Principles and Applications of Systems Biology
ChE 411

Exercise 1: FBA Analysis of a Genome-Scale Metabolic Model of E. Coli

By: Sirine Sayagh, Daniel Pacheco Gutierrez, Jedidi Fatma, Anurag Ranjak

Oct 10th, 2018

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

All results were obtained with the help of MATLAB and more specifically the COBRA Toolbox.

What is the COBRA Toolbox?
COBRA stands for COnstraint-Based Reconstruction and Analysis and it is a Python and MATLAB software used for modelling, analysing and predicting a variety of metabolic phenotypes using genome-scale biochemical networks.

What is the standard structure of a genome-scale model in COBRA?
It is a 1x1 structure array with 26 fields.

What is FBA?
FBA stands for Flux Balance Analysis and is a mathematical method for simulating metabolism in genome-scale reconstructions of metabolic networks => in silico simulation; linear programming; genome scale.

We'll be working with a model of E.Coli.

PART I

3. **Understand the inputs required for the FBA problem and the outputs of it.**
optimizeCbModel solves a flux balance analysis problems, by solving LP (Linear Programming) problems of the form: $min_v\ c^T v$ s.t. $Sv = b,\ l \leq v \leq u$.
It takes a model as an input and will output a solution structure with:

- f, the objective value
- v, the reaction rates
- y, dual
- w, the reduced costs
- s, slack
- stat, the solver status in standarized form (-1: no solution, 1: optimal solution, 2: unbounded solution, 0: infeasible)
- origStat, the original status returned by the specific solver.

Find the maximal growth yield for glucose.
This value can be found with the help of the FBASolution structure. The yield will be defined to be the biomass growth rate [h-1] when the substrate uptake is 10mmol/gDW/h. In this case the maximal growth yield of glucose at unlimited oxygen intake (-1000mmol/gDW/h lower bound) is 0.874 h-1.

4. Create a script to (i) define one-by-one the alternative substrates as the only carbon sources, (ii) calculate the maximal growth yield for each of them. Consider for each carbon source an uptake rate of 10 mmol/gDW/h. Perform the simulations in aerobic conditions: oxygen uptake rate of 20 mmol/gDW/h.
The following figure shows the comparison between all the carbon-based substrates and their maximal growth yield under aerobic and anaerobic conditions.

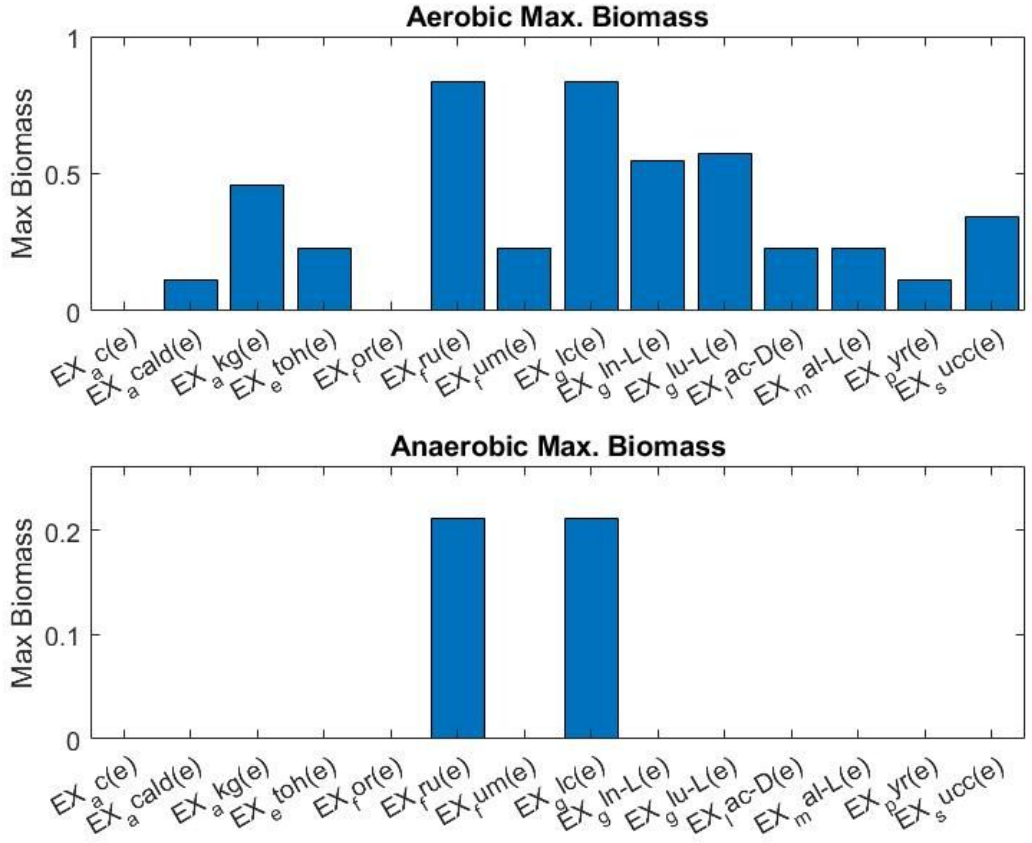

Fig. 1 Comparison Between aerobic and anaerobic Carbon Substrates Maximum Yield

Under aerobic conditions, different substrates can be used for biomass growth. Glucose and fructose are the ones that give the highest yield due to their abundance in carbon and high enthalpy of formation. Some substrates such as acetate give no yield because the excess oxygen inhibits their decomposition due to the production of phosphate, which is an intermediate product in the transformation of acetate to acetyl coa.

5. Repeat Step for Anaerobic Conditions

The anaerobic pathways for biomass growth lead to either acetate, ethanol or formate when the cycle starts at fructose or glucose. Other substrates enter the system directly in the Krebs cycle, which is not activated in the absence of oxygen. As a consequence, only glucose and fructose give biomass growth.

6. Visualize an optimal solution for aerobic and anaerobic growth on 10 mmmol/gDW/h glucose in an ESCHER map and describe the changes in the flux profile qualitatively.

The following figures show the ESCHER maps for aerobic and anaerobic conditions.

