Bioinformatics, 2020, 1–12
doi: 10.1093/bioinformatics/123456
Advance Access Publication Date: 14 Dec 2020
Manuscript Category

Flux Balance Analysis

Validation of *Faecalibacterium prausnitzii* genome-scale metabolic model with in vitro data

C. Joseph[1,*], M. Neilson[2] , P. Juyal[2], T. Blokker[2], B. Florenzi[2], K. Faust[1]

[1]Department of Microbiology, Immunology and Transplantation, Address: Rega - Herestraat 49 - box 1030 - 3000 Leuven,

[2]Department of Bioinformatics Students, Address Leuven/Heverlee/Kessel-Lo

*To whom correspondence should be addressed.

Associate Editor: V. van Noort

Received on 12 Dec 2020; revised on 13 Dec 2020; accepted on 14 Dec 2020

Abstract

Motivation: The gut bacterium Faecalibacterium prausnitzii is one of the most abundant species in the gut, notably in the healthy gut. To understand the role of this species in health and its involvement in gut communities, we have decided to study this species using metabolic modelling. A genome-scale metabolic model (GEM) has been refined for this species and is available. To validate this model, we want to predict species growth in different media and compare the result with biological data.

Results: Our results suggest that the environment (medium) for the model is of utmost importance and further research is needed to make full use of the model and to validate it. Currently, the metabolic model can be used to investigate critical medium components. The hierarchy of the growth rate for the different media found in vitro could not be confirmed in the *in-silico* model. Further refinement of the medium needed to make qualitative or even quantitative predictions on the maximum growth rate that can be obtained in a medium.

Availability: The data is not yet published, questions should be addressed to the e-mail address below.

Contact: joseph.clemence@kuleuven.be

Introduction

Faecalibacterium prausnitzii (F. prausnitzii) is the most abundant bacterium in the human gut and can be approximately 15% of the entire microbiota of a healthy individual (Miquel *et al.,* 2013). It was formerly known as Fusobacterium prausnitzii but was reclassified as Faecalibacterium upon 16s rRNA sequencing (Miquel *et al.,* 2013). F. prausnitzii is extremely oxygen sensitive (EOS) and difficult to grow even in anaerobic conditions (Miquel *et al.*, 2013). Because it is so complex to culture it is interesting to develop an in-silico metabolic map for this bacteria and other anaerobic bacteria to identify essential growth components (Heinken *et al.*, 2014).

In this pursuit, the in-silico metabolic model of ***F. prausnitzii***, which was developed from gene annotation and transcriptomic data - hence also called a Genome Scale metabolic model (GEM), was utilized. A typical workflow includes performing flux balance analysis on the GEM to achieve optimum growth. However, FBA is a mathematically complex and computational intensive task. In this approach to study flow of metabolites through a biochemical network (such as GEM), all the metabolic reactions are represented in the form of a matrix. The stoichiometric coefficients of each metabolite in a metabolic reaction are represented in the individual columns of the matrix. This enforces stoichiometric constraints to ensure conservation of mass under steady state condition. The objective of this analysis can also be represented mathematically in the form of an objective function. For example if maximizing growth is the objective then biomass production is the objective function. The stoichiometric matrix and the objective function constitutively form a system of linear equations which are solved by linear programming using packages like COBRA.py.

In more detail, the objective function is represented as $Z = c^T v$ (Orth *et al*., 2010), where c is a vector of coefficients depicting contributions of individual reactions, and c^T if of the form [1 0 0 0 0...]. c^T is sparse because we are only interested in the reaction that represents biomass production; contributions from other reactions are set to 0. The second component of the objective function is the flux vector v which represents the flux of metabolite through a metabolic reaction. If the value is positive then the metabolite is produced (positive flux). Conversely, if the calculated value is negative then the metabolite is consumed (negative flux). Also, outside steady state conditions ($Sv \neq 0$) the flux cannot be balanced and it is possible to construct near-infinite permutations of *in-silico* media composition that can lead to greatly varied bacterial growth. Mathematically, this can be imagined as an infinite cloud of points in an

n-dimensional space where each dimension corresponds to the involved metabolite. To restrict these possibilities to a realistic scenario, it is necessary to constrain the solution space. The first constraint in FBA is steady state assumption:

$$Sv = 0$$

Secondly, as the flow of metabolites is not infinite, we set flux boundaries for each metabolite to further constrains the solution space. Together, with the stoichiometric constraints explained earlier, these constraints help to find an optimal solution in our flux balance analysis.

