Maximum entropy and population heterogeneity in continuous cell cultures meet experimental data, preliminary results

February 11, 2020

1 Materials and Method

1.1 Model framework

The model framework used is detaily explained in (Fernandez-de Cossio-Diaz et al., 2017) and (Fernandez-de Cossio-Diaz & Mulet, 2019).

1.2 Chemostat experimental data

1.2.1 Human cell line AGE1.HN.AA1

Experimental data were taken from (Rath, 2017). In this work, the author performed 6 continuous cultures, (A, B, C, D, E, F), with the cell line AGE1.HN.AA1. The parental line AGE1.HN was established by the company ProBioGen (ProBioGen AG, Berlin, Germany) from a tissue sample of a human brain. All culture's feed mediums were based on the standard 42-Max-UB-medium, which is serum-free and was specially developed for the AGE1.HN cell line *Table* 1. The experiments were run under various conditions, differing mainly in the dilution rate (D) and the feed medium composition of glucose (GLC), glutamine (GLN) and galactose ($GAL/GALC$) *Table* 2.

For each experiment, a steady-state condition was reached, (steady states were labeled A, B, C, D, E, $F01$), and several observables was reported *Tables* $3-4$. Particularly relevant for this work was the growth rate (μ), D, the viable cell density (Xv) and the medium concentration (s) and derived uptake rate (u/q) for a set of metabolites (GLC, lactose (LAC), GLN, ammonium ($NH4$), GAL, pyruvate (PYR), glutamate (GLU), alanine (ALA), asparagine (ASP)). A unit conversion was required to make experimental data and our model compatible. For this propose the only external data needed was the cell mass density, 0.25 pgDW/ μm^3 (Niklas et al., 2011).

1.2.2 Escherichia coli KJ134

Continuous cultivation data for Escherichia coli ($E.coli$) were taken from (van Heerden & Nicol, 2013). The used Escherichia coli KJ134 strain was genetically modified for succinic acid fermentation. A dozen of steady states was recorded at different dilution rates and GLC availability in the feed medium. Data for the effluent concentration of GLC, succinic acid (SA), acetic acid (AcA), form acid (FA) and malic acid (MA) for the different steady states were given. Also the cell density was reported.

1.3 Preparing GEMs

1.3.1 Human

For modeling the metabolism of the human derived cell AGE1.HN.AA1, it was used a GEM from (Shlomi et al., 2011)(download link: ??). The GEM biomass equation was modified in agreement with the biomass composition for AGE1.HN.AA1 reported in (Niklas et al., 2013). The production of α1-antitrypsin ($A1AT$) was also considered, it was used the protein sequence, $https://www.drugbank.ca/polypeptides/P01009$, and the reported production rate for $A1AT$ (Rath, 2017) to set a maintenance for the required amino acids. Additionally, an atp demand ($ATPM$) was included, taking the maintenance energy demand for mammalian cells reported in (Fernandez-de Cossio-Diaz & Vazquez, 2018).

1.3.2 E. coli

With the same propose of modeling the cellular metabolism, the $E.coli\ GEM$ iJR904 (Reed et al., 2003) (download link: $https://darwin.di.uminho.pt/models$) was modified to match with (van Heerden & Nicol, 2013) data. The reported genetic modifications for the Escherichia coli KJ13 strain were included. It was also added enzymatic cost constraints in concordance with (Beg et al., 2007).

2 Results and Discussion

The main objective of this work is to explore the capability of an application of the Maximum Entropy ($MaxEnt$) framework to the modeling of the cellular metabolism in a continuous cultivations regime. $MaxEnt$ has been used for the analysis of other biology related problems with success. It has become into an useful tool when we are dealing with limited data, which actually is a common scenario in biology (De Martino & De Martino, 2018). In particular, the framework proposed in (Fernandez-de Cossio-Diaz et al., 2017) and (Fernandez-de Cossio-Diaz & Mulet, 2019) allows to link effectively the macroscopic variables that defined the state of a chemostat steady state, commonly accessible, with the underlying cellular metabolic state, harder to determine.