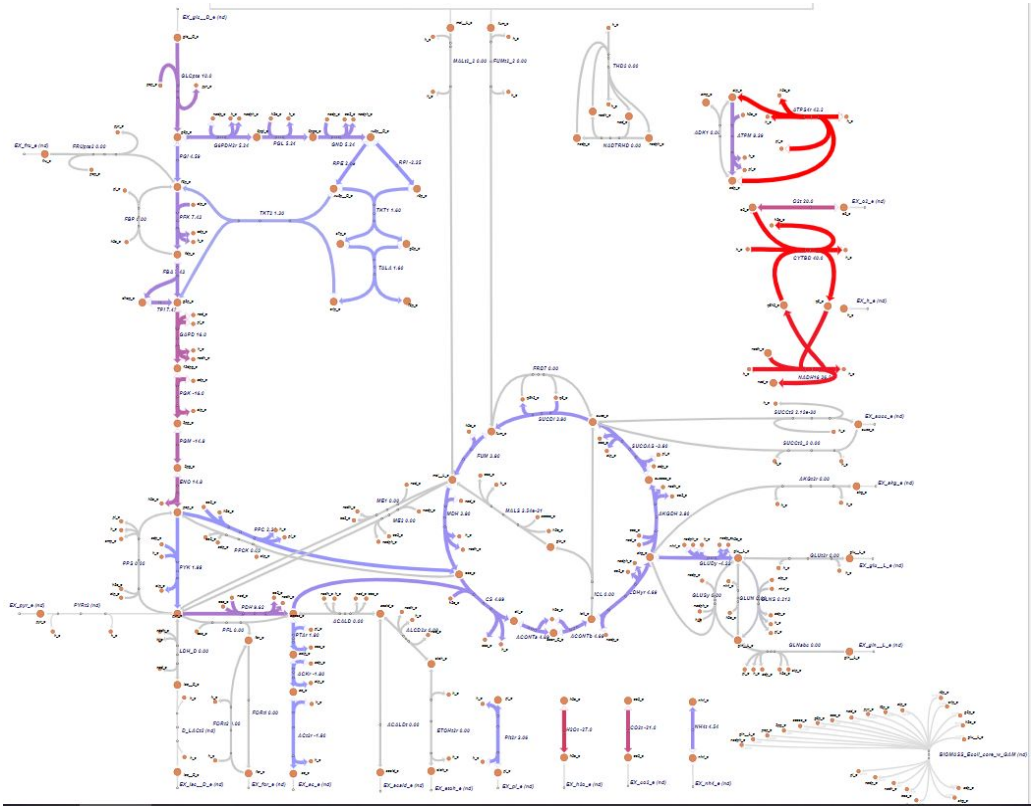

Fig.2 Overview of the ESCHER MAP of E.coli under Aerobic conditions

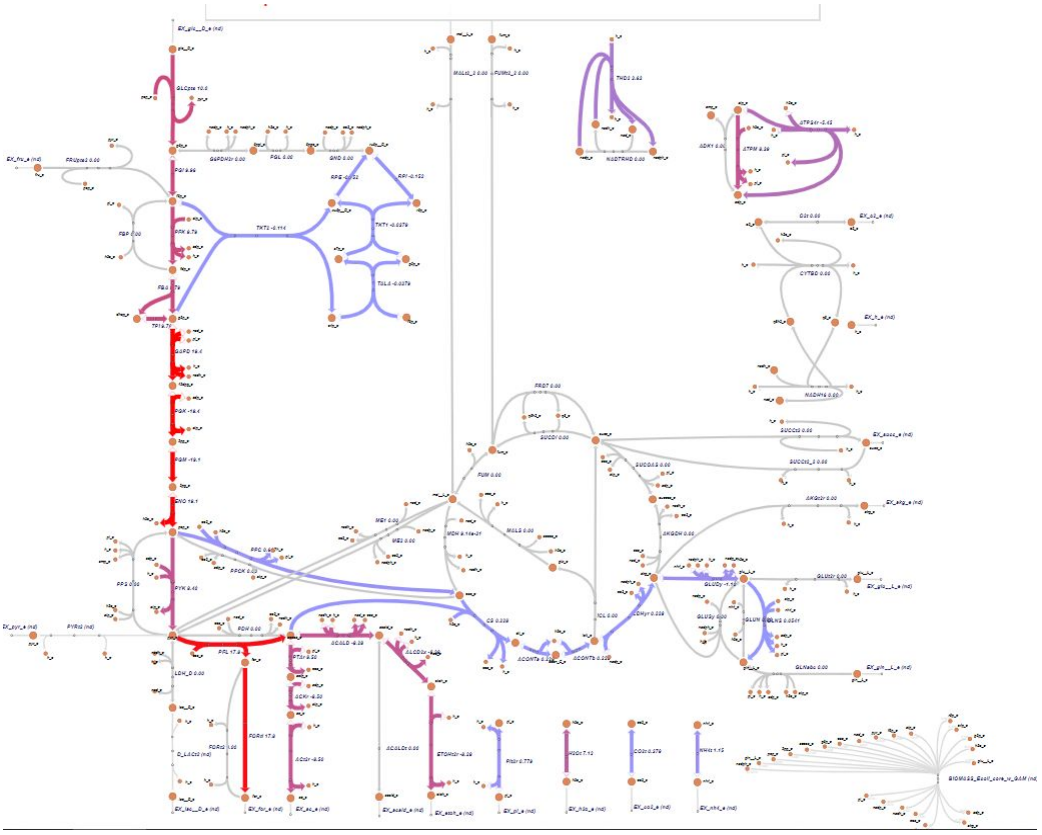

Fig.3 Overview of the ESCHER MAP of E.coli under Anaerobic conditions

In anaerobic conditions, only the lower part of the Krebs cycle is operating. Since no oxygen is available, some of the redox reactions to generate NADH are blocked, shown in the image by the inactive upper part of the Kreb's cycle. Under such conditions, the E.coli metabolism relies mostly on the degradation of D_glucose, generating formate, acetate and ethanol.
We can note that the reactions for the ATP and NADP maintenance are the ones with the higher flux as they have a key role to sustain the metabolism. ~

7. Fill the table below and discuss about the results. What do we learn from the comparison?

Substrates	**Reaction ID**	**Max yield AEROBIC**	**Max. yield ANAEROBIC**
D-Glucose exchange	**EX_glc(e)**	0.8326	0.212
D-Lactate exchange	**EX_lac-D(e)**	0.229	0
D-Fructose exchange	**EX_fru(e)**	0.833	0.212
Ethanol exchange	**EX_etoh(e)**	0.229	0

We learn that D-glucose and D-fructose exchange reactions have the largest yield in both aerobic and anaerobic conditions. Comparing ethanol and lactate for aerobic conditions, it is visible that the lactate yield is slightly larger than the ethanol yield. This might be due to the fact that lactate contains one more carbon, favoring the conversion from pyruvate to acetyl coa opposed to the transition from ethanol to acetyl coa. In anaerobic conditions, only glucose and fructose give any biomass growth. It must be noted that ethanol is one of the final products under anaerobic conditions, but it is not used as an uptake substrate. D-lactate is not a final product nor a possible substrate under anaerobic conditions as shown on the Escher map and the table.

