Can you predict yoghurt fermentation?

Kristen Michelle Nader
kristenmichelle.nader@student.kuleuven.be

Introduction and Background

Yeast and lactic acid bacteria (LAB) are known to cooperate in fermented beverages and food products such as wine, kefir, cheese and sourdough. It is assumed that there is a mutually beneficial interaction between the species. It is relevant to elucidate these interactions and the flow of metabolites that allow mutualistic growth. The Lactic acid bacteria and yeast coexist during the fermentation of yoghurt where such interactions enables the growth of each of the species in the community. Studying this interaction allows a better understanding of yoghurt fermentation and the effects of the environment on growth. When a specific media is presented, we want to know which metabolites will be cross-fed between the organisms of the community to ensure the growth of each of the organisms and thus to allow fermentation to proceed.

This internship aims to answer this question and replicate the findings revealed in the paper entitled "Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow" [1]. Using Flux Balance analysis and SteadyCom, we try to determine which metabolites are required by individual species to grow and which metabolites can be provided by other species when a community model is instantiated. Ponomarova highlighted a set of amino acids that were experimentally shown to be involved in cross-feeding. We will try to replicate and validate these results using the Constraint-Based Reconstruction and Analysis(COBRA) Toolbox in MATLAB. The results obtained form a basis for validation of a newly developed computational approach: dynamic community flux balance analysis(dcFBA). However, due to time constraints, dcFBA was not covered in this internship.

Yeast cannot metabolise lactose but *Lactococcus lactis* can. This is one element of the cross-feeding; *Lactococcus lactis* relies on yeast for nitrogen (amino acids). In turn, yeast will depend on *Lactococcus lactis* for its ability to metabolise lactose – the carbon source that is supplied during yogurt fermentation. The paper covers a 3-community model but is unclear which elements are being cross- fed between *L.plantarum* and the community. This internship will focus on simulating the interaction between *L.lactis* and *S.cerevisiae*. This document aims to give a detailed explanation of the workflow and references a number of useful tutorials. All scripts used in this internship are attached.

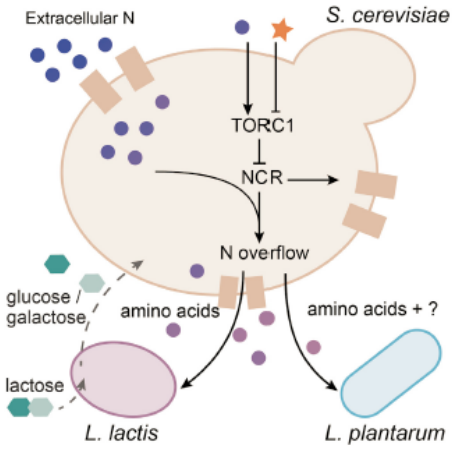

Figure 1: Community model developed by Ponomarova et al.*L.Lactis* metabolises lactose and supplies glucose to *S.cerevisiae* which can then supply amino acids to both LAB species. The elements of cross-feeding between yeast and *L.plantarum* were undetermined.[1]

Requirements

MATLAB is needed to run the attached scripts. The COBRA toolbox can be accessed and installed by cloning from a GitHub repository. Installation steps are given on the opencobra GitHub: opencobra.github.io/cobratoolbox/stable/installation.html

A MATLAB solver is also required. 'glpk' or 'gurobi' are sufficient for running Flux balance analysis, however if SteadyCom is run using one of these solvers, it will take a long time. The Cobra-Toolbox recommends switching to 'ibm_cplex' solver, which reduces runtime signifcantly. This solver can be installed via the following link: www.ibm.com/analytics/cplex-optimizer

Setting up the Media

Included in the paper's supplementary information[1] is a list of the nutrients that are supplied to the community model to determine a cross flow of nutrients - the CDM35 media. Supplied media should lack metabolites that can be exchanged between members of the community, but should contain elements to allow growth of the species.

