# Model Generation and Formulation

# Summary

This manual describes how the data were collected, how the model was reconstructed, and finally how the optimization problem was generated as LP file format. We will describe latter each step in details. In the beginning, we write a key for the main Excel and MATLAB files that used to generate and simulate the model.

## Table S1.xlsx: contains all of the collected information.

* **Annotation:** gene names and information.
* **Intron:** Intron information for the genes in the model
* **Length:** The protein length and scanning length in 5’UTR
* **MW:** The molecular mass for the proteins in the model.
* **ElementalComposition:** number of N, O, H, C, P, and S atoms in each protein
* **Subunit:** stoichiometric composition for the metabolic complexes.
* **Chaperone\_Network:** a chaperone and its substrates from chaperone atlas.
* **Folding:** a protein with its chaperones from the Chaperone\_Networksheet.
* **KcatGEKO:** kcat values from GEKO method.
* **Kcat:** kcat values from Avlant’s paper.
* **PlasmaMembrane:** The complexes that are located in the plasma membrane.
* **Mito\_Intermembrane\_proteins:** The complexes that are located in the inner mitochondrial membrane.
* **Cytosolic\_complex:** The complexes that are located in the cytosol.
* **Amino Acids:** free amino acids compositions.
* **Lipid:** Lipid compositions that are used in biomass.
* **Medium:** The composition of YPD, SC, and SD medium.

## tRNA\_modification\_position\_1.xlsx: this contains all of collected information about tRNA.

* **Sequence:** information about cytosolic tRNA sequence, codons, and charged tRNA in the model.
* **Position:** tRNA modification types and their positions.
* **Reaction:** Chemical reaction for each tRNA modification.
* **MitoSequence:** information about mitochondrial tRNA sequence, codons, and charged tRNA in the model.
* **Position\_mito:** tRNA modification types and their positions.

## Two main MATLAB files:

# Laptop Part (model reconstruction) Main\_final\_thermal.m: This script generates the new model. The running time is about six hours. This file extracts the information in the previous Excel files and uses other functions to make the new model. All filling reactions are also merged. The new model is connected and is ready for simulation.

# The generated model is stored as “YeastnextGeneration.mat” and an Excel file “NewYeastnextGeneration.xlsx”

We run this code on normal Laptop.

**Server part (model simulation)**

We run this part on Linux server, because of Soplex solver.

**GlucoseLimit\_GEKO.m:** This script calls the model generator function.

**modelGenerationGeko.m:** This function generates the model as LP format and calls Soplex to solve the problem. In part III, we describe each constraint in the optimization problem.

# Part I: Information Collecting

## **1.1 mRNA**

The RNA polymerase II transcribes a gene from Transcription Start Sites (TSS) to Poly Adenine Sites (PAS), while the ribosome decodes only the nucleotides only in the coding region (CDS). The 5’UTR length can determine the cost of scanning process in translation initiation process. To determine TSS for each gene,we used a SMORE-seq (Parky et al. 2014)**.** For a gene that has no measured TSS, we assume that its 5’UTR length is equal to 50 nts (median of 5’UTR length).For a gene that has TSS inside the CDS, we assume that its 5’UTR length is equal to zero.Finally, we used the yeast genome R64-1-1 from SGD website to determine the start and end position in CDS and intron regions.

### Excel Files

We compiled all these information in the Excel file TableS1.xlsx, in sheets: *annotation*, *Intron*, and *Length*.

### MATLAB Function

We used our MATLAB function*fiveUTRLength(n,m)* to measure the length of 5’UTR for each gene in the *annotation* sheet in TableS1.xlsx file, where *n* and *m* are indexes of the first and last gene in this sheet.

## **1.2 tRNA**

tRNA carries an amino acid corresponding to a current codon in the ribosome. tRNAs are modified with different modification types (El Yacoubi et al. 2012). We downloaded tRNA sequence from GtRNAdb database (Chan & Lowe 2009), and each nucleotide modification type for each tRNA species from MODOMICS database (Machnicka et al. 2013). We curated each modification reaction based on RHEA database and MetaCyC database. Finally, the number of tRNA genes is less than the number of codons. We adopted the codon table was proposed by Roth (Roth 2012).

### Excel Files

We compiled all these information into the Excel file tRNA\_modification\_position\_1.xlsx contains all of this information.  
  
*Matlab*

We used our MATLAB function*addtRNA2Model\_test (model)* to add the tRNA modification and charged reactions

## **1.3 Cytoplasmic Ribosome**

### Subunits

We used the annotation for cytoplasmic ribosomal subunits from Cruz et al (de la Cruz et al. 2015).

### Excel Files

We compiled all ribosomal subunits in the Excel file TableS1.xlsx, in the *annotation* sheet.

MatlabWe used our two MATLAB functions *addRibosome(model)* to include the cytoplasmic ribosomes into the model.

### **1.4 Translation initiation**

Translation initiation in yeast has been reviewed intensively to reveal translation initiation in eukaryotic

(Hinnebusch 2014). In conclusion, the translation initiation factors (eIF4A, -4B, -4E, -4G) need one ATP molecule to bind with mRNA and form the mRNA activation complex. A charged Met-tRNAi binds with eIF2b-GTP to form the Tenery Complex (TC). The ribosomal subunit 40S binds with the translation initiation factors (eIF1, -1A, -3, -5) and TC to form PIC assembly complex that binds with mRNA activation complex to form 43S/mRNA complex. This new complex scans 5’UTR region until finding the first codon AUG to form the 48S complex. Berthelot et al (Berthelot et al. 2004) proposed a random walk algorithm to estimate number of steps that the 5’UTR region scanning, where each step hydrolysis one ATP molecule. Finally, the complex 48S binds with the ribosomal subunit 60S and the translation initiation factor eIF5B-GTP to form the elongation complex and releases all translation initiation factors.

