unmodified\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_Generic\_RNA\_llmg\_rRNA\_7a\_unmodified\_to\_rRNA\_16S\_unmodified |
| * **Reaction Formula:** llmg\_rRNA\_7a\_unmodified\_c -> rRNA\_16S\_unmodified\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_Generic\_RNA\_llmg\_rRNA\_48a\_unmodified\_to\_rRNA\_16S\_unmodified |
| * **Reaction Formula:** llmg\_rRNA\_48a\_unmodified\_c -> rRNA\_16S\_unmodified\_c |
| * **Catalyst:** - |

## rRNA modification

The formulation of rRNA modification reaction is similar to tRNA modification. We formulated modification reactions only for 16S rRNA and 23S rRNA as modifications to 5S rRNA seems to be infrequent in bacteria. We predicted that in *L. lactis* there are ten modification sites in 16S rRNA and 24 modification events in 23S rRNA. Therefore, we formulated ten reactions for 16S rRNA modification and 24 reactions for 23S rRNA modification. We show the first and the last reaction of 16S rRNA modification below.

The first reaction:

|  |
| --- |
| * **Reaction ID:** R\_rRNA\_modification\_rRNA\_16S\_modified\_1 |
| * **Reaction Formula:** rRNA\_16S\_unmodified\_c -> rRNA\_16S\_modified\_1\_c |
| * **Catalyst:** RNAmod\_llmg\_2518\_Enzyme\_c |

The last reaction:

|  |
| --- |
| * **Reaction ID:** R\_rRNA\_modification\_rRNA\_16S |
| * **Reaction Formula:** 2 M\_amet\_c + rRNA\_16S\_modified\_9\_c -> 2 M\_h\_c + 2 M\_ahcys\_c + rRNA\_16S\_c |
| * **Catalyst:** RNAmod\_llmg\_1882\_Enzyme\_c |

## Ribosomal assembly

We divided ribosomal assembly into two steps, i.e., ribosome protein formation and assembly of protein and rRNA. Firstly, we formulated formation of each ribosome protein subunit, which is spontaneous in the model. It should be noted that in *L. lactis* ribosome protein subunit L33 can be coded by three genes, i.e., “llmg\_0098”, “llmg\_0632”, and “llmg\_2390”. We found in the proteome of *L. lactis* Il1403 that only the protein of “L0096” (llmg\_0098) was detected with its abundance in top 5% of the total detected proteome (PMID: 24739216). Therefore, we assumed that only “llmg\_0098” is used for the production of L33. Besides, the coefficient of L7/L12 (llmg\_1208) was assumed to be four according to the fact that L12 presents in four copies in prokaryotic ribosomes (PMID: 15147176).

The formation of 30S ribosomal protein was formulated by adding all the small subunits:

|  |
| --- |
| * **Reaction ID:** R\_ribosome\_30S\_protein |
| * **Reaction Formula:** llmg\_0251\_Monomer\_c + llmg\_0296\_Monomer\_c + llmg\_0932\_Monomer\_c + llmg\_1724\_assumed\_Monomer\_c + llmg\_1921\_Monomer\_c + llmg\_2078\_Monomer\_c + llmg\_2355\_Monomer\_c + llmg\_2356\_Monomer\_c + llmg\_2364\_Monomer\_c + llmg\_2367\_Monomer\_c + llmg\_2370\_Monomer\_c + llmg\_2374\_Monomer\_c + llmg\_2377\_Monomer\_c + llmg\_2379\_Monomer\_c + llmg\_2384\_Monomer\_c + llmg\_2430\_Monomer\_c + llmg\_2473\_Monomer\_c + llmg\_2475\_Monomer\_c + llmg\_2545\_Monomer\_c + llmg\_2557\_Monomer\_c + llmg\_2558\_Monomer\_c -> ribosome\_30S\_protein\_c |
| * **Catalyst:** - |

The formation of 50S ribosomal protein was formulated by adding all the large subunits:

|  |
| --- |
| * **Reaction ID:** R\_ribosome\_50S\_protein |
| * **Reaction Formula:** llmg\_0098\_Monomer\_c + llmg\_0099\_Monomer\_c + llmg\_0145\_Monomer\_c + llmg\_0204\_Monomer\_c + llmg\_0906\_Monomer\_c + llmg\_1207\_Monomer\_c + 4 llmg\_1208\_Monomer\_c + llmg\_1491\_Monomer\_c + llmg\_1493\_Monomer\_c + llmg\_1671\_Monomer\_c + llmg\_1815\_Monomer\_c + llmg\_2029\_Monomer\_c + llmg\_2030\_Monomer\_c + llmg\_2276\_Monomer\_c + llmg\_2277\_Monomer\_c + llmg\_2353\_Monomer\_c + llmg\_2357\_Monomer\_c + llmg\_2362\_Monomer\_c + llmg\_2363\_Monomer\_c + llmg\_2365\_Monomer\_c + llmg\_2366\_Monomer\_c + llmg\_2371\_Monomer\_c + llmg\_2372\_Monomer\_c + llmg\_2373\_Monomer\_c + llmg\_2375\_Monomer\_c + llmg\_2376\_Monomer\_c + llmg\_2378\_Monomer\_c + llmg\_2380\_Monomer\_c + llmg\_2381\_Monomer\_c + llmg\_2382\_Monomer\_c + llmg\_2383\_Monomer\_c + llmg\_2546\_Monomer\_c -> ribosome\_50S\_protein\_c |
| * **Catalyst:** - |

The 70S ribosome is made up of 30S and 50S ribosome, each of which is made up of ribosomal protein and rRNA. For 30S ribosome, it is made up of 16S rRNA and 30S ribosomal protein. We assumed that the formation of 30S ribosome is catalysed by a complex consisting of GTP-binding protein Era (llmg\_0371), Ribosome-binding factor A (RbfA) (llmg\_1791), and 16S rRNA processing protein RimM (llmg\_0936). Besides, the formation of one molecule of 30S ribosome needs one molecule of GTP as energy.

Below is the formation of 30S ribosome catalyst complex:

|  |
| --- |
| * **Reaction ID:** R\_Ribosome\_30S\_Enzyme |
| * **Reaction Formula:** llmg\_0371\_Monomer\_c + llmg\_1791\_Monomer\_c + llmg\_0936\_2mer\_c -> Ribosome\_30S\_Enzyme\_c |
| * **Catalyst:** - |

Therefore, we formulated the formation of 30S ribosome below:

|  |
| --- |
| * **Reaction ID:** R\_ribosome\_30S |
| * **Reaction Formula:** M\_gtp\_c + rRNA\_16S\_c + ribosome\_30S\_protein\_c -> M\_pi\_c + M\_h\_c + M\_gdp\_c + ribosome\_30S\_c |
| * **Catalyst:** Ribosome\_30S\_Enzyme\_c |

For 50S ribosome, it is made up of 5S rRNA, 23S rRNA, and 50S ribosomal protein. We assumed that the formation of 50S ribosome is catalysed by trigger factor (llmg\_0519) without energy cost.