The paper includes the use of both the CDM35 media where the carbon source is glucose and a CDM35 media where the carbon source was replaced from glucose to lactose. For this internship, the CDM35 media with lactose was used to simulate a community growth model between *Lactococcus lactis* and *Saccharomyces cerevisiae*. To allow a simulation of this cross-feeding, the media must supplied and this will be in the form of an excel file to limit the lower bounds of the fluxes of the reactions; limiting the nutrients that will be uptaken. The BiGG database[2] was used for the COBRA models and also to create the excel file with the metabolites names.

Table 1: BiGG identifiers

BiGG identifier and models used	**Organism[reference]**
iNF517	*Lactococcus lactis* subsp. cremoris MG1363[3]
iMM904	*Saccharomyces cerevisiae* S288C [4]

The excel file should contain in the first column, the list of components from the CDM35 media and the second column, the BiGG identifiers for the base of each reaction. For example, due to

the fact that this simulation will use lactose instead of glucose, it must be explicitly included in the model. See the excel attachment 'CDM35'.

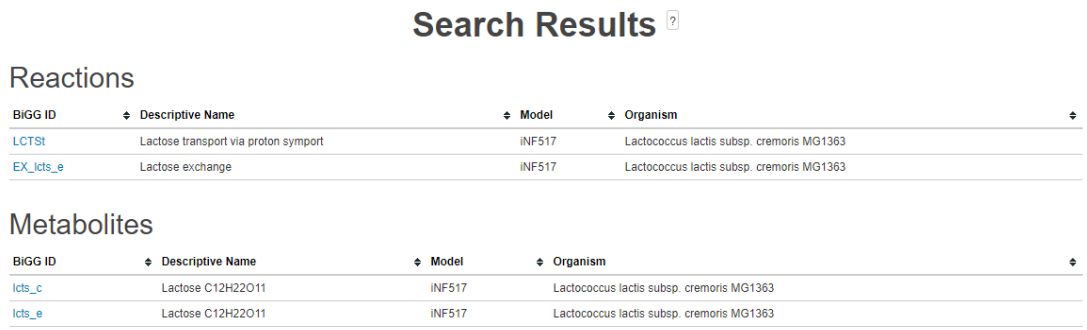

Figure 2: Example case for creating the CDM35 media with lactose as carbon source. '_c' represents lactose present in the cytosol and '_e' represents lactose in the extracellular space. CLicking on the link will reveal reactions that include lactose in each compartment.

In the second column of the excel file, adjacent to "lactose", "lcts" is entered as the BiGG identifier for lactose. This is repeated for all elements of the media. For some compounds, there might not be any identifier, leave it empty.

Flux Balance Analysis

reference tutorial:
https://opencobra.github.io/cobratoolbox/latest/tutorials/tutorialFBA.html

Flux Balance analysis (FBA) is widely used mathematical approach to study the flow of metabolites in metabolic networks. This analysis makes it possible to predict growth rate of individual species (production of biomass) or the rate of production of a metabolite. In order to predict growth rate, FBA determines the rate at which metabolic compounds provided in the surrounding environment is converted into biomass (nucleotides, proteins or lipids)[5].

The first tutorial used from the COBRA toolbox is the Flux Balance Analysis that was done on an *E.coli* model. A modified version will be supplemented for both *Lactococcus lactis* and *Saccharomyces cerevieae*.For any model that is created using this toolbox; the general format should include but is not limited to:

Table 2: The meaning of each Identifier in a standard model[5]

Identifier	**Field Meaning**
S	Stochiometric matrix
mets	Identifiers of the metabolites
b	The coefficients of the constraints of the metabolites(mets)
csense	Indicator in the b vector is lower bounded, upper bounded or hard constraints
rxns	Identifiers for the reactions
lb	the lower bounds of the fluxes of the reactions
ub	the upper bounds of the fluxes of the reactions
c	the linear objective
genes	the list of genes in your model
rules	the Gene-protein-reaction rules in a computer readable format
osenseStr	the objective sense :'max' maximisation or 'min' minimisation

The following equation can be formulated under the assumption of a steady state. It assumes that metabolites cannot accumulate in a specific compartment and the total flux into the network must equal the total outflux. FBA can be considered a multi-compartment model. The equations are given by:

$$Sv = 0 \tag{1}$$

where,

$$v_{i_{min}} \leq v_i \leq v_{i_{max}} \tag{2}$$

where S is the stoichiometric matrix and the constraints reflect uptake of metabolites, v is a vector of the minimum and maximum values of the flux through the model i reactions. Equation 1 represents the contraints by the media.