Total cost  
Translation initiation process requires one ATP molecule for binding mRNA with initiation factors, a number of steps in scanning process) ATP molecules, and two GTP molecules for eIF5b and eIF2b.

### Excel Files

We compiled all translation initiation factors in the Excel file TableS1.xlsx, in the *annotation* sheet.

MATLAB FunctionsWe used our MATLAB function *randomScanning(length,.9,0.002)* to estimate estimated these steps. Where length is the length of 5’UTR (From Table S1) and 0.9 is the probability of moving towards AUG and 0.002 is the probability of leaving mRNA and starting again from the beginning.

The MATLAB function *addTranslationFactors* () adds the translation factors to the model.  
 **1.5 Translation Elongation**Translation elongation also is studied intensify in yeast (Dever & Green 2012). In conclusion, the translation elongation factor 1B (eEF1B) binds eEF1A with one GTP molecule to form eEF1A-GTP that binds with charged-tRNA to form eEF1A-GTP-aatRNA. It binds with accepter site (A) of a ribosome. When eEF1A-GTP-aatRNA recognizes the corresponding code, tRNA hydrolyzes GTP molecule, releases eEF1A and accommodates aatRNA into A-site of a ribosome. After that, GTPase eEF2 assists the ribosome to release tRNA. Additionally, yeast has another translation elongation factor (eEF3A) that assists eEF2 in releasing tRNA from the ribosome (Andersen et al. 2006). eEF3 is two ABC-type ATPase (Chakraburtty 2001). Andersen et al (Andersen et al. 2006) proposed that eEF3 fellows the ATP switch model for ABC transporters (Andersen et al. 2006), so eEF3 needs two ATP molecules to release tRNA from the ribosome.

Total costFor each codon (amino acid), two ATP molecules and two GTP molecules are required.

### Excel Files

We compiled all translation initiation factors in the Excel file TableS1.xlsx, in the *annotation* sheet.

MatlabWe used our MATLAB function *elnogation(seq)* to return with a number of ATP and GTP molecules and returns products and substrate on the protein biosynthesis reaction and estimated these steps, where *seq* is a protein sequence without first and last codon.

### **1.6 Translation termination**

We adopted a termination model proposed by Dever and Green (Dever & Green 2012). In conclusion, the translation termination factor (eRF1) binds with GTP form of eRF3 to form the complex eRF1.eRF3-GTP that recognizes the stop codon and binds with the ribosome. The ATPase Rli1 assist in ribosome recycling.

Total costfor each peptide, one ATP molecule and one GTP molecule are required.

### Excel Files

We compiled all translation termination factors in the Excel file TableS1.xlsx, in the *annotation* sheet.

### **1.7 Reaction for Cytosolic Translation**

MatlabWe used our MATLAB function *addProtein2Model\_test (.)* to include a reaction that describes the cost of translated peptide.

## **1.8 Mitochondrial Ribosome**

### Subunit

We collected mitochondrial ribosomal subunits from Woellhaf et al (Woellhaf et al. 2014), and Greber and Ban (Greber & Ban 2016).

### Excel Files

We compiled all ribosomal subunits in the Excel file TableS1.xlsx, in the *annotation* sheet.

MatlabWe used our MATLAB function *addMitoRibosome(model)* that include the mitochondrial ribosome into the model.

### **1.9 Translation initiation in Mitochondrial**

Mitochondrial translation initiation is simpler than the cytoplasmic translation initiation. The GTP-bound mitochondrial translation initiation factor 2 (IFM1) directs the first Met-tRNA to bind with the mitochondrial ribosome.   
  
Total cost  
One GTP molecule is required

### Excel Files

We compiled all translation initiation factors in the Excel file TableS1.xlsx, in the *annotation* sheet.

**1.10 Translation Elongation in Mitochondrial**Mitochondrial translation elongation is similar with bacterial translation elongation. Yeast mitochondria has two translation elongation factors: MEF1 and TUF1.

Total costTwo GTP molecules are required.

### Excel Files

We compiled all translation initiation factors in the Excel file TableS1.xlsx, in the *annotation* sheet.

MatlabWe used our MATLAB function *elnogationMito(seq)* to return with number of ATP and GTP molecules and returns products and substrate on the protein biosynthesis reaction and estimated these steps, where *seq* is a protein sequence without first and last codon.

### **1.11 Translation termination in Mitochondrial**

Total costfor each peptide, one ATP molecule and one GTP molecule are required.

### Excel Files

We compiled all translation initiation factors in the Excel file TableS1.xlsx, in the *annotation* sheet.

### **1.12 Reaction for Mitochondrial Translation**

MatlabWe used our MATLAB function *addMitoProtein2Model*()to include a reaction that describes the cost of a translated peptide into mitochondria.

## **1.13 Proteasome**

### Subunits

We used the annotation for proteasome subunits from Finley et al (Finley et al. 2012).

### Excel Files

We compiled all proteasome subunits in the Excel file TableS1.xlsx, in the *annotation* sheet.

MatlabWe used our two MATLAB function *addProteasome* () that include the proteasome into the model.

### Degradation cost

**1.14 Reaction for protein degradation**MatlabWe used our the MATLAB function *addProteinDegradation2Model* () that include all degradation reactions into the model.

### **1.15 Cytosolic ribosome assembly** Factors Excel Files

We compiled assembly factors in the Excel file TableS1.xlsx, in the *annotation* sheet.

MatlabWe used our the MATLAB function *addAssemblyFactors* () that include the assembly factors in the model into the model.

## **1.16 Chaperones**

### Subunits

We used the annotation for cytosolic chaperones and their substrates from atlas of chaperones (Gong et al. 2009).

### Excel Files

We compiled all proteasome subunits in the Excel file TableS1.xlsx, in the *annotation* sheet and *Folding* sheets.