Below is the formation of 50S ribosome catalyst:

|  |
| --- |
| * **Reaction ID:** R\_Ribosome\_50S\_Enzyme |
| * **Reaction Formula:** llmg\_0519\_2mer\_c -> Ribosome\_50S\_Enzyme\_c |
| * **Catalyst:** - |

Therefore, we formulated the formation of 50S ribosome below:

|  |
| --- |
| * **Reaction ID:** R\_ribosome\_50S |
| * **Reaction Formula:** rRNA\_5S\_unmodified\_c + rRNA\_23S\_c + ribosome\_50S\_protein\_c -> ribosome\_50S\_c |
| * **Catalyst:** Ribosome\_50S\_Enzyme\_c |

At last, we formulated the formation of 70S ribosome, which is spontaneous in the model:

|  |
| --- |
| * **Reaction ID:** R\_ribosome\_70S |
| * **Reaction Formula:** ribosome\_30S\_c + ribosome\_50S\_c -> ribosome\_70S\_c |
| * **Catalyst:** - |

## tRNA charging

We modelled tRNA charging process for 1) charging tRNA with amino acids and 2) generating charged codon readers for the codons that do not have corresponding tRNA.

In *L. lactis*, there are 35 types of tRNA with various anticodons. It should be noted that both tRNA-Met-CAU and tRNA-fMet-CAU have the anticodon “CAU” but they function differently, i.e., the latter transfers Met for the first position while the former for the others. Firstly, we formulated tRNA charging reactions for all the 35 types of tRNA, and the reactions were based on the KEGG database. Generally, tRNA charging is a multiple-step process and catalysed by tRNA synthetase. We lumped the multiple steps in the model and used one enzymatic reaction to represent this process. It should be noted that we did not include tRNA as a substrate in the reaction as otherwise it will be returned as a product in translation process and thereby ruin the network. Instead, we coupled generic tRNA renaming reactions with tRNA charging reactions to account for the need of tRNA in simulations. Below is an example for charging tRNA with the anticodon “GCA”:

|  |
| --- |
| * **Reaction ID:** R\_A\_charging\_tRNA\_Ala\_GCA |
| * **Reaction Formula:** M\_ala\_\_L\_c + M\_atp\_c + M\_h2o\_c -> A\_charged\_in\_tRNA\_Ala\_GCA\_c + M\_amp\_c + M\_ppi\_c + M\_h\_c |
| * **Catalyst:** tRNA\_Synthetase\_A\_GCA\_Enzyme\_c |

When charging tRNA-fMet, which is responsible for transferring methionine for the first codon, we integrated formyltransferase reaction into the charging reaction. Thereby, the integrated reaction was assumed to be catalysed by a complex consisting of methionyl-tRNA synthetase and methionyl-tRNA formyltransferase. Below is an example:

|  |
| --- |
| * **Reaction ID:** R\_fM\_charging\_tRNA\_fMet\_AUG |
| * **Reaction Formula:** M\_atp\_c + 2 M\_h2o\_c + M\_10fthf\_c + M\_met\_\_L\_c -> M\_amp\_c + M\_ppi\_c + 2 M\_h\_c + M\_thf\_c + fM\_charged\_in\_tRNA\_fMet\_AUG\_c |
| * **Catalyst:** tRNA\_Synthetase\_Complex\_fM\_AUG\_Enzyme\_c |

Notably, there is no glutamine-tRNA synthetase gene in *L. lactis* genome, so tRNA-gln is charged in a different way. Simply, glutamate is bound to tRNA-Gln by glutamyl-tRNA synthetase, and then amidotransferase converts the bound glutamate to glutamine. We lumped the two steps in the model and obtained one reaction catalysed by a complex consisting of glutamyl-tRNA synthetase and amidotransferase. Below is an example:

|  |
| --- |
| * **Reaction ID:** R\_Q\_charging\_tRNA\_Gln\_CAA |
| * **Reaction Formula:** 2 M\_atp\_c + M\_h2o\_c + M\_gln\_\_L\_c -> M\_amp\_c + M\_ppi\_c + M\_h\_c + M\_pi\_c + Q\_charged\_in\_tRNA\_Gln\_CAA\_c + M\_adp\_c |
| * **Catalyst:** tRNA\_Synthetase\_Complex\_Q\_CAA\_Enzyme\_c |

There are 64 (43=64) possible codons, so there should be 61 (excluding three stop codons) types of codon readers (i.e., tRNA) if each one follows base pair rules. However, there are 35 types of tRNA in *L. lactis* which only correspond to 34 anticodons, so we generated codon readers for the remaining codons (61-34=27). The biological explanation behind this is the Wobble hypothesis, which holds that the 5” base on the anticodon (or the 3” base on the codon) could have flexible choices of paired base. Therefore, we formulated reaction for the formation of codon readers for the remaining 27 codons. Considering the fact that a codon could be read by more than one tRNA when following the Wobble hypothesis, e.g., tRNA with anticodon “GCC” and “UCC” can both read the codon “GGU”, we assumed that only the tRNA with the lowest cost in its production was used. For simplification, we just made a spontaneous reaction to obtain a charged codon reader directly from a charged tRNA. We show below tRNA-Gly as an example. There are four codons can be translated to glycine, but only two codon readers in *L. lactis*, i.e., tRNA with anticodon “UCC” and that with “GCC”. When following base pair rules, they can read the codon “GGA” and “GGC”, respectively, and the remaining codons are “GGG” and “GGU”, which can be read by both of them following Wobble hypothesis. However, we only used the tRNA with anticodon “UCC” to produce the codon readers for both the codon “GGG” and “GGU” as it is shorter in length than the tRNA with anticodon “GCC”. Below are reactions:

|  |
| --- |
| * **Reaction ID:** R\_G\_charged\_in\_tRNA\_Gly\_GGG |
| * **Reaction Formula:** G\_charged\_in\_tRNA\_Gly\_GGA\_c -> G\_charged\_in\_tRNA\_Gly\_GGG\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_G\_charged\_in\_tRNA\_Gly\_GGU |
| * **Reaction Formula:** G\_charged\_in\_tRNA\_Gly\_GGA\_c -> G\_charged\_in\_tRNA\_Gly\_GGU\_c |
| * **Catalyst:** - |

## Translation

We divided translation into three steps in the model, i.e., 1) translation initiation, 2) translation elongation, and 3) translation termination.