Part 2 : Essential genes analysis in both aerobic and anaerobic conditions

Single gene knockout is a genetic method to study gene function by knocking out a gene from the organism. The gene is made inoperative and the effect of genes in biomass production under aerobic and anaerobic conditions is computed.
All carbon reactions are set to zero, the glucose uptake is set to -10 mmol/gDw/h and the oxygen uptake to -20mmol/gDw/h. The condition is that the gene is set essential if the maximum growth yield variation is superior or equal to 90 %.

Table2. Essential genes analysis in both aerobic and anaerobic conditions among alternative substrates

Substrates	Reaction ID	Num. essential genes Aerobic	Num. essential genes Anaerobic
D-glucose exchange	EX_glc(e)	3	4
D-lactate exchange	EX_lac-D(e)	16	0
L-Glutamine exchange	EX_gln-L(e)	13	0
L-Glutamate exchange	EX_glu-L(e)	5	0
D-fructose exchange	EX_fru(e)	4	5
Pyruvate exchange	EX_pyr(e)	8	0
Acetate exchange	EX_ac(e)	0	0
Ethanol exchange	EX_etoh(e)	16	0

The reaction with the higher number of essential genes are the ethanol exchange and the L-glutamine exchange and the reactions with the smaller number of genes exchanges are the D-glucose exchange and D-fructose exchange. The number of genes can be associated with the number of crucial enzymes or catalysts needed for their decomposition. Therefore, glucose and fructose have less essential genes because there are multiple pathways for their decomposition, contrary to other substrates such as glutamine, ethanol and lactate. It must be noted that these results do not include those genes that the Toolbox was not able to compute and gave NaN for the biomass. These genes are potentially essential as well but were not included in this model. Additionally, each substrate uptake was only lower-bounded at -10mmol/gDW/h, and the oxygen uptake for aerobic conditions was forced to be -20mmol/gDW/h.

Part 3. Understanding the FVA solution by performing Flux Variability Analysis

Flux variability Analysis (FVA) is a computational method to evaluate upper and lower bounds of all steady state reaction fluxes needed to optimize biomass production.

Table3. Minimum and maximum flux through a set of reaction for aerobic and anaerobic conditions

Reaction that undergo a change	**Reaction ID**	**Aerobic Min Flux (mmol/gDW/h)**	**Aerobic Max Flux (mmol/gDW/h)**	**Anaerobic Min Flux (mmol/gDW/h)**	**Anaerobic Max Flux (mmol/gDW/h)**
Acetate exchange	EX_ace(e)	0	8.795	0	8.803
CO2 exchange	EX_co2(e)	9.258	24.398	-4.125	17.303
D-lactase exchange	EX_lac-D(e)	0	5.101	0	10.192
Aconitase (half-reaction B, isocitrate hydrolase)	ACONTb	0.719	10.595	0.183	1.573
Malate dehydrogenase	MDH	-7.089	16.254	-5.274	1.223

In order to produce biomass, several reaction may occur. We can see from the table that acetate and D-lactase are not really important for cell growth. Very small amount is produced under aerobic and anaerobic conditions. That means that the cell can grow and produce biomass based on other reactions without acetate or D-lactase production.

For the Co_2 exchange reaction, Co_2 is produced under aerobic conditions for both minimum and maximum fluxes. But is also consumed in anaerobic conditions (-7.9901 mmol/gDW/h) due to oxygen absence. This seems not logical because during cell respiration, co2 can not be used to produce biomass.

For Aconitase exchange ; Aconitase is an enzyme that catalyzes the reaction. it is involved especially in the Krebs cycle to achieve the stereospecific isomerization of citrate to isocitrate via cis-aconitate. It’s realized for aerobic and anaerobic conditions.

For the Malate dehydrogenase exchange ; Malate dehydrogenase is an enzyme (catalyst) in many metabolic pathways especially for Krebs cycle. It’s consumed under aerobic and anaerobic conditions. The catalyst says about the reversibility of the reaction. This reaction is reversible because this enzyme can both be produced or consumed. It depends on the enthalpy of the reaction.

PART IV

In this part we performed robustness analysis on the effect of successively glucose, succinate and pyruvate uptakes given with a maximum oxygen uptake of 20mmol/gDW/h. We then plotted the phenotypic phase plane for each cases and determined the shadow prices with respect of each substrate for the different metabolic phenotypes observed in the phase planes.

I-ROBUSTNESS ANALYSIS
The graphs below represent the growth yield as a function of respectively glucose, succinate and pyruvate.