In the former article, a more traditional approach to model the metabolism was taken. It delegates in Flux Balance Analysis (FBA) (Orth et al., 2010) for choosing the vector of reactions fluxes assumed to be determining the current metabolic state of the cell under the cultivation conditions. This method have been heavily used in the past decades with good results and a variety of applications ($citeRequired$!!!). It has the advantage of not requiring kinetic parameters from the cellular metabolism, which are generally unavailable data. This is possible because FBA apply a steady state assumption, justified in the time scale difference between regulatory (slow) and metabolic (fast) processes (De Martino & De Martino, 2018). Another assumption FBA makes is that the cell population in the culture is homogeneous and are optimizing a given metabolic objective. FBA returns the vector of all reactions fluxes that optimize the objective function subject to the applied constraints ($citeRequired$!!!). This vector will be taken as the definition of the metabolic state for all the cells in the culture, and with it all the predictions or analysis will be made (Fernandez-de Cossio-Diaz et al., 2017).

To overcome the limitation of not considering culture heterogeneity, although data shows that two cells in a culture are unlikely to be equal ([$citeRequired$!!!18,]), a probability distribution can be defined over the set of all the possible metabolic states. This distribution describe how probable a cell is found in one of the those states. To infer such probability distribution in agreement with available experimental data, the $MaxEnt$ principle can be used (Fernandez-de Cossio-Diaz & Mulet, 2019). Precisely, in this work $MaxEnt$ was apply in such a way that it returns the probability distribution that maximized the entropy and ensure the expected value of the growth rate to match with the experimentally observed. Now, instead of a vector of reactions fluxes that optimize an objective function, the model will returns a vector containing the expected values for each reaction flux due to the inferred $MaxEnt$ distribution. This is a major advantage with respect to the FBA framework, $MaxEnt$ do not assume that the cells "*have a goal*", an objective function, it claims to compute the bias-less probability distribution in concordance with the imposed, data driven, constraints (De Martino & De Martino, 2018).

2.0.1 Human

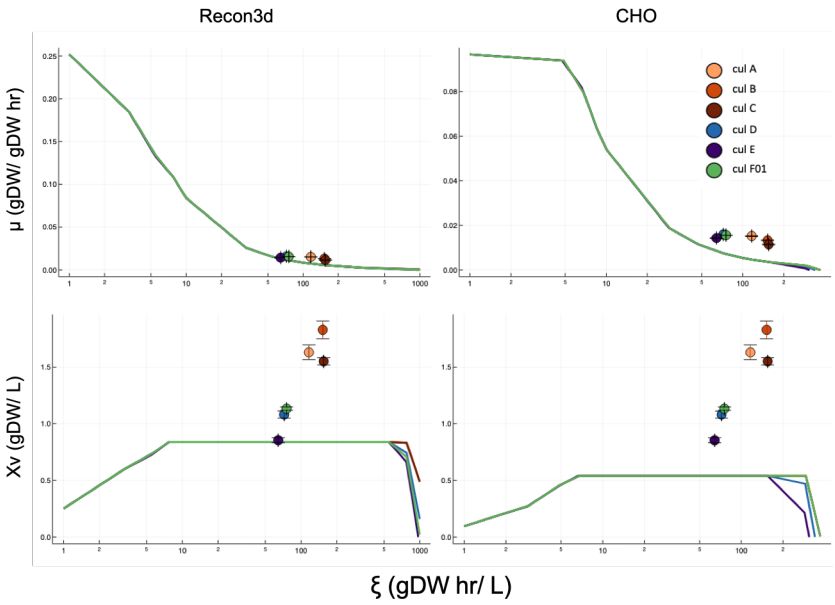

Figure 1: FBA results showing the growth rate, μ, and the viable cell density, Xv, dependence of ξ for the six steady states. The solid lines represent the model predictions and the colored points show the experimental results. The model data was obtained for feed mediums with 0.5mM ($Recon3D$) and 0.0mM (CHO) concentration of phosphatidylethanolamine