### Translation initiation

Translation initiation represents the process that 70S ribosome and the first methionine bind a TU. We assumed that an intact TU or sub-TU is translated without being divided into transcripts of single genes. Therefore, translation of any gene in a TU needs the intact TU as a template, but translation initiation starts from the start codon of the gene of interest. Besides, we assumed that 70S ribosome is used for coupling constraints only in translation elongation step although it should be used during the whole translation process. In the model, translation initiation is catalysed by a complex, which is made up of translation initiation factor 1 (llmg\_2358), 2 (llmg\_1792), and 3 (llmg\_2031).

|  |
| --- |
| * **Reaction ID:** R\_Translation\_Initiation\_Complex\_Enzyme |
| * **Reaction Formula:** llmg\_2358\_4mer\_c + llmg\_1792\_assumed\_Monomer\_c + llmg\_2031\_assumed\_Monomer\_c -> Translation\_Initiation\_Complex\_Enzyme\_c |
| * **Catalyst:** - |

It should be noted that we did not include TU as a substrate in the translation initiation reaction as otherwise it will be returned as a product in the translation termination reaction and thereby ruin the network. Instead, we coupled TU or Sub-TU production reactions with translation initiation reactions to account for the need of mRNA in simulations. We use the gene “llmg\_0145” as an example to show translation initiation reaction below:

|  |
| --- |
| * **Reaction ID:** R\_translation\_initiation\_TU1G1R\_202\_llmg\_0145 |
| * **Reaction Formula:** M\_h2o\_c + M\_gtp\_c + fM\_charged\_in\_tRNA\_fMet\_AUG\_c -> M\_h\_c + M\_pi\_c + M\_gdp\_c + TU1G1R\_202\_llmg\_0145\_initiated\_c |
| * **Catalyst:** Translation\_Initiation\_Complex\_Enzyme\_c |

### Translation elongation

Translation elongation represents the process that amino acids are added till the ribosome reaches a stop codon. There are two steps in this process: binding of a charged tRNA to the ribosome and translocation of the charged tRNA. The former is catalysed by EF-Tu (llmg\_2050) together with EF-Ts (llmg\_2429), and the latter by EF-G (llmg\_2556). We formulated reactions representing the binding and translocation of tRNA, resulting in a pool of amino acids charged to bound and translocated tRNA, from which any amino acids can be used directly for translation elongation. In other words, we only changed the state of amino acids instead of adding tRNA to the reactions.

Firstly, we formulated reactions representing the binding of a charged tRNA to the ribosome, catalysed by a complex consisting of EF-Tu and EF-Ts.

|  |
| --- |
| * **Reaction ID:** R\_EF\_Tu\_EF\_Ts\_Complex\_Enzyme |
| * **Reaction Formula:** llmg\_2050\_Monomer\_c + llmg\_2429\_Monomer\_c -> EF\_Tu\_EF\_Ts\_Complex\_Enzyme\_c |
| * **Catalyst:** - |

This was done for all the charged tRNA, but we only show “tRNA\_Ala-GCA\_charged” as an example below:

|  |
| --- |
| * **Reaction ID:** R\_Activate\_A\_charged\_in\_tRNA\_Ala\_GCA |
| * **Reaction Formula:** M\_h2o\_c + A\_charged\_in\_tRNA\_Ala\_GCA\_c + M\_gtp\_c -> M\_h\_c + A\_charged\_in\_tRNA\_Ala\_GCA\_activated\_c + M\_pi\_c + M\_gdp\_c |
| * **Catalyst:** EF\_Tu\_EF\_Ts\_Complex\_Enzyme\_c |

Secondly, we formulated reactions representing the translocation of the bound tRNA (i.e., activated tRNA in the model), catalysed by EF-G.

|  |
| --- |
| * **Reaction ID:** R\_EF\_G\_Enzyme |
| * **Reaction Formula:** llmg\_2556\_Monomer\_c -> EF\_G\_Enzyme\_c |
| * **Catalyst:** - |

The activated tRNA is then converted to an elongated state (representing translocation), and the product can be directly used in translation elongation process (still using “tRNA\_Ala-GCA\_charged” as an example):

|  |
| --- |
| * **Reaction ID:** R\_Elongate\_A\_charged\_in\_tRNA\_Ala\_GCA |
| * **Reaction Formula:** M\_h2o\_c + M\_gtp\_c + A\_charged\_in\_tRNA\_Ala\_GCA\_activated\_c -> M\_h\_c + M\_pi\_c + M\_gdp\_c + A\_charged\_in\_tRNA\_Ala\_GCA\_elongated\_c |
| * **Catalyst:** EF\_G\_Enzyme\_c |

As a result, we generated a pool of tRNA prepared for adding amino acids in the model. The pool contains 61 charged tRNA which can read all the possible codons in nucleotide sequences. Besides, all the tRNA in the pool are already in a “bound” and “translocated” state, meaning that no catalyst or energy is needed when using them to add amino acids. Then we formulated reactions to add the remaining amino acids based on nucleotide sequences. The reaction uses the initiated product produced in translation initiation process and produces an elongated product. Note that 70S ribosome is used in this step as catalyst, and there is no tRNA involved in the reaction. We still use the gene “llmg\_0145” as an example below:

|  |
| --- |
| * **Reaction ID:** R\_translation\_elongation\_TU1G1R\_202\_llmg\_0145 |
| * **Reaction Formula:** 2 A\_charged\_in\_tRNA\_Ala\_GCA\_elongated\_c + 2 A\_charged\_in\_tRNA\_Ala\_GCU\_elongated\_c + F\_charged\_in\_tRNA\_Phe\_UUC\_elongated\_c + 3 G\_charged\_in\_tRNA\_Gly\_GGA\_elongated\_c + 2 H\_charged\_in\_tRNA\_His\_CAC\_elongated\_c + 5 K\_charged\_in\_tRNA\_Lys\_AAA\_elongated\_c + K\_charged\_in\_tRNA\_Lys\_AAG\_elongated\_c + 2 L\_charged\_in\_tRNA\_Leu\_CUU\_elongated\_c + M\_charged\_in\_tRNA\_Met\_AUG\_elongated\_c + N\_charged\_in\_tRNA\_Asn\_AAC\_elongated\_c + P\_charged\_in\_tRNA\_Pro\_CCA\_elongated\_c + Q\_charged\_in\_tRNA\_Gln\_CAA\_elongated\_c + R\_charged\_in\_tRNA\_Arg\_CGC\_elongated\_c + 9 R\_charged\_in\_tRNA\_Arg\_CGU\_elongated\_c + S\_charged\_in\_tRNA\_Ser\_AGC\_elongated\_c + S\_charged\_in\_tRNA\_Ser\_UCA\_elongated\_c + S\_charged\_in\_tRNA\_Ser\_UCU\_elongated\_c + T\_charged\_in\_tRNA\_Thr\_ACA\_elongated\_c + 4 T\_charged\_in\_tRNA\_Thr\_ACU\_elongated\_c + V\_charged\_in\_tRNA\_Val\_GUC\_elongated\_c + V\_charged\_in\_tRNA\_Val\_GUU\_elongated\_c + Y\_charged\_in\_tRNA\_Tyr\_UAC\_elongated\_c + TU1G1R\_202\_llmg\_0145\_initiated\_c -> 43 M\_h2o\_c + TU1G1R\_202\_llmg\_0145\_elongated\_c |
| * **Catalyst:** ribosome\_70S\_c |