MatlabWe used our MATLAB function *addChaperone ()* to that include the chaperones into the model.

## **1.17 TOM/TIM complexes**

### Subunits

We used the annotation for TOM/TIM complexes subunits from Fox (Fox 2012).

### Excel Files

We compiled all proteasome subunits in the Excel file TableS1.xlsx, in the *annotation* sheet.

MatlabWe used our three MATLAB functions: *addTOM* (),*addTIM22()*, and *addTIM23()* to that include the TOM/TIM complexes into the model.

## **1.18 Metabolic complexes**

### Subunits

We collected the metabolic subunits from UniProt, MetaCyc, and Protein Data bank databases.

### Excel Files

We compiled all metabolic subunits in the Excel file TableS1.xlsx, in the *subunit* sheet.

MatlabWe used our two MATLAB function *addMetabolicComplex* () that include the metabolic complexes into the model.

## **1.19 Kcats**

We collected Kcat from GEKO methods in the sheet kcat in the Excel file Table S1.xlsx.

**Part II: Model Reconstruction**

## **2.1 New reactions**

Our MATLAB function *Main\_final\_thermal* () will generate the new model. First, the function converts the reversible reactions to two irreversible reactions. After that, the function includes the next reaction types to the model (See the next figures):

1. Translation reaction ( => peptide)

The reaction name has the keyword “\_translation” for cytosolic translation and the keyword “\_translation\_mitochondrion” for translation in mitochondrial

1. Degradation reaction ( peptide => )

The reaction name has the keyword “\_subunit\_degradation” for a subunit degradation

1. Transporting peptide to a compartment

The reaction name has the keyword “\_importing\_cytoplasm\_compartment” for transporting peptide from a cytoplasm to a compartment.

1. Folding reaction ( peptide => folding peptide)

The reaction name has the keyword “\_folding\_compartment” for a peptide folding. The compartment is replaced by a compartment name, such mitochondrion.

1. Reaction for a complex assembly ( folding peptide 1 + … + folding peptide n => complex

The reaction name has the keyword “\_complex\_formation” for assembling a complex from its subunits.

1. Reaction for complex dilution

The reaction name has the keyword “\_complex\_dilution” for a subunit dilution.

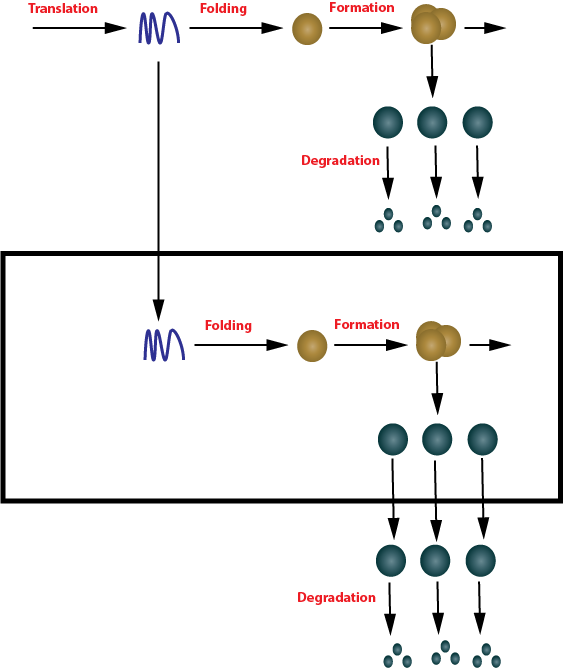
1. Reaction for complex degradation

The reaction name has the keyword “\_complex\_degradation” for a subunit degradation

1. Reaction for transporting a degraded peptide from a compartment to cytosol.

The reaction name has the keyword “\_subunit\_degradation” for a subunit degradation

In addition to these reactions, we added reactions that describe the charging tRNA and un-charging tRNA.



## **2.2 un-modelled protein**

The model does not contain all proteins. To include these proteins, we propose the protein YDRUMD1. This protein has 447 amino acids, where the amino acids composition of this protein is similar to amino acids in the biomass reaction. We assumed that YDRUMD1 has 50 nt in 5’UTR.

**2.3 Biomass reaction**

The biomass reaction is a lumped reaction that describes the concentration of biomass precursors as mmol/gDW (protein precursors are represented by their corresponding amino acids) (Thiele & Palsson 2010). Förster et al (Forster et al. 2003) described how these precursor concentrations were computed from published data. The amino acid concentrations were the major components in the biomass reaction. They determined them from monomer compositions (Oura 1972) and amino acids composition was proposed by Bruinenberg et al (Bruinenberg et al. 1983). These concentrations are conserved to the last model yeast 7.6. Additionally, they computed also the carbohydrates, DNA, and RNA concentrations. Heavner et al (Heavner et al. 2012) used a lipidomics study that analyzes the global yeast lipidome (Ejsing et al. 2009) to curate more lipid species. They developed an “isa” reaction that describes how a lipid class (i.e. Complex Sphingolipid) is formed from its lipid species (i.e. MIPC). Finally, Nookaew et al (Nookaew et al. 2008) computed the biomass compositions when yeast grows in nitrogen-limited condition.

In GEMs, the composition of the biomass reaction does not change with the growth rate. However, when yeast grows faster, it needs more proteins. The increase in protein composition leads to increase in RNA and decrease in carbohydrates and lipid compositions. Additionally, the biomass compositions are also changed from a strain to a strain. Canelas et al (Canelas et al. 2011) measured the biomass composition for the strain CEN.PK 113-7D which is considered a laboratory strain. We collected the biomass composition from their supplementary data and applied basic curve fitting to predict the biomass composition at different growth rates. The proportion of DNA, phosphate, sulfate, and metals is constant at 0.036 g/gDW. For the metals, we divided its ratio to riboflavin, chitin, and heme A, which are biomass components in the model yeast 6.5. The estimated total biomass value is between 0.975 to 1.02 g/gDW.