Methods

F. prausnitzii was grown in anaerobic conditions on the media RCM, mGAM, mMCB with the addition of acetate. The composition of the three media are as follows:

1. RCM: yeast extract, peptone, glucose, soluble starch, NaCl, CH3COONa, cysteine hydrochloride, agar *(Oxoid Ltd.)*;
2. mGAM: bacteriological peptone, soy peptone, proteose peptone, digested serum, yeast extract, meat extract, liver extract, dextrose, soluble starch, L-tryptophan, L-cysteine hydrochloride, Sodium thioglycolate, L-arginine, vitamin K1, hemin, potassium dihydrogen phosphate, sodium chloride *(Himedia Lab);*
3. mMCB: bacteriological peptone, soy peptone, yeast extract, tryptone, NaCL, K2PO4, KH2PO4, Na2SO4, Mg2SO4, CaCl2*2H2O, NH4Cl, Cysteine-HCL, NaHCO3, MnSO4*H2O, FeSOMnSO4*7H2O, ZnSO4*7H2O, hemin, menadione, resazurin, with the addition of acetate (kevin *et. al* 2018).

In the experimental data obtained from Joseph *et al*., the concentration of the bacterium on the different media was measured at OD 600 nm, and recorded over the course of 24 or 48 hours. These readings were then used to plot growth curves and estimate maximum growth rate using the "growthrates" R package. The fit_easy_linear method was used from the package which uses a heuristic approach for the estimations:

Fit_easylinear (time, y, h=5, quota=0.95)

The method initially log-transforms the OD values, then fits a linear regression for each of the subsets of *h* adjacent timepoints; it finds the one with the highest slope and also includes neighbouring subsets with a slope at least 95% of the highest one. The data window so obtained is fitted with a new linear model and the slope is the maximum growth rate for the dataset.

Metabolic model

The metabolic model that was validated in the course of the present study was developed by Heinken *et al.* for F.prausnitzii A2-165 and accounts for 1,030 reactions and 602 genes.

Medium-Composition

One of the major tasks while assessing metabolic models is to optimise the medium recipe. The nutritional environment for each media has been recreated in the metabolic network model of our interest following the guidelines set in Marinos *et al*., 2020. The composition of the three media used in the model was obtained from the website of the medium manufacturers like Oxoid, HyServe and ThermoFisher, while some literature provided detailed inputs. However these websites in some cases do not provide detailed 'molecular formulations' for example from a mass spectroscopy experiment, which is required in the model. Consequently, where required, we made informed assumptions about the constituents of a metabolic component. For example, it is quite safe to assume that Yeast extract, which is a major component of all these 3 mediums, is a source of amino acids; albeit, quantitative information about presence of trace elements/ vitamins in these often animal derived nutrients is also essential for the in-silico model to converge.

After refining the metabolite composition, both in terms of its quantity and molecular composition, we determined their BiGG IDs in preparation to include the metabolite into our model. These Ids are additionally prefixed by "EX_", to represent them in the exchange reactions. After this initial draft of the medium component list, we refined it iteratively with our analysis on the in-silico model that was guided by the experimental data. So, initially we started from the complete list of possible metabolites in the model, and performed step by step filtering of the metabolites as in when we found they did not cause any change in the model. Dipeptides were excluded from the media definitions, given that in simulation no ATP was consumed for splitting and thus their presence has no impact on growth. D-amino acids were also excluded, since they are not naturally occuring in the niche bacteria environment. If a certain molecule in the medium mapped to more than one metabolite in the model, for instance iron to Fe2+ and Fe3+, we tried to infer from literature the best representation. If a certain molecule in the experimental medium was not available in the model, we looked for similar molecules or excluded the element, after establishing that it would not cause significant changes to biomass.

Flux Constraints

Each metabolite is available at a certain concentration in the media and thus provides an input flux to the model. For analysis, the metabolites are included in the model through an exchange reaction and have bounded (lower and upper bound) flux value to enforce constraint explained previously. Consumption of a

compound by the bacterium corresponds to a negative value of the flux, while the production corresponds to a positive value; therefore, in order to define the media, the flux lower-bound was set to a pseudo-lower bound of -0.001 to avoid 0 growth, except for acetate in the mGAM medium and oxygen in all media (both 0) if the metabolite was not available for that reaction, and to a specific negative value if it was present. In case of unknown concentrations of metabolites in the media, we assumed the flux to be a mean of remaining flux values.