### Translation termination

Translation termination represents the process that the ribosome reaches a stop codon and then peptide synthesis is terminated with the release of ribosome and peptide from mRNA. There are four key factors involved in translation termination process, i.e., RF1 (llmg\_0557), RF2 (llmg\_1547), RF3 (llmg\_0368), and ribosome recycling factor Rrf (llmg\_2284). Considering the fact that RF1 recognises the stop codons “UAA” and “UAG” while RF2 recognises “UAA” and “UGA”, i.e., both RF1 and RF2 can read “UAA”, we assumed that only RF1 recognises “UAA” as it needs less cost. Then adding RF3 and Rrf, we formulated reactions for the formation of translation termination complex.

|  |
| --- |
| * **Reaction ID:** R\_RFRrf\_UAA\_UAG\_Enzyme |
| * **Reaction Formula:** llmg\_0557\_Monomer\_c + llmg\_0368\_Monomer\_c + llmg\_2284\_Monomer\_c -> RFRrf\_UAA\_UAG\_Enzyme\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_RFRrf\_UGA\_Enzyme |
| * **Reaction Formula:** llmg\_0368\_Monomer\_c + llmg\_2284\_Monomer\_c + llmg\_1547\_Monomer\_c -> RFRrf\_UGA\_Enzyme\_c |
| * **Catalyst:** - |

In the model, translation termination reaction uses the elongated product produced in translation elongation process and produces a nascent peptide with a TU returned. The reaction is catalysed by one of the two translation termination complexes, depending on the stop codon of the gene of interest. We still use the gene “llmg\_0145”, which ends with the stop codon “UAA”, as an example below:

|  |
| --- |
| * **Reaction ID:** R\_translation\_termination\_TU1G1R\_202\_llmg\_0145 |
| * **Reaction Formula:** M\_h2o\_c + M\_gtp\_c + TU1G1R\_202\_llmg\_0145\_elongated\_c -> M\_h\_c + M\_pi\_c + M\_gdp\_c + llmg\_0145\_nascent\_c |
| * **Catalyst:** RFRrf\_UAA\_UAG\_Enzyme\_c |

## Protein maturation

We modelled protein maturation process to remove formyl-group for all the translated peptides, and methionyl-group for the case that the mature protein does not have methionine at the first position. It depends on the N-terminus prediction whether methionine should be removed for a peptide, e.g., if it is predicted that a mature protein does not start with methionine, methionine should be removed from the nascent peptide.

The cleavage of the formyl-group is catalysed by peptide deformylase Def (llmg\_0532). We formulated the formation of Def in the model:

|  |
| --- |
| * **Reaction ID:** R\_Peptide\_Deformylase\_Enzyme |
| * **Reaction Formula:** llmg\_0532\_Monomer\_c -> Peptide\_Deformylase\_Enzyme\_c |
| * **Catalyst:** - |

The cleavage of the methionyl-group is catalysed by methionine aminopeptidase Map (llmg\_0577). We formulated the formation of Map in the model:

|  |
| --- |
| * **Reaction ID:** R\_Methionine\_Aminopeptidase\_Enzyme |
| * **Reaction Formula:** llmg\_0577\_2mer\_c -> Methionine\_Aminopeptidase\_Enzyme\_c |
| * **Catalyst:** - |

There are two cases in protein maturation process, i.e., 1) only formyl-group should be removed, and 2) both formyl-group and methionyl-group should be removed. Below, we show two examples for the two cases.

1. Only formyl-group should be removed (using “llmg\_0145” as an example)

|  |
| --- |
| * **Reaction ID:** R\_deformylase\_llmg\_0145 |
| * **Reaction Formula:** M\_h2o\_c + llmg\_0145\_nascent\_c -> M\_for\_c + llmg\_0145\_c |
| * **Catalyst:** Peptide\_Deformylase\_Enzyme\_c |

1. Both formyl-group and methionyl-group should be removed (using “llmg\_0073” as an example)

Firstly, the formyl-group is removed:

|  |
| --- |
| * **Reaction ID:** R\_deformylase\_llmg\_0073 |
| * **Reaction Formula:** M\_h2o\_c + llmg\_0073\_nascent\_c -> M\_for\_c + llmg\_0073\_for\_M\_excision\_c |
| * **Catalyst:** Peptide\_Deformylase\_Enzyme\_c |

Secondly, the methionyl-group is removed:

|  |
| --- |
| * **Reaction ID:** R\_Met\_aminopeptidase\_llmg\_0073 |
| * **Reaction Formula:** M\_h2o\_c + llmg\_0073\_for\_M\_excision\_c -> M\_met\_\_L\_c + llmg\_0073\_c |
| * **Catalyst:** Methionine\_Aminopeptidase\_Enzyme\_c |

## Protein assembly

We modelled protein assembly process for protein folding based on protein stoichiometry information. We assumed that this process is spontaneous. The protein stoichiometry determines how many nascent peptides should be used for a functional protein. We assumed the protein without stoichiometry information to be a monomer. Below are examples for monomer, assumed monomer, and oligomer:

|  |
| --- |
| * **Reaction ID:** R\_llmg\_0013\_Monomer |
| * **Reaction Formula:** llmg\_0013\_c -> llmg\_0013\_Monomer\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_llmg\_0022\_assumed\_Monomer |
| * **Reaction Formula:** llmg\_0022\_c -> llmg\_0022\_assumed\_Monomer\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_llmg\_0020\_4mer |
| * **Reaction Formula:** 4 llmg\_0020\_c -> llmg\_0020\_4mer\_c |
| * **Catalyst:** - |

## Enzyme formation

In the model, we formulated enzyme formation reactions to represent the production of functional enzymes catalysing other reactions in metabolism and protein expression process. We formulated such reactions for all the catalysts no matter whether it is a complex containing subunits, and all of them are spontaneous. We have shown some examples above and will also show some below. If a catalyst is made up of only one folded protein, e.g., acetylglutamate kinase, the reaction is:

|  |
| --- |
| * **Reaction ID:** R\_M\_ACGK\_Enzyme |
| * **Reaction Formula:** llmg\_1755\_6mer\_c -> M\_ACGK\_Enzyme\_c |
| * **Catalyst:** - |

If a catalyst is a complex and made up of several subunits, e.g., glycogen phosphorylase, the reaction is:

|  |
| --- |
| * **Reaction ID:** R\_M\_GLCP\_Enzyme |
| * **Reaction Formula:** llmg\_1869\_Monomer\_c + llmg\_1871\_Monomer\_c -> M\_GLCP\_Enzyme\_c |
| * **Catalyst:** - |

Note that we did not include the formation of 70S ribosome in this process. Instead, we formulated such a reaction in another process, i.e., ribosomal assembly, as it is a special catalyst made up of not only proteins but also rRNA.