The goal of FBA is to optimize an objective function Z given by equation 3.

$$Z = c^T v \tag{3}$$

Where Z can be described as any linear combination of v, the fluxes, and c is a vector of weights revealing the contribution of a reaction to the overall objective function Z.

Through linear programming, FBA can be used to solve equation 1 given a set of constraints on v in equation 2 and equation 3 [5].

Each of the models used were derived from the BiGG database (see Table 1). Once the model is read, the media can be set using the "setMedia" function that will take the excel file created in part 1 along with the sheet name.

The simulation will also be run under Anaerobic conditions. This is important for extrapolating and using SteadyCom to simulate the community model.The next two parts will be based on understanding the compounds that are required to allow each of the species to grow **independently** and what will be needed to allow **mutualistic growth** in a community setting.

Lactococcus lactis

Under anaerobic conditions and without the media, there is a growth rate of 0.0426hr^{-1}. However, upon restricting the environment to the CDM35 media, there is no growth of *L.lactis*. This is because there are some compounds that must be supplied by yeast. Table 3 shows the elements that *L.lactis* requires in addition to the CDM35 media to sustain growth. These are saved under the 14x1 cell array "missing".

reference: fba_lactis.m

Ponomarova[1] determined the most abundant amino acids accumulating in the environments by yeast were threonine, glutamine, alanine, glutamate, serine and glycine.

Table 3 displays the missing nutrientswhich consists of a mixture of nucleobases and amino acids that are required by *L.lactis* to sustain growth independently. This is supported by the Ponomarova paper which explicitly states in the STAR methods section entitled 'Metabolic Modeling of Community Cross-Feeding':

> *"Manually curated model of Lactococcus lactis IL1403 was updated in order to reconcile model's in silico growth with L.lactis IL1403 inability to growth in the CDM35 media."*

It is important to see if yeast is able to provide these nutrients while still maintaining growth for itself. This is discussed in the section that follows.

Table 3: Missing metabolites required by L.lactis iNF517 model for growth on CDM35 media. Metabolites may be supplied by other species in the community or supplemented.

Identifier	**Meaning**
EX_ade_e	Adenine exchange
EX_ala__L_e	L-Alanine exchange
EX_asp__L_e	L-Asparagine exchange
EX_cys__L_e	L-Cysteine exchange
EX_glc__D_e	D-Glucose exchange
EX_glu__L_e	L-Glutamate exchange
EX_gly_e	Glycine exchange
EX_gua_e	Guanine exchange
EX_ins_e	Inosine exchange
EX_lys__L_e	L-Lysine exchange
EX_phe__L_e	L-Phenylalanine exchange
EX_ser__L_e	L-Serine exchange
EX_thr__L_e	L-Threonine exchange
EX_ura_e	Uracil exchange

Saccharomyces cerevisiae

The same protocol is performed with the yeast model iMM904 to determine growth rate under different conditions. Table 4 summarises these values:

Table 4: iMM904 model growth on different conditions. Situation 3 and 6 resulted in 0hr-1.

	CDM35	**Glucose**	**Oxygen**	**Growth Rate(hr-1)**
FBAsolution 1	-	+	+	0.5365
FBAsolution 2	+	+	+	0.7074
FBAsolution 3	+	+	-	0
FBAsolution 4	+	-	+	0.4017
FBAsolution 5	+	-	-	0
FBAsolution 6	-	+	-	0

reference: fba_cerevisiae.m

From table 4, yeast appears to grow best when supplied with glucose and oxygen. In FBAsolution 3, yeast is simulated to grow under CDM35+lactose media, supplemented by glucose but under anaerobic conditions. *S.cerevisiae* is capable of performing anaerobic and aerobic respiration. A growth rate of $0\text{hr} - 1$ reveals no growth in this situation - essentially the situation that it will be placed under when simulated in a community model. Trial FBAsolution 6 further removes the CDM35 lactose media, supplements glucose under anaerobic conditions but no growth is observed.