To make the biomass composition depend on growth rate, we constrained our model to produce the protein ratio (g/gDW) and removed the protein component from the biomass reaction. Additionally, we constrained the new biomass reaction flux (without protein) to equal the corresponding growth rate. For carbohydrates and RNA, we followed the same computation as in Forster et al (Forster et al. 2003), but with different carbohydrates and RNA ratios (g/gDW). To make a lipid composition depend on growth rate, we made a lumped reaction consisting of 64 lipid species (Ejsing et al. 2009), where the data were provided as mol/mol lipid \*100. First, we used  the Yeast Metabolome Database (YMDB) (Ramirez-Gaona et al. 2016) and LIPID MAPS (http://www.lipidmaps.org/) to find the molecular mass for each metabolite. We then computed the molecular mass of lipid as g/mol. Finally, we got the concentration of each metabolite as mmol/gDW **(Table S1),** where the lipid ratio in Table S1 is 0.07g/gDW. We updated these compositions with the lipid ratio at different growth rates.

**MATLAB functions**

The MATLAB function *biomass\_CEN\_PK(μ)* returns the biomass reaction for the strain CEN.PK 113-7D, when growth rate equals *μ*.

The MATLAB function *proteinRatio(μ)* returns the protein ratio for the strain CEN.PK 113-7D, when growth rate equals *μ*.

**2.4 Growth Media**

We compiled the media compositions (SC, YPD, and SD) from Snitkin et al (Snitkin et al. 2008) in sheet media in the Excel file.

**2.5 Converting protein concentration as Molecules/cell**

The protein concentration in the model is expressed as mmol/gDW. We use the next calculations to convert 1 protein molecules to mmol/gDW, where a cell weights 13 pg.

At steady state, protein concentration where and is a flux value in a reaction of a translated protein We convert as Molecules/cell by multiplying [e] by 13\*6.023e8.

**2.6 Protein volume and occupied area**

Erickson (Erickson 2009) assumed that the protein molecule is a sphere. He estimated the radius ) of this sphere as a function of molecular weights where:

We then use to estimate the volume and area occupied of this protein.

**Part III: Model Generation**

We converted our model as LP format (CPLEX format) and used the Soplex solver to solve the problem with high accuracy. The advantage of writing our model as LP format directly is to add constraints directly without any restrictions. Additionally, most solvers read LP format, so we are not restricted to the solvers in RAVEN or COBRA packages. In the following, we describe in detail, how we added each constraint in the model.

**3.1 General Parameters setting [Code Lines 45 - 150 in the file modelGenerationGeko.m]:**

* **Growth rate:** the model is generated for each growth rate independently, so growth must be given to generate a model at this growth rate.
* **Metabolic enzyme catalytic rate (Kcat):** We used an average Kcat (110 per second).
* **Sigma factor: We fitted this parameter at growth rate.** A cell has a specific part for a proteome. A low growth rate can be achieved by a small part of the proteome, so the cell fill the proteome part with more proteins. However, a high growth rate can be achieved with a large part of proteome, so the amount of filled protein will be decreased.
* **Degradation rate (kdeg):** the average of proteins degradation rate is 0.045 per hour, at growth rate = 0.1 (1/h) (Petri’s data).
* **Cytosolic Ribosomal catalytic rate (kcat\_ribo):** 9 aa/S, Shah et al predicted ribosomal elongation rate as 9.5 aa/S (Shah et al. 2013). We can find a formula that depends on the growth rate.
* The rate of protein synthases (Protein/h) = [P] (mu +kdeg) gram aa /h
* The molecular weight of amino acid MW(aa) is 110 g/mol.
* The rate of protein synthases (aa/h) = [P] (mu +kdeg) / MW(aa) mol aa/h
* Number of ribosome = fraction of rRNA \* [RNA] /MW(rRNA) in ribosome.
* The rate of ribosome (aa/h) = = constant [P]/[RNA]
* [P]/[RNA] = a \* growth rate + b, we fitted a = (1/0.37) and b = 0.19.
* **Mitochondrial Ribosomal elongation rate (kcat\_ribo\_mit):** 15 aa/S, as the average value of bacterial and cytosolic ribosome.
* **Protesome catalytic rate:** protesome can degrade 2.7 substrate molecule per min (Peth et al. 2013).
* **Assembly Factor rate:** yeast makes about 2000 per minute (Warner 1999).
* **aa-tRNA synthetases (AARSs) catalytic rate (kcat\_AARS):** 3.5 (1/S). Their kcat vary between 1-6 (1/S)
* **ATP Growth Associated Maintenance (GAM):** 30 mmol/gDW/h. GAM in yeast model equal 59 mmol/gDW/h. Our model uses ATP molecules in protein translation and degradation. We reduced the GAM value to 30 mmol/gDW/h.
* **ATP non-Growth Associated Maintenance (NGAM): 0.5** mmol/gDW/h. NGAM in yeast model equal 1 mmol/gDW/h. Our model uses ATP molecules in protein translation and degradation. We reduced the GAM value to 0.5 mmol/gDW/h.

**3.2 Model preprocessing**

**Biomass reaction [Code Lines 159 - 173 in the file modelGenerationGeko.m]**

In our model, the biomass reaction has no protein ratio and depends on a growth rate. The function biomass\_CEN\_PK(mu) return the biomass reaction. We added this reaction to the model and constrained the upper and lower bound for this reaction to growth rate value.

**Growth media [Code Lines 183 - 200 in the file modelGenerationGeko.m]**

We fixed upper and lower bounds for exchange reaction based on the growth media.