## Protein degradation

Protein degradation process in the model includes two steps, i.e., breakdown of catalyst (complex) into sub-proteins (subunits) and proteolysis of protein into amino acids. Considering the fact that we formulated formation reaction even for the catalyst with only one type of protein (e.g., llmg\_2556\_Monomer\_c -> EF\_G\_Enzyme\_c), all the catalysts in the model should go through the two steps for degradation. We assumed that the first step is spontaneous while the second step is not. In *L. lactis*, several proteases have been identified, including ClpP (PMID: 9987112), FtsH (PMID: 8000529), and HtrA (PMID: 10712686). As ClpP seems to be central in both the proteolysis of misfolded proteins and adjusting regulatory proteins in the cell (PMID: 16628446), we assumed that protein degradation is catalysed by Clp protease complex, which is made up of one proteolytic subunit ClpP (llmg\_0638) and three regulatory subunits ClpB (llmg\_0986), ClpC (llmg\_0615), and ClpE (llmg\_0528) in *L. lactis* (PMID: 10846223). Regarding energy cost, it was estimated that approximately 0.25 to 1.0 ATP is needed for the recycling of each amino acid during protein degradation (PMID: 26575626). Therefore, we adopted a low value (i.e., 0.25 ATP per amino acid) for degradation of each protein as not all the proteins are degraded with costing energy in reality.

Below is the formation of Clp protease complex:

|  |
| --- |
| * **Reaction ID:** R\_Clp\_Protease\_Complex\_Enzyme |
| * **Reaction Formula:** llmg\_0638\_14mer\_c + llmg\_0986\_3mer\_c + llmg\_0615\_10mer\_c + llmg\_0528\_10mer\_c -> Clp\_Protease\_Complex\_Enzyme\_c |
| * **Catalyst:** - |

Here we use “tRNA\_Synthetase\_F\_UUC\_Enzyme\_c” as an example to show two steps of protein degradation.

The first step is breakdown of the complex:

|  |
| --- |
| * **Reaction ID:** R\_tRNA\_Synthetase\_F\_UUC\_Enzyme\_degradation |
| * **Reaction Formula:** tRNA\_Synthetase\_F\_UUC\_Enzyme\_c -> llmg\_2195\_2mer\_degradation\_c + llmg\_2196\_2mer\_degradation\_c |
| * **Catalyst:** - |

The second step is proteolysis of sub-proteins into amino acids:

|  |
| --- |
| * **Reaction ID:** R\_llmg\_2195\_2mer\_degradation |
| * **Reaction Formula:** 398 M\_atp\_c + 1990 M\_h2o\_c + llmg\_2195\_2mer\_degradation\_c -> 136 M\_ala\_\_L\_c + 1990 M\_h\_c + 398 M\_pi\_c + 6 M\_cys\_\_L\_c + 90 M\_asp\_\_L\_c + 140 M\_glu\_\_L\_c + 48 M\_phe\_\_L\_c + 108 M\_gly\_c + 20 M\_his\_\_L\_c + 112 M\_ile\_\_L\_c + 100 M\_lys\_\_L\_c + 146 M\_leu\_\_L\_c + 398 M\_adp\_c + 48 M\_gln\_\_L\_c + 8 M\_trp\_\_L\_c + 56 M\_arg\_\_L\_c + 82 M\_asn\_\_L\_c + 40 M\_tyr\_\_L\_c + 92 M\_ser\_\_L\_c + 44 M\_met\_\_L\_c + 68 M\_pro\_\_L\_c + 94 M\_thr\_\_L\_c + 156 M\_val\_\_L\_c |
| * **Catalyst:** Clp\_Protease\_Complex\_Enzyme\_c |

|  |
| --- |
| * **Reaction ID:** R\_llmg\_2196\_2mer\_degradation |
| * **Reaction Formula:** 172 M\_atp\_c + 860 M\_h2o\_c + llmg\_2196\_2mer\_degradation\_c -> 38 M\_ala\_\_L\_c + 860 M\_h\_c + 172 M\_pi\_c + 8 M\_cys\_\_L\_c + 48 M\_asp\_\_L\_c + 62 M\_glu\_\_L\_c + 34 M\_phe\_\_L\_c + 58 M\_gly\_c + 20 M\_his\_\_L\_c + 36 M\_ile\_\_L\_c + 44 M\_lys\_\_L\_c + 66 M\_leu\_\_L\_c + 172 M\_adp\_c + 26 M\_gln\_\_L\_c + 4 M\_trp\_\_L\_c + 42 M\_arg\_\_L\_c + 26 M\_asn\_\_L\_c + 18 M\_tyr\_\_L\_c + 28 M\_ser\_\_L\_c + 32 M\_met\_\_L\_c + 20 M\_pro\_\_L\_c + 42 M\_thr\_\_L\_c + 38 M\_val\_\_L\_c |
| * **Catalyst:** Clp\_Protease\_Complex\_Enzyme\_c |

Regarding the special catalyst, 70S ribosome, we described its degradation in this process. The first step is also the breakdown into rRNA and ribosomal proteins:

|  |
| --- |
| * **Reaction ID:** R\_ribosome\_70S\_degradation |
| * **Reaction Formula:** ribosome\_70S\_c -> rRNA\_5S\_unmodified\_c + rRNA\_16S\_c + rRNA\_23S\_c + ribosome\_30S\_protein\_degraded\_c + ribosome\_50S\_protein\_degraded\_c |
| * **Catalyst:** - |

Then, the second step is breakdown of 30S ribosomal protein and 50S ribosomal protein into sub-proteins. The degraded rRNA will never be degraded further as they seem to be stable. Below are the reactions:

|  |
| --- |
| * **Reaction ID:** R\_ribosome\_30S\_protein\_degraded\_degradation |
| * **Reaction Formula:** ribosome\_30S\_protein\_degraded\_c -> llmg\_0251\_Monomer\_degradation\_c + llmg\_0296\_Monomer\_degradation\_c + llmg\_0932\_Monomer\_degradation\_c + llmg\_1724\_assumed\_Monomer\_degradation\_c + llmg\_1921\_Monomer\_degradation\_c + llmg\_2078\_Monomer\_degradation\_c + llmg\_2355\_Monomer\_degradation\_c + llmg\_2356\_Monomer\_degradation\_c + llmg\_2364\_Monomer\_degradation\_c + llmg\_2367\_Monomer\_degradation\_c + llmg\_2370\_Monomer\_degradation\_c + llmg\_2374\_Monomer\_degradation\_c + llmg\_2377\_Monomer\_degradation\_c + llmg\_2379\_Monomer\_degradation\_c + llmg\_2384\_Monomer\_degradation\_c + llmg\_2430\_Monomer\_degradation\_c + llmg\_2473\_Monomer\_degradation\_c + llmg\_2475\_Monomer\_degradation\_c + llmg\_2545\_Monomer\_degradation\_c + llmg\_2557\_Monomer\_degradation\_c + llmg\_2558\_Monomer\_degradation\_c |
| * **Catalyst:** - |