The same script is run for iMM904 to determine which compounds are required to sustain growth independently(similar to the previous subsection on *L.lactis* iNF517 model).

Table 5: Misisng metabolites recquired by the yeast iMM904 model for growth on CDM35 media.

Identifier	**Meaning**
EX_glc__D_e	D-Glucose exchange
EX_o2_e	O2 exchange

Glucose will be supplied by *L.lactis* by lactose metabolism. As stated before, an assumption that

is being made is that this simulation occurs without the presence of oxygen. Table 5 displays some inconsistencies and provides a base for further work.

For a community model to be successful, there must exist some dependencies. *Lactococcus Lactis* and *Saccacromyces cerevisae* must not be able to grow independently and must rely on each other for growth. Table 4 revealed that this model of yeast was unable to grow without the presence of oxygen and therefore represents some problems due to the fact that this simulation of fermentation takes place under an aerobic environmental condition. Oxygen is constrained (the lower bound of the reaction is set to 0 to prevent oxygen from entering the system). This can also be interpreted as the maximum amount of a compound that is allowed for uptake by the species.

In this way, both species are unable to grow independently and must rely on cross-feeding for growth. Suppose glucose is provided by *Lactococcus Lactis*, yeast must be able to provide the missing nutrients for *L.lactis*. This is checked individually to determine which nutrients can be produced by yeast and which causes a loss of biomass(0hr-1) under conditions of oxygen, glucose present and no constraints on media.

In this test, yeast is being simulated independently under the assumption that *L.lactis* will be able to provide glucose. Therefore, the lower bound is set to -10 to allow yeast to uptake glucose from the media.

Table 6: Forcing iMM904 to synthesise the set of metabolites individually to determine which compounds yeast is able to produce and which cannot be produced(will result in stopping production of biomass)

Identifier	**SC growth rate (hr-1)**
EX_ade_e	NaN
EX_ala__L_e	0.2878
EX_asp__L_e	0.2877
EX_cys__L_e	0.2872
EX_glc__D_e	NaN
EX_glu__L_e	0.2877
EX_gly_e	0.2877
EX_gua_e	0.2863
EX_ins_e	NaN
EX_lys__L_e	0.2873
EX_phe__L_e	0.2871
EX_ser__L_e	0.2877
EX_thr__L_e	0.2875
EX_ura_e	NaN

Table 6 reveals that yeast is unable to supply a set of nucleobases and glucose. For this reason, the nucleobases adenine, inosine and uracil must be supplemented into the media manually. This caused some unease due to the fact that the paper used as reference[1] did not mention supplementing the media for *Lactococcus lactis* but for *Lactobacillus planetarium*. The inability to supply glucose can be ignored.

Two core assumptions to this point are:

1. Anaerobic conditions
2. Nucleobases (adenine,inosine and uracil) will be supplemented.

SteadyCom

reference tutorial:
https://opencobra.github.io/cobratoolbox/latest/tutorials/tutorialSteadyCom.html

SteadyCom[6] is used to study a community system between yeast and *L.lactis* at Steady-State. Flux Balance Analysis(FBA) treats a community model as a multi-compartment model. SteadyCom explicitly introduces the biomass variables to describe the relationships between biomass, biomass production rate, growth rate and fluxes- more widely used for modelling community systems. Unlike FBA, SteadyCom aims to create a Joint model and the interactions between the organisms are taken into account to maximise the biomass growth rate for the community under steady state assumptions.