3 References

Beg, Q. K., Vazquez, A., Ernst, J., Menezes, M. A. d., Bar-Joseph, Z., Barabási, A.-L., & Oltvai, Z. N. (2007, July). Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences*, *104*(31), 12663–12668. Retrieved 2019-11-06, from `https://www.pnas.org/content/104/31/12663` doi: 10.1073/pnas.0609845104

De Martino, A., & De Martino, D. (2018, April). An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon*, *4*(4), e00596. Retrieved 2019-10-25, from `https://linkinghub.elsevier.com/retrieve/pii/S2405844018301695` doi: 10.1016/j.heliyon.2018.e00596

Fernandez-de Cossio-Diaz, J., Leon, K., & Mulet, R. (2017). Characterizing steady states of genome-scale metabolic networks in continuous cell cultures. *PLoS Computational Biology*, *13*(11), 1–22. doi: 10.1371/journal.pcbi.1005835

Fernandez-de Cossio-Diaz, J., & Mulet, R. (2019, February). Maximum entropy and population heterogeneity in continuous cell cultures. *PLOS Computational Biology*, *15*(2), e1006823. Retrieved 2019-10-31, from `https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006823` doi: 10.1371/journal.pcbi.1006823

Fernandez-de Cossio-Diaz, J., & Vazquez, A. (2018, May). A physical model of cell metabolism. *Scientific Reports*, *8*(1), 1–13. Retrieved 2019-12-09, from `https://www.nature.com/articles/s41598-018-26724-7` doi: 10.1038/s41598-018-26724-7

Niklas, J., Priesnitz, C., Rose, T., Sandig, V., & Heinzle, E. (2013). Metabolism and metabolic burden by a 1 -antitrypsin production. *Metabolic Engineering*, *16*, 103–114. Retrieved from `http://dx.doi.org/10.1016/j.ymben.2013.01.002` doi: 10.1016/j.ymben.2013.01.002

Niklas, J., Schräder, E., Sandig, V., Noll, T., & Heinzle, E. (2011). Quantitative characterization of metabolism and metabolic shifts during growth of the new human cell line AGE1.HN using time resolved metabolic flux analysis. *Bioprocess and Biosystems Engineering*, *34*(5), 533–545. doi: 10.1007/s00449-010-0502-y

Orth, J. D., Thiele, I., & Palsson, B. O. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. Retrieved from `http://dx.doi.org/10.1038/nbt.1614` doi: 10.1038/nbt.1614

Rath, A. (2017). *Characterisation of cell growth, metabolism and recombinant protein production during transient and steady state conditions for the human cell line AGE1.HN-AAT* (PhD Thesis).

Reed, J. L., Vo, T. D., Schilling, C. H., & Palsson, B. O. (2003, August). An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biology*, *4*(9), R54. Retrieved 2019-12-13, from `https://doi.org/10.1186/gb-2003-4-9-r54` doi: 10.1186/gb-2003-4-9-r54

Shlomi, T., Benyamini, T., Gottlieb, E., Sharan, R., & Ruppin, E. (2011, March). Genome-Scale Metabolic Modeling Elucidates the Role of Proliferative Adaptation in Causing the Warburg Effect. *PLOS Computational Biology*, *7*(3), e1002018. Retrieved 2019-11-05, from `https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1002018` doi: 10.1371/journal.pcbi.1002018

van Heerden, C. D., & Nicol, W. (2013, September). Continuous and batch cultures of Escherichia coli KJ134 for succinic acid fermentation: metabolic flux distributions and production characteristics. *Microbial Cell Factories*, *12*(1), 80. Retrieved from `https://doi.org/10.1186/1475-2859-12-80` doi: 10.1186/1475-2859-12-80