|  |
| --- |
| * **Reaction ID:** R\_ribosome\_50S\_protein\_degraded\_degradation |
| * **Reaction Formula:** ribosome\_50S\_protein\_degraded\_c -> llmg\_0098\_Monomer\_degradation\_c + llmg\_0099\_Monomer\_degradation\_c + llmg\_0145\_Monomer\_degradation\_c + llmg\_0204\_Monomer\_degradation\_c + llmg\_0906\_Monomer\_degradation\_c + llmg\_1207\_Monomer\_degradation\_c + 4 llmg\_1208\_Monomer\_degradation\_c + llmg\_1491\_Monomer\_degradation\_c + llmg\_1493\_Monomer\_degradation\_c + llmg\_1671\_Monomer\_degradation\_c + llmg\_1815\_Monomer\_degradation\_c + llmg\_2029\_Monomer\_degradation\_c + llmg\_2030\_Monomer\_degradation\_c + llmg\_2276\_Monomer\_degradation\_c + llmg\_2277\_Monomer\_degradation\_c + llmg\_2353\_Monomer\_degradation\_c + llmg\_2357\_Monomer\_degradation\_c + llmg\_2362\_Monomer\_degradation\_c + llmg\_2363\_Monomer\_degradation\_c + llmg\_2365\_Monomer\_degradation\_c + llmg\_2366\_Monomer\_degradation\_c + llmg\_2371\_Monomer\_degradation\_c + llmg\_2372\_Monomer\_degradation\_c + llmg\_2373\_Monomer\_degradation\_c + llmg\_2375\_Monomer\_degradation\_c + llmg\_2376\_Monomer\_degradation\_c + llmg\_2378\_Monomer\_degradation\_c + llmg\_2380\_Monomer\_degradation\_c + llmg\_2381\_Monomer\_degradation\_c + llmg\_2382\_Monomer\_degradation\_c + llmg\_2383\_Monomer\_degradation\_c + llmg\_2546\_Monomer\_degradation\_c |
| * **Catalyst:** - |

At last, these sub-proteins are degraded into amino acids, which are similar to the proteolysis reactions of other sub-proteins and will not be shown here.

## Enzyme dilution

We formulated enzyme dilution reactions in the model to represent the dilution of functional enzymes to daughter cells during cell division. In other words, these reactions partly represent protein composition in a cell, which varies as a function of growth rate. We formulated such reactions for all the catalysts but not for the subunits. Below is an example:

|  |
| --- |
| * **Reaction ID:** R\_dilution\_Clp\_Protease\_Complex\_Enzyme |
| * **Reaction Formula:** Clp\_Protease\_Complex\_Enzyme\_c -> |
| * **Catalyst:** - |

Besides, dilution of 70S ribosome was also included in this process:

|  |
| --- |
| * **Reaction ID:** R\_dilution\_ribosome\_70S |
| * **Reaction Formula:** ribosome\_70S\_c -> |
| * **Catalyst:** - |

Note that rRNA will never be diluted in the model as its final state 70S ribosome has been diluted.

## RNA dilution

We also formulated RNA dilution reactions in the model to represent the dilution of RNA to daughter cells during cell division, which partly represent a growth rate-dependent RNA composition in a cell. We formulated such reactions only for the TUs used for translation and generic tRNA. Dilution of rRNA is considered in the dilution of 70S ribosome. Below is an example of TU dilution:

|  |
| --- |
| * **Reaction ID:** R\_dilution\_mRNA\_TU1G1R\_100 |
| * **Reaction Formula:** TU1G1R\_100\_c -> |
| * **Catalyst:** - |

Below is an example of generic tRNA dilution:

|  |
| --- |
| * **Reaction ID:** R\_dilution\_generic\_tRNA\_Ala\_GCA |
| * **Reaction Formula:** tRNA\_Ala\_GCA\_c -> |
| * **Catalyst:** - |

# Formulation of other reactions

There are some other reactions should be formulated in the model, including synthesis of unmodeled protein and a biomass dilution reaction with growth-associated maintenance (GAM) involved.

## Adding an unmodeled protein

By considering enzyme and RNA dilution reactions, the model is able to account for the dilution of protein and RNA in biomass. But the model does not contain all the proteins in *L. lactis*. An unmodeled protein should be added to represent the total of the remaining proteins. When synthesizing the protein, the rest of proteins and RNA will be considered in the model. We assumed that the unmodeled protein has the same amino acid composition as that in biomass equation used in the M model and the number of the amino acids is 250, which is the median of the proteins in *L. lactis*. Besides, we assumed that the codon of each amino acid is the most frequently used one in *L. lactis* based on the GtRNAdb database (PMID: 26673694).

## Adding a biomass dilution reaction

Considering the fact that the model is able to describe the dilution of RNA and proteins to daughter, we formulated the dilution of other biomass materials. The biomass dilution reaction was derived from the biomass equation in the M model, in which protein and RNA were removed as they were represented in the model by dilution reactions. The unmodeled protein was, however, added in the biomass dilution reaction with a specific stoichiometric coefficient. Besides, we assumed that the unmodeled protein accounts for 42% of the *L. lactis* proteome by mass based on published data (PMID: 24739216) and then determined the stoichiometric coefficient of the unmodeled protein. We assumed that all these materials in the biomass dilution reaction are growth rate independent.

We calculated GAM and non-growth-associated maintenance (NGAM) in *L. lactis*, as described by (PMID: 17062565), using physiological data from different specific growth rates (PMID: 25828364). We calculated that GAM is 34 mmolATP/gCDW and NGAM is 3.7 mmolATP/gCDW/h as can be seen in the figure below. Thereby we used these two values in the M model.