**GAM and NGAM reactions [Code Lines 202 - 212 in the file modelGenerationGeko.m]**

We added the next two reactions in the model for losing energy that represents energy in un- modeled processes such as transcription. We constrained the upper and lower bound for these reactions to growth rate value.

**3.3 Objective function [Code Lines 213 - 228 in the file modelGenerationGeko.m]**

**The objective function in our model is to minimize glucose uptake rate.**

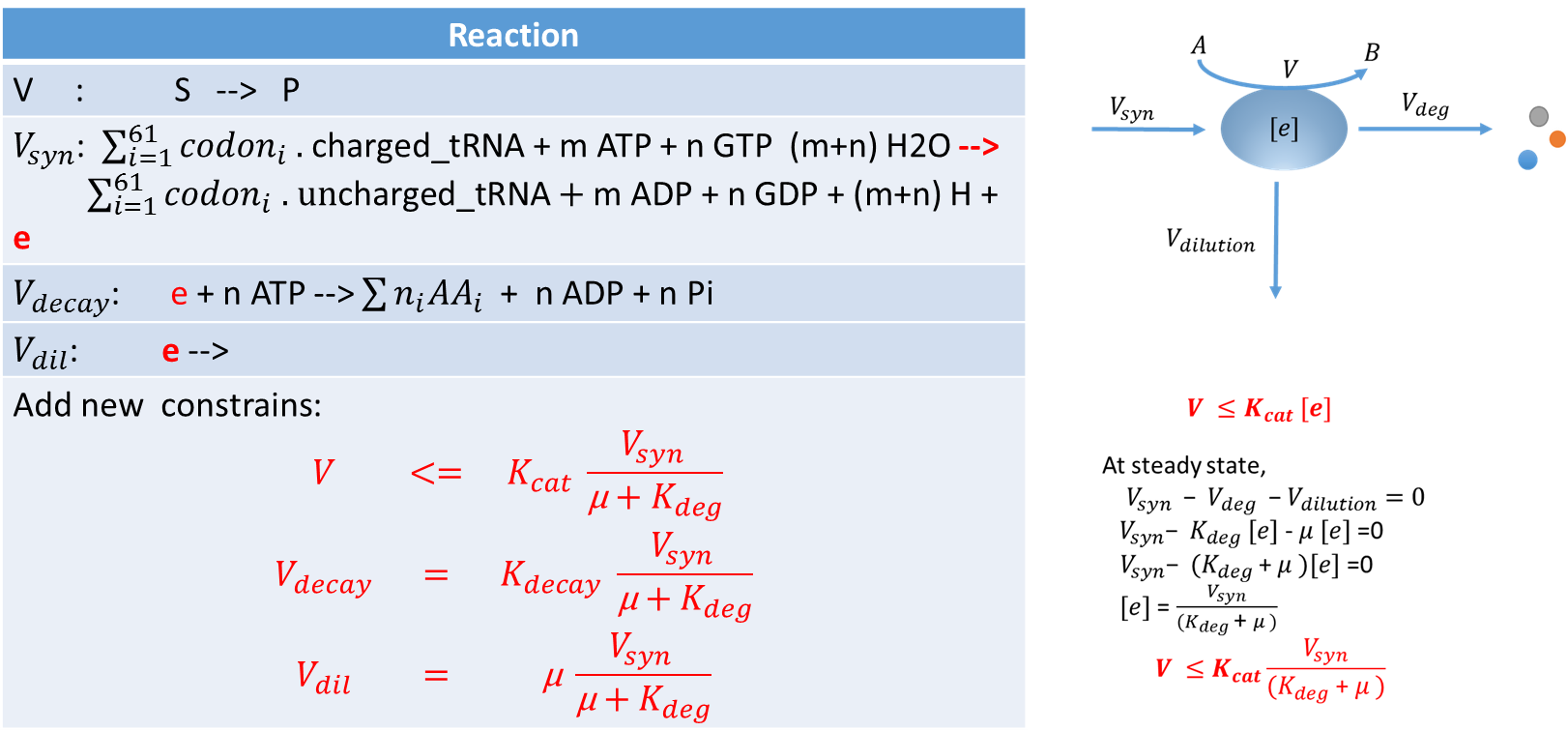
**3.4 Constrains**

**Constraint 1: SV = 0 [Code Lines 229 - 255 in the file modelGenerationGeko.m]**  
The first constraint in our model is the steady-state condition (SV=0). This constraint is the main constraint in metabolic models. Each reaction in the model has the variable name *Xi*, where *i* is the ith reaction (*ith* column in *S*). For example, the reaction number 100 is represented as X100. For each metabolite *j* (where *j* is the *jth* row in S), we print Sji Xi = 0 in the LP file, only and only if *Sji* is not equal zero.

**Constraint 2: Metabolic coupling (V <= kcat [e]) [Code Lines 256 - 474 in the file modelGenerationGeko.m]**

*V* represents a flux value in a metabolic reaction. We looped for all metabolic reactions that are catalyzed by an enzyme concentration [e], where

Each reaction is associated with a complex, where the complex assembly reaction has a keyword “\_complex\_formation”. We search for this reaction describing the enzyme complex formation (Vsyn). In the case, the enzyme can carry out more than one reaction Xi, Xi1, … Xin, we print in LP file



* **Constraint 3: Vdil = mu [e]**

The model needs to dilute with rate equal the growth rate. Therefore we constrain the flux value in dilution reaction of complex biosynthesis is to equal

=0

* **Constraint 4: Vdeg = kdeg [e]**

The model needs to degrade with rate equal the degradation rate. Therefore we constrain the flux value in dilution reaction of complex biosynthesis is to equal

=0

**Constraint 5: Cytosolic Ribosome and translation factors capacity [Code Lines 477 - 646 in the file modelGenerationGeko.m]**

To couple the ribosome biosynthesis reaction with translated proteins, we constrained that each translated protein (including ribosomal proteins) must take a ratio of total amount of ribosomes as:

Where is a flux value in a translated protein reaction . is the elongation translation rate, and is a length of a protein . These equations were adopted from Pavlov and Ehrenberg (Pavlov & Ehrenberg 2013)  
  
We then searched for each reaction ID ending with “\_translation” to find . We added the coupling constraint in the model LP file:

he ribosome biosynthesis, dilution, and degradation reactions are known, so we added the total ribosome constraint three constraints in the model LP file.