4 Appendix

Substance	Value	Dimension	Analytical method
Pluronic	1.0	g/L	as stated by Xell
NaHCO3	2.1	g/L	as stated by Xell
Osmolality	290.0	$^{mOsm}/_{kg}$	FPDO[a]
pH value	7.4	-	pH meter
GALC	0.5	g/L	AEC[b]
GLC	5.5	g/L	Bioprofile
AMM	0.3	mM	Bioprofile
LAC	0.0	g/L	Bioprofile
PYR	2.9	mM	AEC
GLU	636.9	µM	AEC
ALA	437.1	µM	AEC
ARG	1588.2	µM	AEC
ASN	920.4	µM	AEC
ASP	2197.9	µM	AEC
CYS	963.1	µM	AEC
GLY	1196.0	µM	AEC
HIS	642.7	µM	AEC
ILE	1744.9	µM	AEC
LEU	1893.2	µM	AEC
LYS	1256.0	µM	AEC
MET	601.3	µM	AEC
PHE	1039.4	µM	AEC
PRO	1040.5	µM	AEC
SER	3027.4	µM	AEC
THR	1502.4	µM	AEC
TRP	383.8	µM	AEC
TYR	1109.7	µM	AEC
VAL	1811.9	µM	AEC

a: freezing point depression osmometer (FPDO); b: anion exchanger chromatography (AEC);

Table 1: Measured composition of the 42-MAX-UB standard medium. Extracted from (?, ?)

Exp. ID	*DR* ($^{1}/_{h}$)	Preculture (passage no.)	GLC (mM)	GLN (mM)	GALC (mM)
A	0.0140	7	10	5	3
B	0.0120	7	10	5	3
C	0.0100	5	10	5	3
D	0.0150	8	10	2	3
E	0.0133	4	8	2	3
F01	0.0150	10	10	5	0

Table 2: The dilution rates, preculture ages and the 42-Max-UB-medium modified components concentrations used in (?, ?) for the 6 steady states. Table adapted from (?, ?)