GAM in the proteome constrained model was calculated by removing the parts contributed by protein synthesis. There are two major ATP consumption processes included in the proteome constrained model but not in the M model, i.e., tRNA charging and translation. It was estimated that the tRNA charging process accounts for around 3 to 6 mmol ATP per gram CDW, and translation for around 6 to 10 mmol ATP per gram CDW, but the M model itself has already contained around 18 mmol ATP per gram CDW. Accordingly, GAM is set as 34 + 18 – 12 = 40 mmol/gCDW. NGAM is decreased to 3 mmol/gCDW/h as protein and mRNA turnover are assumed to cost little energy in the model.

|  |
| --- |
| * **Reaction ID:** R\_biomass\_dilution |
| * **Reaction Formula:** 42 M\_atp\_c + 42 M\_h2o\_c + 0.002 M\_nad\_c + 0.0002 M\_coa\_c + 1e-05 M\_thf\_c + 6.1e-05 M\_pg\_LLA\_c + 0.000138 M\_clpn\_LLA\_c + 0.0064 M\_CPS\_LLA\_c + 9.6e-05 M\_d12dg\_LLA\_c + 0.00074 M\_DNA\_LLA\_c + 0.00015 M\_LTAAlaGal\_LLA\_c + 1.3e-05 M\_lyspg\_LLA\_c + 0.119 M\_PG\_c + 0.0002 M\_udcpdp\_c + 1e-05 M\_thmpp\_c + 1.3e-05 M\_m12dg\_LLA\_c + 0.00682909 unmodeled\_protein\_biomass\_c -> 42 M\_h\_c + 42 M\_pi\_c + 42 M\_adp\_c |
| * **Catalyst:** - |

# Constraints

## Collection of turnover rates for metabolic reactions

Turnover rates are used for coupling constraints and thereby should be collected and assigned. The turnover rates of metabolic enzymes, i.e., kcats, were assigned for enzymatic reactions using the GECKO toolbox (10.15252/msb.20167411). For the enzymes without collected kcats, we assumed their kcats to be the median of the collected ones, which is 100 /s. Note that the kcats should be multiplied by the stoichiometry coefficients. Besides, we manually assigned kcat for glucose transporter based on *E. coli* data (PMID: ﻿28755958), which is 180 /s.

## Estimation of catalytic rates for protein expression process

According to a previous study, the equation of ribosomal catalytic rate follows a Michaelis-Menten-type (PMID: ﻿21097934), and we assumed that it is also the case in *L. lactis*. Below we will show how to obtain parameters and estimate catalytic rates for several machineries.

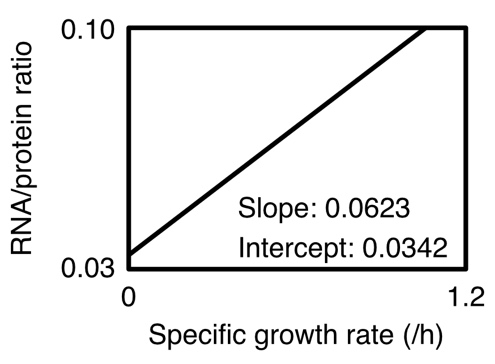
In *E. coli*, the RNA-to-Protein ratio follows the equation (PMID: ﻿21097934):

in which is total cellular RNA mass (g/gCDW), is total cellular protein mass (g/gCDW).

1. and in *L. lactis*

The study (PMID: 7504067) shows that there is a linear correlation between specific growth rate and RNA/protein ratio in *L. lactis* NCDO 712 with the slope being 0.0623 while the intercept 0.0342 as shown in the figure below. In another study (PMID: 10648541) we found that the total protein accounts for 0.45 g/gCDW and total RNA 0.08 g/gCDW at the growth rate of 0.8 /h. With the relationship in the figure, we estimated total protein and RNA for the growth rate of 0.8 /h. We found that if the total protein is 0.45 g/gCDW, then the total RNA is only 0.0378 g/gCDW, which is lower than the reported value in the study (PMID: 10648541). It seems that the RNA measurements were underestimated at that time. Therefore, we adjusted the relationship with a factor of 0.08/0.0378 = 2.1, i.e., both slope and intercept were increased. As a result, the adjusted slope is 0.13 while intercept 0.072. Accordingly, we can estimate and based on equation (1):

Now the equation of ribosomal catalytic rate follows a Michaelis-Menten-type. Therefore, we can calculate the key parameters in the equation if data available for *L. lactis*.



1. Ribosomal catalytic rate

Ribosomal catalytic rate (aa/ribosome/s) can be formulated as

in which is protein synthesis rate (aa/s), is number of ribosomes, is molecular weight of average amino acid (g/mol), is mass of rRNA per ribosome (g/mol\_ribosome), is fraction of rRNA in total RNA.

Let

then

Using the RNA-to-Protein ratio equation, then

in which

Now the equation of ribosomal catalytic rate follows a Michaelis-Menten-type. Therefore, we can calculate ribosomal catalytic rate for *L. lactis*.

Given that is 1700000(1480887) g/mol\_ribosome, is 109(113) g/mol, is 0.86(0.85), we can calculate the value of and then and :

So ribosomal catalytic rate (aa/ribosome/s) is:

1. RNA polymerase catalytic rate:

In *E. coli*, the transcription rate is exactly 3 times the translation rate at all growth rates (PMID: 20413502). Therefore, the RNA polymerase catalytic rate is three times ribosomal catalytic rate (nucleotide/rnap/s):

1. mRNA catalytic rate:

mRNA catalytic rate (protein/mRNA/s) is

in which

Given that (molecular weight of average mRNA nucleotide) is 324 g/mol, is 109(113) g/mol, (fraction of mRNA in total RNA) is 0.02, is 7.7 /h, is 0.072, average protein length is 350, average transcription unit length is 1591, we can calculate

1. tRNA catalytic rate:

tRNA catalytic rate (aa/tRNA/s) is

in which

Given that (molecular weight of average tRNA) is 25000(24606) g/mol, is 109(113) g/mol, (fraction of tRNA in total RNA) is 0.12(0.13), is 7.7 /h, is 0.072, we can calculate

## Constraints

Simulations are usually performed by solving optimization problems, in which one or several reaction rates are maximized or minimized. The simulation will probably move closer to the real phenotype if more constrains are added. Accordingly, imposing constraints on reaction rates is crucial when performing simulations. We discuss below the constraints used in the model.

1. Basic constraints

The proteome constrained model is an expanded version of the M model, and thereby is also able to include two basic constraints of the M model. The first one is SV = 0 at the steady-state condition. When performing simulations, we assume that the concentration of each component constant, meaning that the sum of the rates producing it equals the sum of the rates consuming it. The second constraint is the lower and upper bounds of each reaction rate. This enables a feasible solution space where the simulated state must be located. Besides, the lower and upper bounds make it possible to easily constrain the rates of the reactions of interest, e.g., some exchange reactions.