**Constraint 6: Translation initiation factors:**

We collected all initiation translation factors in one complex (). We simulated the scanning process, where a ribosome can scan about 10 nt per second. During the model reconstruction, we estimated the total number of steps that are required to reach to the first codon. Therefore, we constrained that each translated protein takes a ratio of total initiation complexes as such as the previous ribosomal constraint.

Where is a flux value in a translated protein reaction . is the scanning rate, and is the number of steps in the scanning process (see section 1.4 translation intialiation).  
  
We then searched for each reaction ID ending with “\_translation” to find . We added the coupling constraint in the model LP file:

he ribosome biosynthesis, dilution, and degradation reactions are known, so we added the total ribosome constraint three constraints in the model LP file.

**Constraint 7: Translation elongation factors (Binding EF1A with GTP and aa-tRNA: 6 (1/S)** Gromadski et al (Gromadski et al. 2007) estimated that the rate of exchange eEF1A with GTP is 6 per second. Additionally, they estimated also the association rate of Phe-tRNA with eEF1A as 5.7 ± 0.3. Therefore, we assumed the rate of overall process is equal to 6 (1/S).  
  
To couple the elongation translation factor (eEF1A) biosynthesis reaction with translated proteins, we constrained that each translated protein (including ribosomal proteins) must take a ratio of eEF1A as:

Where is a flux value in a translated protein reaction . is the catalytic rate of and is a length of a protein .   
  
We then searched for each reaction ID ending with “\_translation” to find . We added the coupling constraint in the model LP file:

he eEF1A biosynthesis, dilution, and degradation reactions are known, so we added the total eEF1A constraint three constraints in the model LP file.

**Constraint 8: eEF2**

To couple the elongation translation factor (eEF2) biosynthesis reaction with translated proteins, we constrained that each translated protein (including ribosomal proteins) must take a ratio of total amount of eEF2as:

Where is a flux value in a translated protein reaction . is the catalytic rate of and is a length of a protein .   
  
We then searched for each reaction ID ending with “\_translation” to find . We added the coupling constraint in the model LP file:

he eEF2 biosynthesis, dilution, and degradation reactions are known, so we added the total eEF2 constraint three constraints in the model LP file.

**Constraint 9: eEF3**

**eEF2 has the rate of elongation rate 10 aa/. We added the next constraint.**To couple the elongation translation factor (eEF3) biosynthesis reaction with translated proteins, we constrained that each translated protein (including ribosomal proteins) must take a ratio of total amount of eEF3as:

Where is a flux value in a translated protein reaction . is the catalytic rate of and is a length of a protein .   
  
We then searched for each reaction ID ending with “\_translation” to find . We added the coupling constraint in the model LP file:

he eEF3 biosynthesis, dilution, and degradation reactions are known, so we added the total eEF3 constraint three constraints in the model LP file.

* **Constraint 10: Translation Termination factors:**

To couple the ribosome biosynthesis reaction with translated proteins, we constrained that each translated protein (including ribosomal proteins) must take a ratio of total amount of eRFas:

Where is a flux value in a translated protein reaction . is the catalytic rate of

We then searched for each reaction ID ending with “\_translation” to find . We added the coupling constraint in the model LP file:

he ribosome biosynthesis, dilution, and degradation reactions are known, so we added the total constraint three constraints in the model LP file.

**Constraint 11: Mitochondrial ribosome capacity: [Code Lines 647 - 703 in the file modelGenerationGeko.m]**

To couple mitochondrial ribosome biosynthesis reaction with the mitochondrial translated proteins (7 proteins for oxidative phospyralation systems and one subunit in mitochondrial ribosome), we constrained that these proteins should be less than all ribosomes in the model (. We used the same constraints as in cytosolic ribosomes (constraint 5). The mitochondrial proteins were translated by reactions “\_translation\_ mitochondrial” to find .

**Constraint 12: Total proteins weight: [Code Lines 704 - 708 in the file modelGenerationGeko.m]**

All concentration of proteins should be equal the protein ratio at the given growth rate. In the previous two steps (cytosolic and mitochondrial ribosome capacity constrained), we searched for andand collected all protein concenertations as in the next equation:

We added this constraint in LP file.

**Constraint 13: Total number of proteins molecules per cell:** all concentration of proteins should be equal the protein ratio at the given growth rate. In the previous two steps (cytosolic and mitochondrial ribosome capacity constrained), we searched for andand collected all protein concenertations as in the next equation:

We added this constraint in LP file.

**Constraint 14: Chaperone capacity [Code Lines 7028 - 804 in the file modelGenerationGeko.m]:** We implemented a chaperon atlas (Gong et al. 2009) as a binary matrix, where a row is a substrate and a column is a cytosolic chaperones (RAC, SSB, SSA, SSE, HSP40, HSP90, PFD, and CCT) (see the file Table S1.xlsx sheet Folding). We added three constraints for each chaperone.

**Constraint 15: Cytoplasmic volume constraint : [Code Lines 825 - 836 in the file modelGenerationGeko.m]**

We collected all cytosolic proteins in the metabolic model and translation machinery proteins. We added this constraint to the generated LP file:

We assumed that all proteins in the model are cytosolic proteins except the proteins that are translated by mitochondrial ribosome.