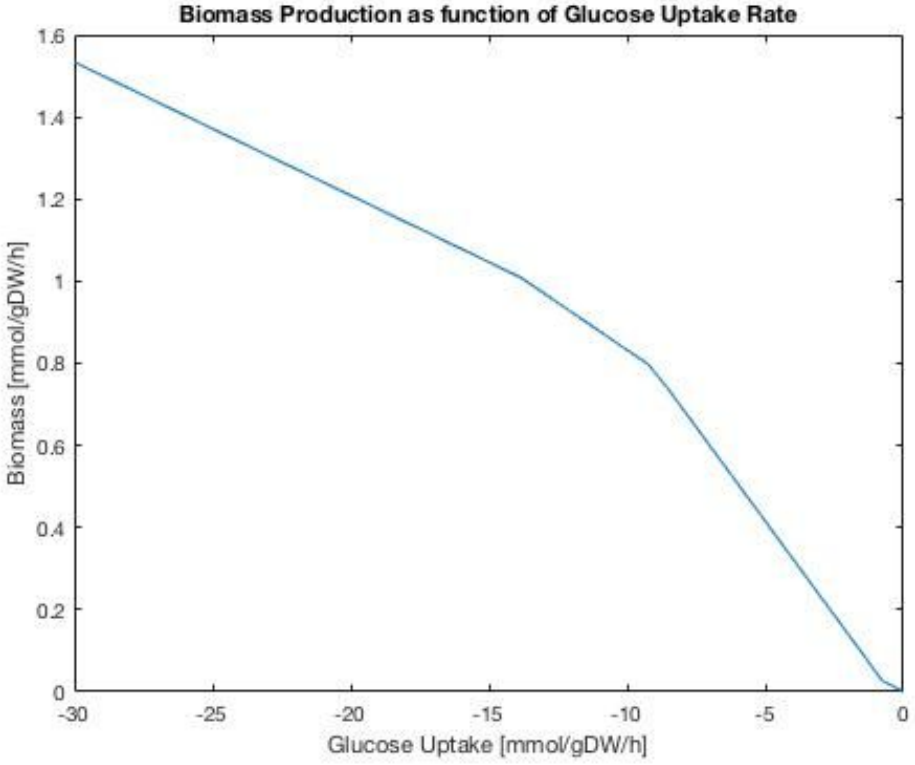

Fig.4 Glucose robustness analysis

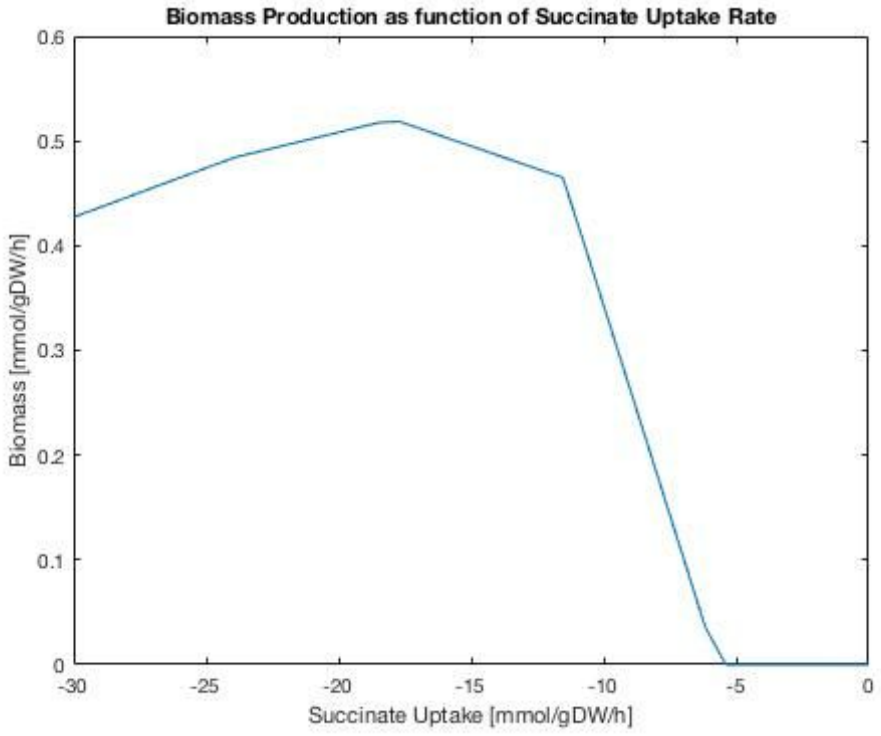

Fig.5 Succinate robustness analysis

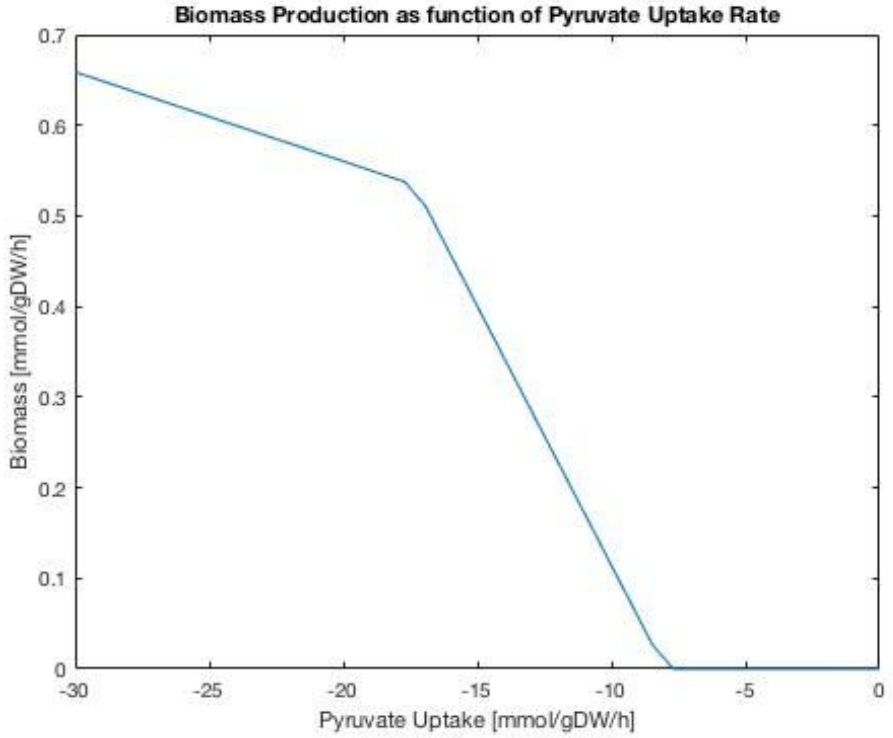

Fig. 6 Pyruvate robustness analysis

2.

From the graphics above, at the maximum oxygen uptake, one can see that the glucose uptake yields to higher biomass. The glucose is the best carbon source in this case. When 30 mmol/gDw/h are consumed, the cell produces 1.53 mmol/gDw/h of biomass. The pyruvate and the succinate are good nutrient for the cell too, one can notice that logically succinate should yield to higher biomass because it contains more carbon source than the pyruvate. The reason can be that the reaction is reversible at a certain amount of succinate at a fixed oxygen uptake.

II-PHENOTYPIC PHASE PLANES

In order to see distinct phases of optimal growth with different use of oxygen and respectively glucose, succinate and pyruvate as substrates, we plotted the phenotype phase planes shown below.