1. Coupling metabolic reactions and enzymes

In reality, the reaction rate in the cell cannot reach the maximum value that set in the model, i.e., 1000 mmol/gCDW/h, as the cell cannot maintain abundant enough molecules of the enzyme in the limited intracellular space to hold such a high flux. Therefore, there comes the first coupling constraint, i.e., the rate of a metabolic reaction is constrained by the concentration of the enzyme catalysing it:

In which represents the rate (mmol/gCDW/h) of the metabolic reaction, is the turnover rate (/h), and represents the concentration (mmol/gCDW) of the enzyme that catalyses the reaction. The turnover rate has already been collected and can thereby be used directly. Below we focus on the enzyme concentration. At the steady-state condition, the concentration of the enzyme is constant:

in which represents the synthesis rate (mmol/gCDW/h) of the enzyme, represents its degradation rate (mmol/gCDW/h), and represents its dilution rate (mmol/gCDW/h). The equation (22) can then be converted to

in which, is the degradation constant (/h), and is the growth rate (/h). We assume the average degradation constant to be 0.1 in the model. By combining (21) and (24), we can have

Accordingly, the inequation (25) is used for coupling the metabolic reactions and enzyme formation reactions.

For the case that one enzyme catalyses two or more metabolic reactions, inequation can also be generated. For example, if an enzyme can catalyse two metabolic reactions, we can have

There is, however, no such case in the metabolism part as we assigned the man-made enzyme, based on the reaction ID, for each enzymatic reaction in the M model.

Note that there will be some cases that optimal solution cannot be achieved probably due to some low kcats, we assume the kcat with value below 10% lowest kcats (7056 /h) to be 7056 /h.

1. Coupling some protein expression reactions and enzymes

Similar to metabolic reactions, some reactions in protein expression process can also be coupled with the enzyme synthesis reactions using the turnover rate as one of the coupling parameters. These enzymes include tRNA synthetases, methionine aminopeptidase, peptide deformylase, and enzymes in RNA modification process. These reactions follow the same way as metabolic reactions to be coupled with their enzyme formation reactions. We used the median turnover rate in the metabolic part for the protein synthesis part. Moreover, the turnover rates of some enzymes or catalysts depend on the length of the substrates (e.g., the turnover rate of RNA polymerase relies on the length of the TU and thereby differs from TUs with various sequences) and growth rate. Next, we will show coupling constraints about the specific machineries.

1. Coupling transcription reactions and RNA polymerase

In the model, we use the RNA polymerase together with the sigma factor as the complex to catalyse transcription reactions. The turnover rate (nucleotide molecules/complex molecule/s) of RNA polymerase changes with growth rate as shown in the equation (14). Given that the turnover rate is for nucleotides, we can have the turnover rate for TU molecules, which should be divided by the number of nucleotides in the TU:

which has the unit of “TU molecules/complex molecule/s”. By balancing the concentration of the complex of RNA polymerase with sigma factor, we can have:

Then we can have:

1. Coupling mRNA degradation reactions and mRNA degradation complex

In the model, the mRNA degradation complex is made up of degradosome together with oligoribonuclease. Similarly, we can have the constraint:

We assume that the turnover rate of the mRNA degradation complex is calculated as follow:

in which the median turnover rate of metabolic enzymes is 100 /s, average transcription unit length is 1591, and represents the saturation of the machinery, which increases with growth rate and hit the maximum at the maximal growth rate 0.7 /h. The unit is nucleotide molecules/complex molecule/s.

1. Coupling tRNA

As we did not add tRNA as reactants in the tRNA charging and translation reactions, tRNA production reactions (i.e., generic RNA renaming reactions in the model) and tRNA charging reactions should be coupled to enable active fluxes of tRNA. The tRNA can be regarded as the catalyst in tRNA charging reactions, and thereby can be assigned with the catalytic rate or turnover rate, which has been shown in the equation (20). Accordingly, the tRNA charging reaction is coupled with not only the synthesis of tRNA synthetase, but also the production of tRNA. Below show the two coupling constraints:

It should be noted that tRNA is assumed to be stable and thereby do not be degraded. So, there is no degradation factor in the inequation (33).

1. Coupling mRNA

There are two types of TUs in the model, one is for translation (which will be degraded and diluted), and the other is for stable RNA cleavage (which will be cleaved).

For the TUs for translation, we did not add them as reactants in the translation reactions, so TU production and translation reactions should be coupled. The catalytic rate of mRNA has been shown in the equation (17). Theoretically, the TUs can be produced by transcription reactions and stable RNA cleavage reactions. There is, however, no TU from stable RNA cleavage reactions that contain intact genes in the model. So, we can just couple translation initiation reactions and transcription reactions:

Regarding the TUs for stable RNA cleavage, the transcription rate equals to the cleavage rate at the steady-state condition. There is nether degradation nor dilution reaction for them.

1. Coupling translation reactions and ribosome

In addition to mRNA, translation reactions are constrained by the ribosome production. Therefore, translation reaction should also be coupled with ribosome production reaction. In the model, we couple the translation elongation reactions with ribosome. The turnover rate of the 70S ribosome has been shown in the equation (8). Similar to the RNA polymerase case, the coupling constraint can be set as:

1. Coupling protein degradation reactions and protease

We assume that the turnover rate of the protease complex is calculated as follow:

in which the median turnover rate of metabolic enzymes is 100 /s, average protein length is 351, and represents the saturation of the machinery, which increases with growth rate and hit the maximum at the maximal growth rate 0.7 /h. The unit is amino acid molecules/complex molecule/s. Similar to mRNA degradation complex, we can have the constraint:

1. Degradation of mRNA, 70S ribosome and enzymes

The model accounts for degradation of mRNA, 70S ribosome and enzymes. All these degradation reactions are assumed to follow the first order rate constant:

For mRNA degradation, ﻿the median half-lives of mRNA in *L. lactis* were measured to be 5.8, 11.4 and 15.5 min, at 0.80, 0.51 and 0.11 h-1 respectively (23516597). Accordingly, the degradation constant of mRNA is:

In the model, we assume the global protein degradation constant (1/h) to be:

and ribosome degradation constant to be:

Note that the TUs used for translation should be degraded, TUs for cleavage should not be degraded, and sub-TUs should also be degraded. Besides, the unmodeled protein should be degraded. The coupling constraint follows:

1. Dilution of mRNA, ribosome and enzymes

The model also accounts for the dilution of mRNA, tRNA, ribosome and enzymes into daughter cells. The dilution constant rate is the growth rate. Then the dilution rates of them should be:

As there is no reaction for tRNA degradation in the model, all the tRNA will be diluted. So there is no need to formulate coupling. Besides, the TUs used for translation should be diluted, TUs for cleavage should not, but sub-TUs should also be diluted.

1. Constraint of total protein and RNA

In the model, there is a biomass dilution reaction representing a constant biomass composition, which do not account for protein and RNA. This means that most of the components, except protein and RNA, are assumed to be growth rate-independent. On the other hand, the total protein pool is assumed to be fixed at 0.46 g/gCDW, i.e.,

Besides, we do not constrain the total RNA pool in the model, which is determined by the total protein requirement together with catalytic rates of ribosome, tRNA and mRNA for each specific simulation.

1. Constraint of glucose transporter

We assumed in the model that the total concentration of glucose transporter should be not greater than a given value, i.e.,