Yeast cell volume is constant at different growth rate. Yeast volume is about . Cytoplasmic volume is about 50% of whole cell volume. From molecular crowding view, the occupied volume in cytoplasm is about 30% of cytoplasmic volume. Additionally, we assume that 50% of occupied volume is filled by proteins. Therefore, we assumed that cytoplasmic proteins occupy about .

**Constraint 16: Mitochondria volume constraint**

We collected all mitochondrial proteins in the metabolic model and mitochondria ribosome and import machinery. We added this constraint to the generated LP file:

**Constraint 17: Plasma Membrane**

At minimum medium, the model consumes main five nutrients: glucose, ammonia, phosphate, sulfate and oxygen. The oxygen can only be transported by diffusion but the other need transporting proteins. The cell volume is a constant at different growth rate, so we assume the plasma membrane is also a constant at different growth rate.

|  |  |  |
| --- | --- | --- |
| **Metabolite** | **Comment** | **kcat (1/s)** |
| Glucose | Ye et al (Ye et al. 2001) and Kruckeberg et al (Kruckeberg et al. 1999) estimated that the catalytic rate () for HXT7 and HXT2 are equal to 200 and 53 Molecules/Sec, respectively. HXT7 has a paralog HXT4. | 200 |
| Phosphate | Shokrollahzadeh et al (Shokrollahzadeh et al. 2007) reported that Pho84 has the specific activity as 6 µ mol/gDW/min.  We assume that the concentration of Pho84 is about 300,000 molecules/cell, so we can estimate as: | 3 |
| Ammonia | Wacker et al (Wacker et al. 2014) measured the rate of transported ammonia as 30–300 **ions** per s per trimer. | 300 |
| Sulfate | Holt et al (Holt et al. 2017) reported that the is equal to 1.01 nmol/mg cells/min.  From **Petri’s data,** thetotal concentration of mRNA for sulphate transporters is equal to 4.5 molecules/pg. one cell weights 13 pg, so the mRNA concentration is equal to 58.5 molecules/cell. We assume that 1 mRNA molecule is translated to 10,000 protein copies. We assume that all sulfate transporters has about 60,000 molecules/cell. Therefore we can estimate as: | 2.5 |
| H2O |  |  |
| Proton (H+) | Sampedro et al (Sampedro et al. 2007) measured that the kcat for Plasma Membrane H+-ATPase in *KluyVeromyces lactis* is equal to 36 per second. | 36 |

**Constraint 18: Ribosome Assembly constraints [Code Lines 839 - 857 in the file modelGenerationGeko.m]**

Ribosome assembly factors make a new ribosome. We couple a ribosome biosynthesis reaction with ribosome assembly factor reaction as

**Constraint 19: Proteasome capacity [Code Lines 858 - 900 in the file modelGenerationGeko.m]**

To couple the proteasome biosynthesis reaction with degradation proteins, we constrained that all fluxes in the degradation protein reactions (is less than of [Proteasome] as:

Where is the ’s catalytic rate.

**Constraint 20: Un-modelled protein: [Code Lines 901 - 915 in the file modelGenerationGeko.m]**

The model should translate and decay proteins that are not in the model. These protein represents a ratio (

**Constraint 21: TOM complex capacity [Code Lines 916 - 928 in the file modelGenerationGeko.m]:**

**Constraint 22: Over expression protein (such as GFP) [Code Lines 946 - 955 in the file modelGenerationGeko.m]:**

The model contains three fluorescence proteins: GFP, YFP, and mCherry proteins, and hemoglobin complex. We converted number of molecules to mmol/gDW. For GFP example, we added the three constraints.

**References**

Andersen, C.B.F. et al., 2006. Structure of eEF3 and the mechanism of transfer RNA release from the E-site. *Nature*, 443(7112), pp.663–668. Available at: http://www.nature.com/doifinder/10.1038/nature05126.

Berthelot, K. et al., 2004. Dynamics and processivity of 40S ribosome scanning on mRNA in yeast. *Molecular Microbiology*, 51(4), pp.987–1001.

Bruinenberg, P.M., Van Dijken, J.P. & Scheffers, W. a., 1983. A Theoretical Analysis of NADPH Production and Consumption in Yeasts. *Microbiology*, 129(4), pp.953–964.

Canelas, A.B. et al., 2011. An in vivo data-driven framework for classification and quantification of enzyme kinetics and determination of apparent thermodynamic data. *Metabolic Engineering*, 13(3), pp.294–306.

Chakraburtty, K., 2001. Translational regulation by ABC systems. *Research in Microbiology*, 152(3–4), pp.391–399.

Chan, P.P. & Lowe, T.M., 2009. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic acids research*, 37(Database issue), pp.D93-7. Available at: http://nar.oxfordjournals.org/content/37/suppl\_1/D93.abstract [Accessed August 19, 2015].

Dever, T.E. & Green, R., 2012. The elongation, termination, and recycling phases of translation in eukaryotes. *Cold Spring Harbor Perspectives in Biology*, 4(7), pp.1–16.

Ejsing, C.S. et al., 2009. Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proc. Natl. Acad. Sci. {U.S.A.}*, 106(7), pp.2136–2141.

Erickson, H.P., 2009. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biological Procedures Online*, 11(1), pp.32–51.

Finley, D. et al., 2012. The ubiquitin-proteasome system of Saccharomyces cerevisiae. *Genetics*, 192(2), pp.319–360.

Forster, J. et al., 2003. Genome-Scale Reconstruction of the Saccharomyces Cerevisie Metabolic Network. *Genome Research*, (13), pp.244–253.

Fox, T.D., 2012. Mitochondrial protein synthesis, import, and assembly. *Genetics*, 192(4), pp.1203–1234.

Gong, Y. et al., 2009. An atlas of chaperone-protein interactions in Saccharomyces cerevisiae: implications to protein folding pathways in the cell. *Molecular systems biology*, 5(275), p.275. Available at: http://dx.doi.org/10.1038/msb.2009.26.