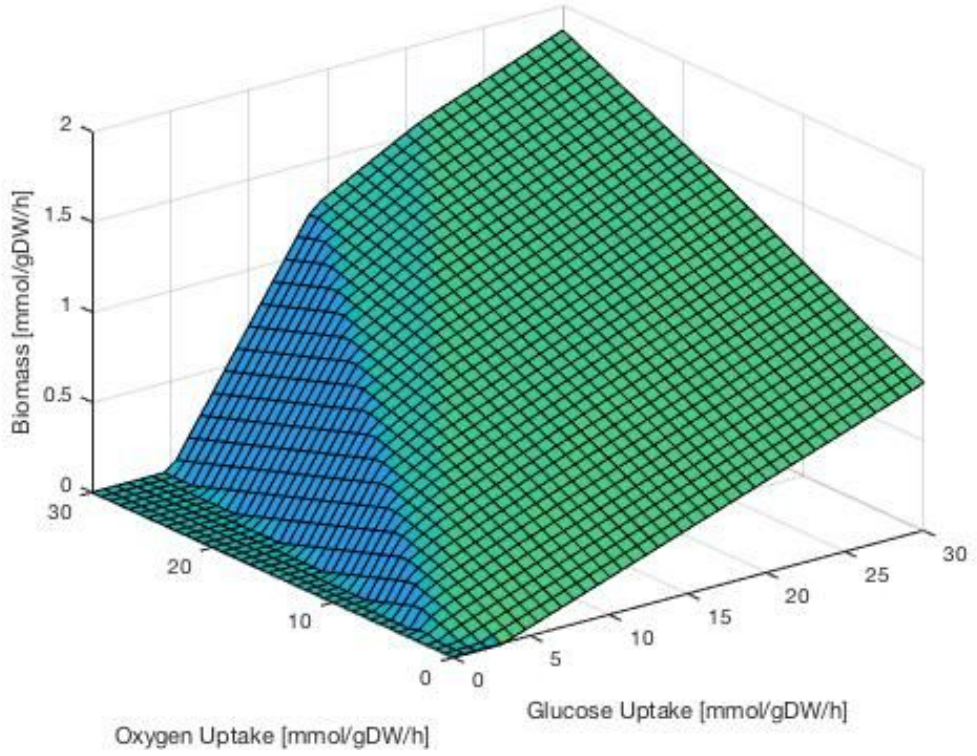

Fig. 7 Phenotype phase plane for the substrates oxygen and glucose

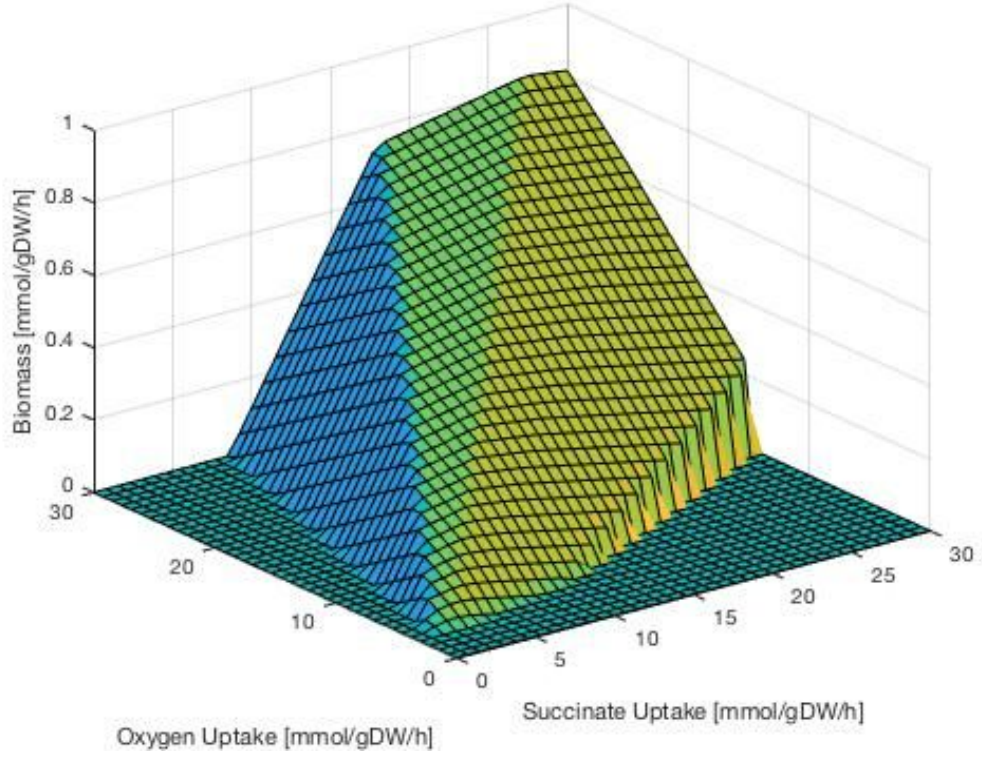

Fig. 8 Phenotype phase plane for the substrates oxygen and succinate

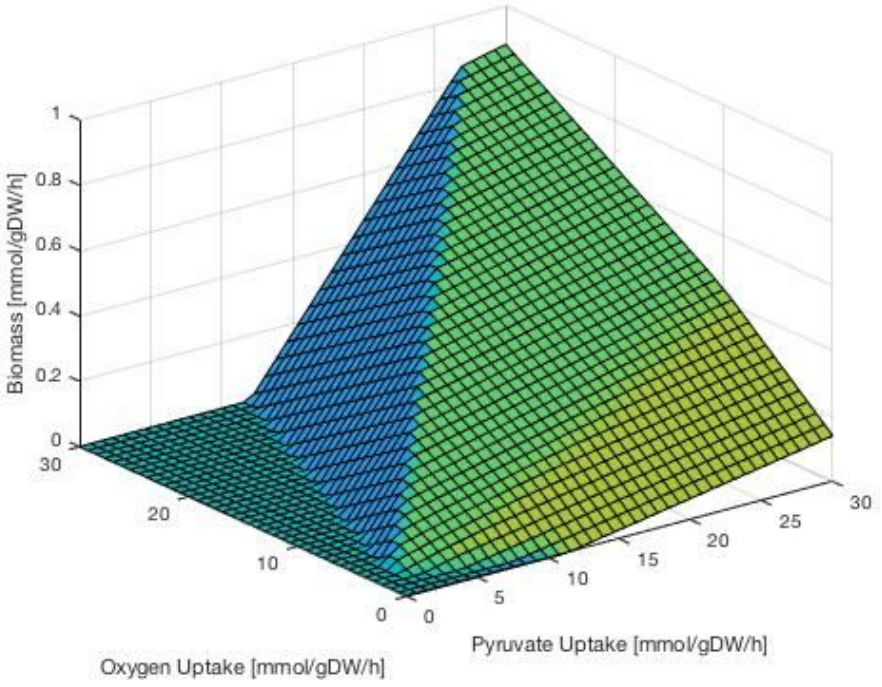

Fig. 9 Phenotype phase plane for the substrates oxygen and pyruvate

III- SHADOW PRICES

Finally, we would like to determine the different metabolic phenotypes for each region observed on the planes. To do so we look at the shadow prices below, which give us the

slope of the isoclines in each region which is the same value as the objective function. We numbered the different regions of the shadow prices for further analysis of the graphs.