Both models are initialised and the maximum oxygen uptake rate is set to 0 mmol gDW-1 hr-1 (millimoles per gram dry cell weight per hour) to prevent the uptake to oxygen – this is to simulate an anaerobic environment. The nucleobases (adenine, inosine and uracil) are supplemented by setting the maximum uptake rate to 10 mmol gDW-1 hr-1. In the script, this is written as -10, this is due to the fact that exchange reactions are often written in terms of a reactions. Therefore, importing a metabolite from the surrounding environment is a negative flux whereas exporting a metabolite from the cell to the environment is a positive flux. This image below illustrates this concept in MATLAB.

```
% uptake reaction
model = changeRxnBounds(model, 'EX_glc__D_e', -10, 'l');

% force model to produce and export
model = changeRxnBounds(model, 'EX_glc__D_e', 10, 'l');

% prevent uptake
model = changeRxnBounds(model, 'EX_glc__D_e', 0, 'l');
```

Listing 1: MATLAB code to set the lower bound of the glucose exchange reaction for uptake,production and prevent uptake.

Setting a lower bound for the exchange reaction to 10 can be interpreted as the maximum amount of glucose available to the model: 10mmol per hour for 1gdw of the model biomass. Uptake reactions have negative flux.

After restricting the individual models to the CDM35 media(supplemented by a set nucleobases for *L.lactis*), a very important next step is to ensure that the Joint Model is also restricted to the CDM35. This is a major problem that was encountered. Initially, not restricting the Joint Model and running SteadyCom, the biomass of yeast was 1000 but that of L.lactis was 0. This is contradictory to what was revealed in the reference paper. Ponomarova described that when both LAB species and yeast were plated on CDM35 media, *L.lactis* was the most abundant species. The Joint Model was not restricted and was thus allowing nutrients to enter the system. In this situation, yeast achieves optimum growth and does not require *L.lactis* for glucose and therefore is not required to produce and export nutrients(mainly amino acids) to sustain *L.lactis* growth. Exporting nutrients is energetically unfavourable if the relationship is one sided(in the case where yeast supplies nutrients to *L.lactis* but does not receive anything in return). No cross-feeding is occurring in that situation. To ensure no uptake of external nutrients, the exchange reaction for the Joint Model, yeast model and *L.lactis* model were further restricted. Maximum uptake rate for lactose for the Joint model was set to 10 mmol gDW-1 hr-1 and 0 mmol gDW-1 hr-1 for the maximum uptake rate for glucose. Doing this ensured that there was no glucose or oxygen being supplied to the yeast model. Nucleobases should also be supplemented to the Joint model as they would be for the individual *L.lactis* model.

Attached to this documentation will be two scripts entitled "JointModel_trial_x.m".The reason there are 2 scripts is that towards the end of the internship, I was still using the scripts to get them to work. The difference in the scripts is the way I have restricted the Joint Model. JointModel_trial_3.m is where I manually set each reaction lower bound (if the reaction is an exchange reaction for metabolites in the media)to -1 using a for loop. SteadyCom however would throw an error if these were set to 0. Figure 3 displays the output of SteadyCom for a community system of *L.lactis* and yeast under restrictive CDM35 media in the absence of oxygen. The biomass will vary as different parameters are taken for different environmental constraints.

```
Maximum community growth rate: 2.052787
X_LL:   0.774310
X_SC:   0.225690
```

Figure 3: Maxiumum community growth rate and the individual species biomass.

SteadyCom FVA changes the objective function to incorporate the sum of fluxes/biomases as the community growth rate is fixed between 0 and a maximum growth rate. The function'steadycomFVA' was used to derive Figure 4. This graph presents the relative abundance of each species as a fuction of differnt community growth rates. *L.lactis* displays a higher relative abundance with respect to yeast, Figure 3 and Figure 4, which is in line with findings in the Ponomarova paper.