Variable	Exp. A		Exp. B		Exp. C		Exp. D		Exp. E	
	Average ± SD	Rel. SD (%)	Average ± SD	Rel. SD (%)	Average ± SD	Rel. SD (%)	Average ± SD	Rel. SD (%)	Average ± SD	Rel. SD (%)
Setpoints:										
DR (1/h)	0.0140	-	0.0120	-	0.0100	-	0.0150	-	0.0133	-
GLC feed conc. (mM)	10.0	-	10.0	-	10.0	-	10.0	-	8.0	-
GLN feed conc. (mM)	5.0	-	5.0	-	5.0	-	2.0	-	2.0	-
GALC feed conc. (mM)	3.0	-	3.0	-	3.0	-	3.0	-	3.0	-
GLC/GLN ratio (mol/mol)	2	-	2	-	2	-	5	-	4	-
X_V (E6 cells/mL)	2.490 ± 0.100	4.0	2.745 ± 0.118	4.3	2.735 ± 0.056	2.0	1.489 ± 0.043	2.9	0.998 ± 0.025	2.5
X_D (E6 cells/mL)	0.210 ± 0.032	15.2	0.291 ± 0.021	7.4	0.378 ± 0.047	12.5	0.086 ± 0.006	6.5	0.075 ± 0.014	18.5
CV_v (μL/mL)	6.56 ± 0.48	7.4	7.19 ± 0.67	9.4	6.23 ± 0.21	3.4	4.33 ± 0.27	6.3	3.40 ± 0.17	4.9
μ (1/h)	0.0152 ± 0.0002	1.2	0.0133 ± 0.0001	1.1	0.0114 ± 0.0002	1.4	0.0159 ± 0.0001	0.6	0.0143 ± 0.0002	1.4
CD (μm)	17.1 ± 0.25	1.5	17.2 ± 0.19	1.1	16.3 ± 0.24	1.5	17.7 ± 0.30	1.5	18.7 ± 0.20	1.1
GLC (mM)	0.0 ± 0.00	-	0.0 ± 0.00	-	0.0 ± 0.00	-	0.0 ± 0.00	-	0.0 ± 0.00	-
LAC (mM)	2.1 ± 1.52	-	0.0 ± 0.00	-	0.0 ± 0.00	-	12.5 ± 0.60	5.0	12.7 ± 0.40	3.0
GLN (mM)	0.8 ± 0.08	10.0	0.5 ± 0.02	3.5	0.6 ± 0.04	7.3	0.5 ± 0.10	11.1	0.5 ± 0.10	13.7
AMM (mM)	3.4 ± 0.22	6.7	5.5 ± 0.19	3.5	5.9 ± 0.09	1.5	1.5 ± 0.03	2.1	1.6 ± 0.02	1.6
GALC (mM)	1.9 ± 0.08	4.5	1.9 ± 0.06	2.9	1.7 ± 0.02	0.9	2.1 ± 0.10	2.5	2.4 ± 0.10	3.2
PYR (mM)	0.2 ± 0.00	1.8	0.1 ± 0.00	3.2	0.1 ± 0.00	3.2	0.3 ± 0.04	12.6	1.1 ± 0.10	5.3
GLU (mM)	1.2 ± 0.08	6.2	1.3 ± 0.01	0.5	1.0 ± 0.05	5.0	0.7 ± 0.10	12.9	0.8 ± 0.10	12.8
ALA (mM)	1.4 ± 0.09	6.4	0.2 ± 0.03	13.8	0.1 ± 0.01	14.2	0.3 ± 0.02	8.9	1.0 ± 0.05	5.2
ASP (mM)	1.6 ± 0.06	3.9	1.5 ± 0.05	3.2	1.3 ± 0.02	1.2	2.4 ± 0.17	7.1	2.8 ± 0.30	10.8
A1AT (mg/L)	57.9 ± 3.03	5.2	81.0 ± 9.21	11.4	72.1 ± 1.09	1.5	50.9 ± 1.50	2.9	44.6 ± 2.80	6.2
$Y_{CVv/glc}$ (μL/mmol)	711 ± 47	6.6	827 ± 77	9.4	753 ± 17	2.3	482 ± 30	6.3	518 ± 18	3.5
$Y_{CVv/gln}$ (μL/mmol)	1679 ± 140	8.3	1770 ± 145	8.2	1601 ± 52	3.3	3617 ± 215	6.0	2837 ± 201	7.1
$Y_{lac/glc}$ (mol/mol)	0.2 ± 0.15	-	0.0 ± 0.01	-	0.0 ± 0.00	-	1.2 ± 0.10	5.3	1.8 ± 0.10	3.2
$Y_{amm/gln}$ (mol/mol)	0.8 ± 0.05	6.6	1.2 ± 0.04	3.6	1.3 ± 0.03	1.9	1.0 ± 0.05	4.9	1.1 ± 0.10	4.8
$Y_{ala/gln}$ (mol/mol)	0.2 ± 0.02	9.5	0.0 ± 0.00	-	0.0 ± 0.00	-	0.0 ± 0.00	-	0.6 ± 0.04	7.9
q_{GLC} (nmol/μL h)[a]	21.43 ± 1.56	7.3	16.17 ± 1.76	10.9	15.13 ± 0.51	3.4	33.05 ± 2.17	6.6	27.64 ± 1.33	4.8
q_{LAC} (nmol/μL h)[a]	-4.43 ± 3.05	-	0.00 ± 0.00	-	0.00 ± 0.00	-	-41.23 ± 4.13	10.0	-49.68 ± 1.30	2.6
q_{GLN} (nmol/μL h)[a]	9.10 ± 0.82	9.0	7.55 ± 0.72	9.5	7.12 ± 0.31	4.3	4.40 ± 0.28	6.4	5.07 ± 0.41	8.1
q_{AMM} (nmol/μL h)[a]	-7.19 ± 0.67	9.3	-9.21 ± 0.97	10.5	-9.55 ± 0.30	3.2	-4.48 ± 0.35	7.8	-5.66 ± 0.33	5.9
q_{GALAC} (nmol/μL h)[a]	1.94 ± 0.21	10.8	1.55 ± 0.22	14.3	2.46 ± 1.26	51.1	1.92 ± 0.24	12.3	1.61 ± 0.37	23.3
q_{PYR} (nmol/μL h)[a]	5.82 ± 0.52	8.9	4.70 ± 0.57	12.1	4.48 ± 0.11	2.4	9.10 ± 0.64	7.1	7.04 ± 0.55	7.8
q_{GLU} (nmol/μL h)[a]	-1.31 ± 0.18	14.0	-1.05 ± 0.10	9.4	-0.52 ± 0.09	17.9	-0.71 ± 0.31	43.5	-0.88 ± 0.38	43.8
q_{ALA} (nmol/μL h)[a]	-2.14 ± 0.27	12.8	0.39 ± 0.03	6.7	0.54 ± 0.03	5.9	0.10 ± 0.12	127.8	-2.81 ± 0.26	9.4
q_{ASP} (nmol/μL h)[a]	1.19 ± 0.19	15.9	1.20 ± 0.10	8.1	1.49 ± 0.05	3.5	-0.41 ± 0.62	150.7	-2.16 ± 1.15	53.1
q_{A1AT} (pg/cell d)[a]	-7.87 ± 0.19	2.4	-8.54 ± 1.34	15.7	-6.33 ± 0.19	3.0	-12.31 ± 0.60	4.9	-14.29 ± 1.08	7.5