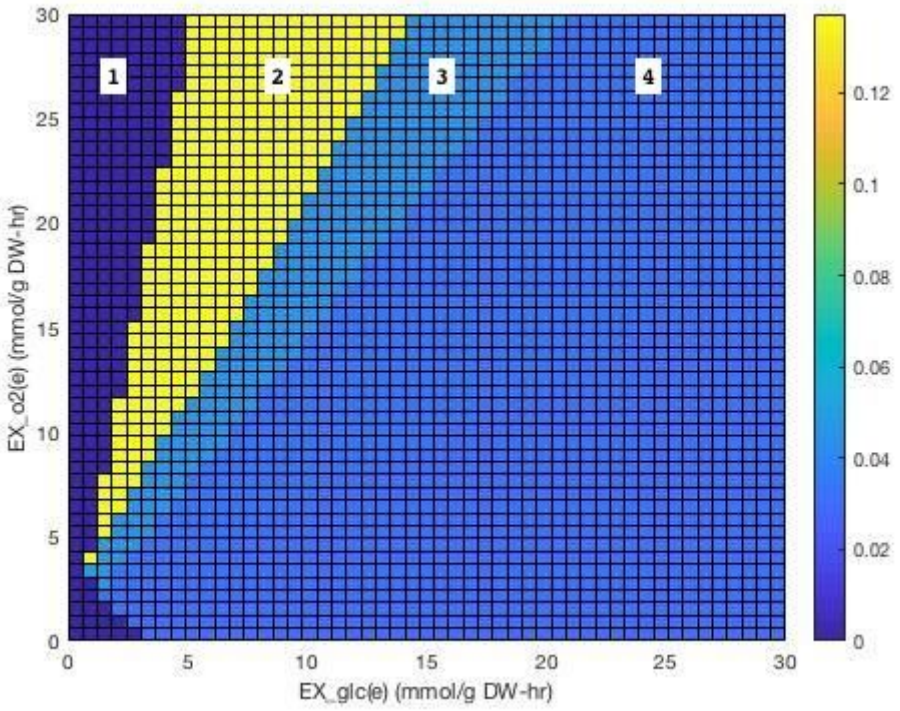

Fig. 10 Shadow prices for the substrate glucose

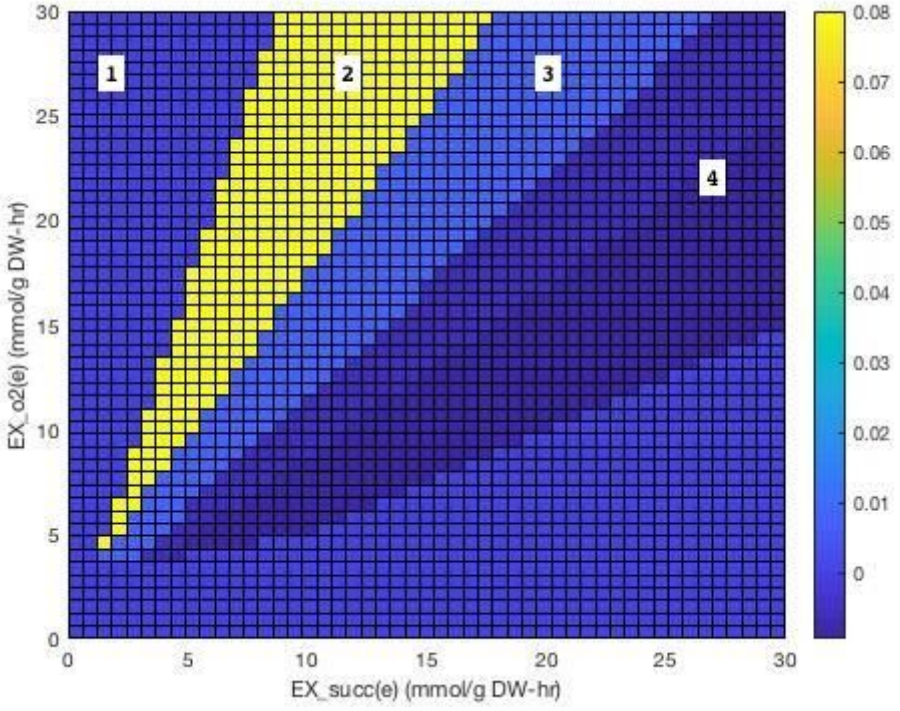

Fig. 11 Shadow prices for the substrate succinate

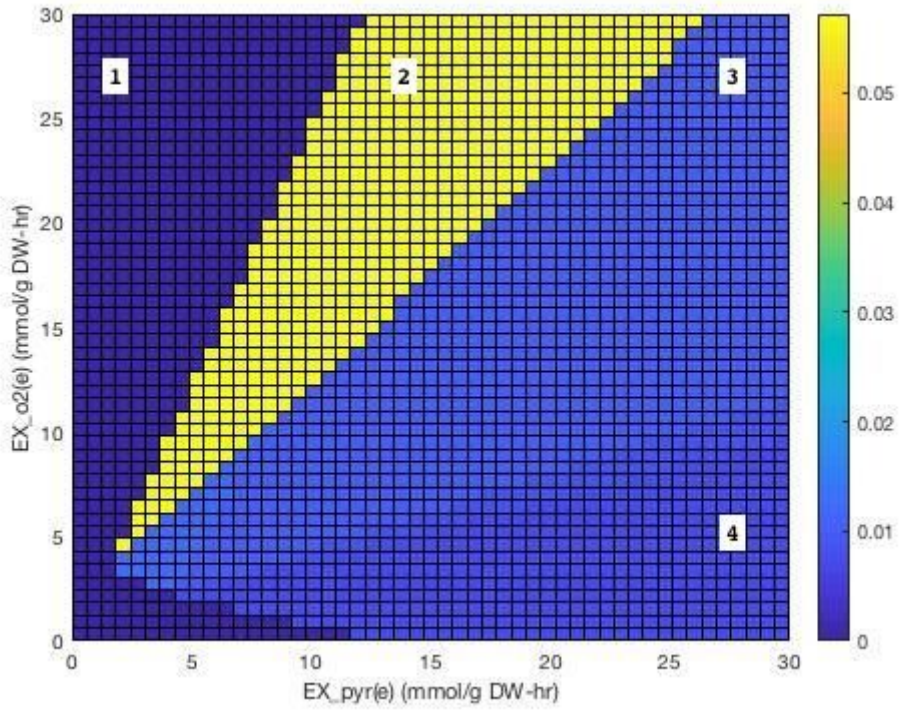

Fig. 12 Shadow prices for the substrate pyruvate

3. Which phenotypes are the most sensitive to a variation in carbon uptake? Find the differences of the flux distribution between the metabolic phenotypes you observe in the phase planes. (Hint: Use ESCHER maps to visualize the flux profiles)

The regions with the highest values regions on the Shadow Prices are where the slope is the biggest meaning a small change of glucose/pyruvate/succinate will lead to a greater change in the biomass yield.

We are now looking at the different colored areas of the Shadow Prices displayed above. We use the same numbering.

<u>Glucose:</u>
Area 1 is the infeasible region as there is either too little oxygen or too little glucose.
Area 2 represents where we obtain the highest biomass yield. The Krebs cycle steps not doned in this metabolic phenotype are overruled by a more direct step preventing the formation of more intermediate products and thus the spending of too much energy. This area is defined as the optimal case.