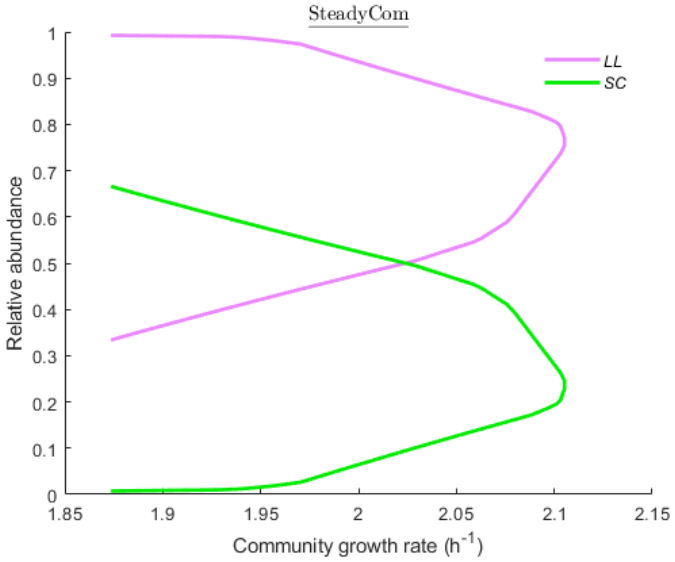

Figure 4: The graph displays the maximum and minimum relative abundance rate for each species in the community model at various community growth rates.

JointModel_trial_2.m successfully sets the lower bounds for the exchange reactions to 0 without a for loop. The difference is in the SteadyCom results.This method will result in vastly different biomass for each of the species in the community model shown in Figure 5.I have included it because I do think it was important to determine which method of setting the bounds would accurately fit this problem.

```
Maximum community growth rate: 0.000000
X_LL:  1000.000000
X_SC:  29.714286
```

Figure 5: Alternative method of setting the lower bounds r*reference: JointModel_trial_2.m*

Conclusion

Ponomarova et al described a flow of amino acids: glutamine, glutamate, proline, threonine, and phenylalanine from yeast to the *L.lactis*. However, this experiment used the different strains and different genome-scale metabolic models (see Table 7) for both yeast and *L.lactis* and this should also be kept in mind when analysing results. The genomic scale metabolic model used to simulate yeast growth was derived to improve the iMM904 model that was used in this internship. Table 7 shows the metabolic models used in the paper and the reference to the model. The difference in the models must be taken into account when analysing discrepancies between results derived from the paper and those from the simulation.

Table 7: Models used in the paper

identifier and models used	**Organism[reference]**
iAO358	*Lactococcus lactis* [3]
iAZ900	*Saccharomyces cerevisiae* [4]

The simulation derived from both FBA and SteadyCom revealed that when *L.lactis* and *S.cerevisiae* exist on CDM35 media ,where the carbon source has been modified to lactose to mimic a yogurt fermentation environment, the elements that are being cross-fed consist of a subset the metabolites in Table 6 (excluding the nucleobases and glucose) and by using the 'printUptakeBoundCom(JointModel, 1)' the uptake bounds can be visualised for both the JointModel and for individual species.

need to incluthese as refernces https://www.patil.embl.de/media/models/ponomarova/iAOP358.bioopt

Zomorrodi, A.R., and Maranas, C.D. (2010). Improving the iMM904 S. cerevisiae metabolic model using essentiality and synthetic lethality data. BMC Syst. Biol. 4, 178.

References

[1] O. Ponomarova, N. Gabrielli, D. C. Sévin, M. Mülleder, K. Zirngibl, K. Bulyha, S. Andrejev, E. Kafkia, A. Typas, U. Sauer, *et al.*, "Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow," *Cell systems*, vol. 5, no. 4, pp. 345–357, 2017.

[2] Z. A. King, J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. O. Palsson, and N. E. Lewis, "Bigg models: A platform for integrating, standardizing and sharing genome-scale models," *Nucleic acids research*, vol. 44, no. D1, pp. D515–D522, 2016.

[3] "Bigg models inf517 -http://bigg.ucsd.edu/models/inf517."

[4] "Bigg models imm904-http://bigg.ucsd.edu/models/imm904."

[5] J. D. Orth, I. Thiele, and B. Ø. Palsson, "What is flux balance analysis?," *Nature biotechnology*, vol. 28, no. 3, pp. 245–248, 2010.

[6] S. H. J. Chan, M. N. Simons, and C. D. Maranas, "Steadycom: Predicting microbial abundances while ensuring community stability," *PLoS computational biology*, vol. 13, no. 5, p. e1005539, 2017.