a: Substrate uptake is indicated by a positive rate, whereas a negative value indicates a production rate.

Table 3: Steady-state values reported by (?, ?) for different parameters from continuous cultivations, varying *GLC* and *GLN* feed concentrations and with 3 mM *GAL*. Table taken from (?, ?)

Variable	Exp. F01	
	Average ± SD	Rel. SD (%)
Setpoints:		
DR (1/h)	0.0150	-
GLC feed conc. (mM)	10	-
GLN feed conc. (mM)	5	-
GALC feed conc. (mM)	0	-
GLC/GLN ratio (mol/mol)	2	-
X_V (E6 cells/mL)	1.701 ± 0.023	1.3
X_D (E6 cells/mL)	0.081 ± 0.004	5.1
CV_v (μL/mL)	4.54 ± 0.24	5.2
μ (1/h)	0.0156 ± 0.0000	0.2
CD (μm)	17.2 ± 0.33	1.9
GLC (mM)	0.0 ± 0.00	-
LAC (mM)	7.8 ± 0.64	8.2
GLN (mM)	1.9 ± 0.18	9.3
AMM (mM)	3.4 ± 0.17	5.1
PYR (mM)	0.4 ± 0.02	6.6
GLU (mM)	1.0 ± 0.05	5.2
ALA (mM)	0.9 ± 0.07	8.2
ASP (mM)	1.9 ± 0.21	10.8
A1AT (mg/L)	44.0 ± 1.71	3.9
$Y_{CVv/glc}$ (μL/mmol)	626 ± 25	4.0
$Y_{CVv/gln}$ (μL/mmol)	1865 ± 201	10.8
$Y_{lac/glc}$ (mol/mol)	0.9 ± 0.07	8.2
$Y_{amm/gln}$ (mol/mol)	1.3 ± 0.04	3.1
$Y_{ala/gln}$ (mol/mol)	0.2 ± 0.02	8.5
q_{GLC} (nmol/μL h)[a]	24.9 ± 1.08	4.3
q_{LAC} (nmol/μL h)[a]	-22.6 ± 1.27	5.6
q_{GLN} (nmol/μL h)[a]	8.4 ± 1.07	12.7
q_{AMM} (nmol/μL h)[a]	-10.9 ± 1.13	10.3
q_{PYR} (nmol/μL h)[a]	4.5 ± 0.31	6.9
q_{GLU} (nmol/μL h)[a]	-0.7 ± 0.15	20.4
q_{ALA} (nmol/μL h)[a]	-1.99 ± 0.33	16.3
q_{ASP} (nmol/μL h)[a]	1.65 ± 0.64	39.0
q_{A1AT} (pg/cell d)[a]	-9.2 ± 0.25	2.7

a: Substrate uptake is indicated by a positive rate, whereas a negative value indicates a production rate.

Table 4: Steady-state values reported by (?, ?) of different parameters from continuous cultivations, varying GLC and GLN feed concentrations and without GAL. Table taken from (?, ?)