Area 3 is where the whole Krebs cycle is functioning and we obtain succinate as a side product.
Area 4 is where acetate and formate are generated only half of the Krebs cycle is functioning.

From this we can conclude that the line of optimality would be the one delimiting area 2 from area 3.
Below are the Escher maps from every metabolic phenotypes:

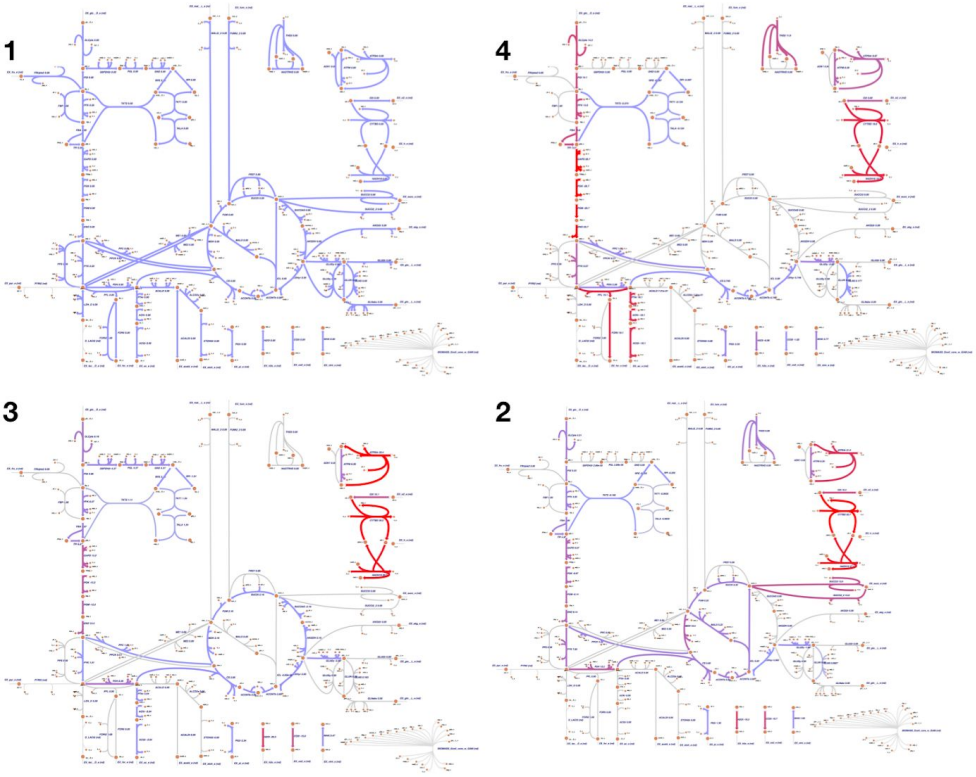

Fig. 13 Overview of the ESCHER map under the conditions of the different areas numbered in Fig. 10 for glucose

Succinate:
Area 1 is where there is no biomass yield. However the metabolic phenotypes is one where three fourth of the Krebs cycle is functioning and formate and acetate are produced.
Area 2 is the optimal region where we have the maximal biomass yield, without any side products.
Area 3 leads to the whole Krebs cycle functioning and the formation of acetate.
Area 4 is the infeasible regions. As seen in the robustness analysis, it seems that past a certain value of succinate uptake (approximately 17.69 mmol/gDW/h) the value of the biomass decreases. Here we are in a scenario where succinate is in excess and no metabolic reactions can take place.
As before, we can understand that the line of optimality would be the one delimiting area 2 from area 3.
Below are the Escher maps we used for our reasoning:

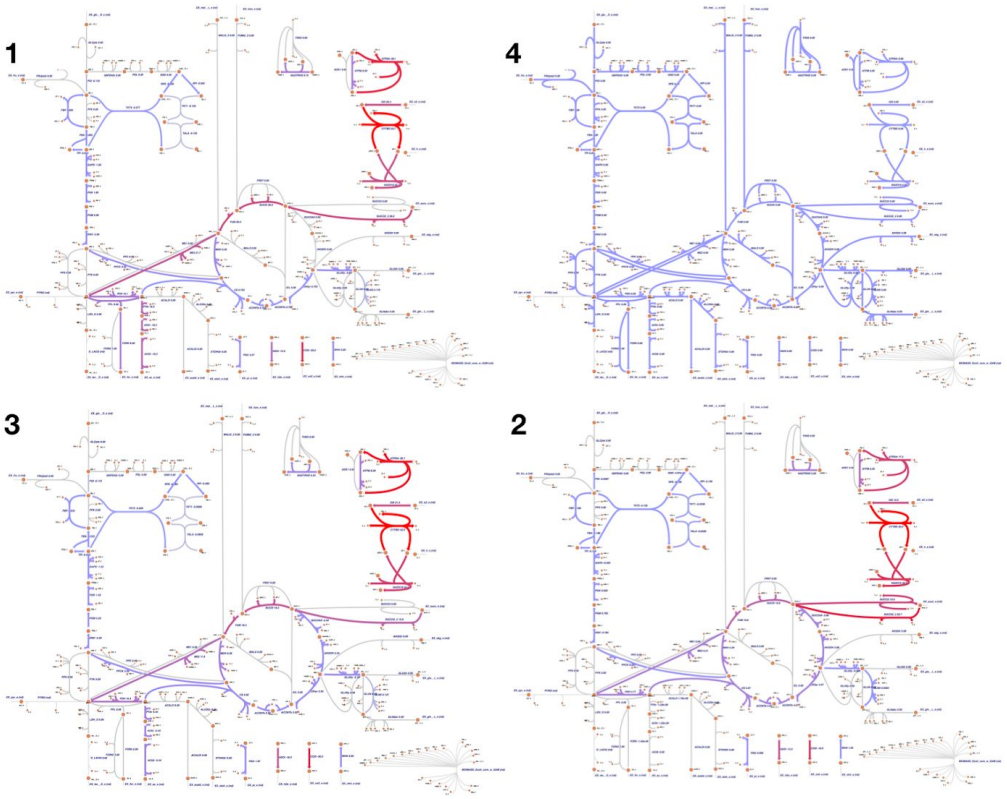

Fig. 14 Overview of the ESCHER map under the conditions of the different areas numbered in Fig. 11 for succinate

Pyruvate:

Area 1 is the infeasible region where there is either too little oxygen or too little carbon substrate: pyruvate, for any metabolic reaction to occur.

Area 2 is where the full Krebs cycle is functioning without forming any side products: it is where we have the optimal conditions.

Area 3 defines the case where the whole Krebs cycle can be performed with the generation of acetate.

Area 4 delimits the conditions under which only half of the Krebs cycle works and formate as well as acetate are generated as side products.

Thus, once again, the line of optimality is the one delimiting areas 2 and 3.

Below are the Escher maps from every metabolic scenarios seen on the shadow prices:

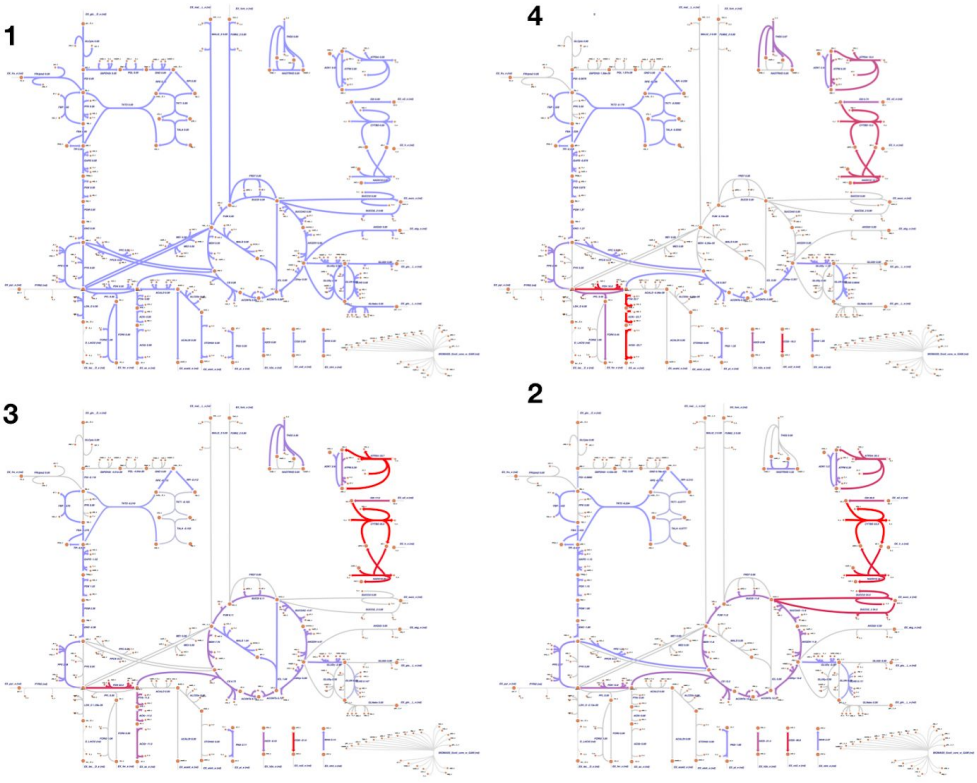

Fig. 14 Overview of the ESCHER map under the conditions of the different areas numbered in Fig. 12 for pyruvate

5. Compare the curves and the phase planes. What can we learn from these results? What is the nutrient to which biomass is the most sensitive? What are the optimal conditions for growth?

The phase planes (Figures 7,8 and 9) show the biomass production as a function of oxygen uptake and substrate uptake. From these three graphs it is visible that the glucose biomass production is generally higher than the production for the other substrates. In fact, it is the substrate less affected by lack of oxygen or excess oxygen. Pyruvate is in turn the substrate most affected by excess of oxygen since its “excess oxygen” area, at the top left corner, is the largest of the three. Succinate is the substrate most affected by the lack of oxygen, which is expected given that there are no anaerobic pathways for this substrate.

The gradient of these surfaces indicate the sensitivity of the biomass to the studied substrate. The curve with the steepest gradients is that of succinate. This is due to the lack of flexibility in the consumption pathways for growth. It is only under specific oxygen and succinate concentrations that significant biomass is observed. In contrast, the curve for glucose is much flatter, indicating its flexibility for alternative pathways including the production of other substrates that could be consumed as well. For all the studied substrates, the optimal conditions for growth are given by stoichiometric amounts of substrate and oxygen. At stoichiometric conditions there will be minimal side reactions occuring. For glucose this curve corresponds to y=6x, for pyruvate it is y=3x, and for succinate it is y=3.5x.