Greber, B.J. & Ban, N., 2016. Structure and Function of the Mitochondrial Ribosome. *Annual review of biochemistry*, 85(March), pp.103–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27023846.

Gromadski, K.B. et al., 2007. Kinetics of the interactions between yeast elongation factors 1A and 1Bα, guanine nucleotides, and aminoacyl-tRNA. *Journal of Biological Chemistry*, 282(49), pp.35629–35637.

Heavner, B.D. et al., 2012. Yeast 5 - an expanded reconstruction of the Saccharomyces cerevisiae metabolic network. *BMC systems biology*, 6(1), p.55. Available at: http://www.biomedcentral.com/1752-0509/6/55.

Hinnebusch, A.G., 2014. The scanning mechanism of eukaryotic translation initiation. *Annual review of biochemistry*, 83, pp.779–812. Available at: http://www.annualreviews.org/doi/full/10.1146/annurev-biochem-060713-035802?url\_ver=Z39.88-2003&rfr\_id=ori:rid:crossref.org&rfr\_dat=cr\_pub=pubmed&.

Holt, S. et al., 2017. Major sulfonate transporter Soa1 in Saccharomyces cerevisiae and considerable substrate diversity in its fungal family. *Nature Communications*, 8, p.14247. Available at: http://www.nature.com/doifinder/10.1038/ncomms14247.

Kruckeberg, A.L. et al., 1999. Functional expression, quantification and cellular localization of the Hxt2 hexose transporter of Saccharomyces cerevisiae tagged with the green fluorescent protein. *The Biochemical journal*, 339 ( Pt 2, pp.299–307. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1220158&tool=pmcentrez&rendertype=abstract.

de la Cruz, J., Karbstein, K. & Woolford, J.L., 2015. Functions of ribosomal proteins in assembly of eukaryotic ribosomes in vivo. *Annual review of biochemistry*, 84, pp.93–129. Available at: http://www.ncbi.nlm.nih.gov/pubmed/25706898.

Machnicka, M.A. et al., 2013. MODOMICS: A database of RNA modification pathways - 2013 update. *Nucleic Acids Research*, 41(D1), pp.262–267.

Nookaew, I. et al., 2008. The genome-scale metabolic model iIN800 of Saccharomyces cerevisiae and its validation: a scaffold to query lipid metabolism. *BMC systems biology*, 2(1), p.71. Available at: http://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-2-71.

Oura, E., 1972. *The effect of aeration on the growth energetics and biochemical composition of baker’s yeast. 1972*. PhD thesis. Helsinki University, Helsinki, Finland.

Parky, D. et al., 2014. Simultaneous mapping of transcript ends at single-nucleotide resolution and identification of widespread promoter-associated non-coding RNA governed by TATA elements. *Nucleic Acids Research*, 42(6), pp.3736–3749.

Pavlov, M.Y. & Ehrenberg, M., 2013. Optimal control of gene expression for fast proteome adaptation to environmental change. *Proceedings of the National Academy of Sciences*, 110(51), pp.20527–20532. Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1309356110.

Peth, A., Nathan, J.A. & Goldberg, A.L., 2013. The ATP costs and time required to degrade ubiquitinated proteins by the 26 S proteasome. *Journal of Biological Chemistry*, 288(40), pp.29215–29222.

Ramirez-Gaona, M. et al., 2016. YMDB 2.0: a significantly expanded version of the yeast metabolome database. *Nucleic acids research*, 45(November 2016), p.gkw1058. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27899612.

Roth, A.C., 2012. Decoding properties of tRNA leave a detectable signal in codon usage bias. *Bioinformatics*, 28(18), pp.340–348.

Sampedro, G. et al., 2007. Fluorescence Quenching by Nucleotides of the Plasma Membrane H + -ATPase from. , pp.5616–5622.

Shah, P. et al., 2013. Rate-limiting steps in yeast protein translation. *Cell*, 153(7), pp.1589–1601. Available at: http://dx.doi.org/10.1016/j.cell.2013.05.049.

Shokrollahzadeh, S. et al., 2007. Growth kinetics and Pho84 phosphate transporter activity of Saccharomyces cerevisiae under phosphate-limited conditions. *Journal of Industrial Microbiology and Biotechnology*, 34(1), pp.17–25.

Snitkin, E.S. et al., 2008. Model-driven analysis of experimentally determined growth phenotypes for 465 yeast gene deletion mutants under 16 different conditions. *Genome biology*, 9(9), p.R140. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2592718&tool=pmcentrez&rendertype=abstract.

Thiele, I. & Palsson, B.Ø., 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols*, 5(1), pp.93–121. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3125167&tool=pmcentrez&rendertype=abstract.

Wacker, T. et al., 2014. Direct observation of electrogenic NH4(+) transport in ammonium transport (Amt) proteins. *Proc Natl Acad Sci U S A*, 111(27), pp.9995–10000. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24958855.

Warner, J.R., 1999. The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences*, 24(11), pp.437–440.

Woellhaf, M.W. et al., 2014. Import of ribosomal proteins into yeast mitochondria. *biochemistry and cell biology - Biochimie et biologie cellulaire*, 10(April), pp.1–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24943357.

El Yacoubi, B., Bailly, M. & de Crécy-Lagard, V., 2012. Biosynthesis and Function of Posttranscriptional Modifications of Transfer RNAs. *Annual Review of Genetics*, 46(1), pp.69–95. Available at: http://www.annualreviews.org/doi/10.1146/annurev-genet-110711-155641.

Ye, L. et al., 2001. Expression and activity of the Hxt7 high-affinity hexose transporter of Saccharomyces cerevisiae. *Yeast*, 18(13), pp.1257–1267.