The flux value was then calculated as follows:

1. Assume density of 1 g/mL to transform ppm of metabolite into g/L;
2. Use the molecular weight (g/mol) to obtain the concentration of metabolite in mMol/L;
3. Assume the number of bacterial cells per volume to be equal to the number of cells indicated in the experimental data for the mMCB medium at an arbitrary time point in the exponential phase (nb/L);
4. Assume the cell dry weight of fb to be equal to 1,172 fg, the larger value given for a E.Coli cell (Loferer-Krößbacher *et al*., 1998) (gDw/cell);
5. Divide the molar concentration of the metabolite per the concentration of the bacteria in the medium multiplied by dry weight of a bacterial cell, to obtain the flux lower bound for each metabolite (Fig. 1).

$$Flux\ value = \frac{C_{metabolite}}{C_{bacteria} * \frac{gDw}{cell}} * \frac{1}{t_{experiment}}$$

$$\frac{mmol}{gDw * h} = \frac{\frac{mmol}{ml}}{\frac{nb}{ml} * \frac{gDw}{nb}} * \frac{1}{h} = \frac{mmol * \cancel{ml} * \cancel{nb}}{\cancel{ml} * \cancel{nb} * gDw} * \frac{1}{h}$$

Figure 1: Calculation of estimated flux lower bound

COBRApy

To perform analysis and prediction of growth of the model, the Python package COBRApy (Constraint-based reconstruction and analysis) was used. COBRApy is the object-oriented open-source Python implementation of basic methods from the COBRA Toolbox for Matlab. Indeed, FBA computations are constraint-based methods, for which COBRA Toolbox provides comprehensive support. Model, Metabolite, Reaction and Gene are the core classes of COBRApy, and the attributes of their objects can be directly accessed in an object-oriented fashion.

FBA Pipeline in Python

Using the model in question and the medium composition information available, it was feasible to recreate the media *in silico* to reflect the growth of *F.prausnitzii* based on the metabolic reactions of the model in 6 major steps (Fig. 2).

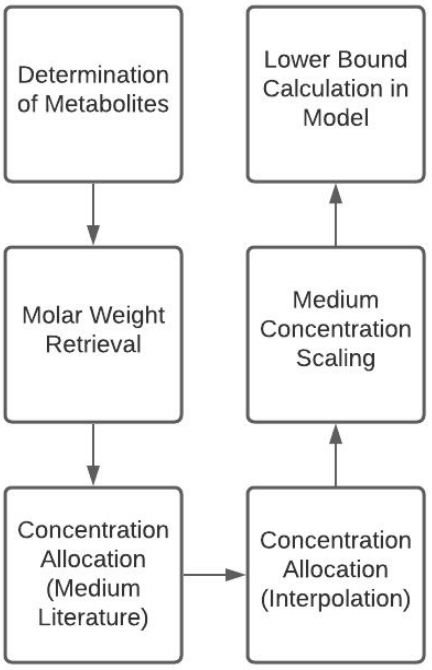

Figure 2: Conceptual flow of FBA calculation pipeline in Python script

Initially, the list of metabolites was needed to be determined from the literature available. In essence, this process is done by inspecting each ingredient of the given media and determining suitable constituent metabolites in the acceptable BiGG database format. For example, yeast extract contains biotin, which can be referenced in the BiGG database as 'btn' as a metabolite for models of metabolic reconstruction. Upon the first attempt of trying to deconstruct the media based on the literature, it was virtually impossible to determine the correct metabolites used in the exchange reactions. The media definition process then changed to extracting all metabolites used in exchange reactions directly from the model and manually filtering through each metabolite to justify if each was present in the media ingredients. Once each media ingredient had been deconstructed into its constituent metabolites, the molar masses for each metabolite were retrieved from another online resource (chEBI, PubChem) for later use in estimating the flux calculation. It is worth noting that some of the molar masses of the metabolites that are defined in the model's exchange reactions were not clearly defined in these searches, and needed to be manually asserted to avoid ambiguity (ie. starch, capsular polysaccharide). Taking all of the metabolites present in the medium and their molar masses, the next step was to define the concentration of each metabolite in each ingredient, and further determine the concentration of each ingredient in each media. Using earlier mentioned resources (HyServe, Oxoid, and Thermo Fisher data sheets) the concentration of specific metabolites was recorded in a dataframe for the ingredients of interest in each of the 3 media.

With this information available, there was still a one-to-many relationship between metabolites in the media ingredient and metabolites in the media that was attempting to be defined. For example, some of the ingredients used contained a broadly defined "glucose" (BiGG ID: 'glc'), but several exchange reactions in the model consume glucose in multiple variants, such as glucuronate (BiGG IDs: 'glcur'). To solve this issue, we assumed that the concentration of metabolite within the ingredient could supply the medium with all variants of the metabolite used in the exchange reactions. With each metabolite in our medium having its corresponding molar mass and concentration in medium ingredients, the next step was to define the concentration of each metabolite in the medium based on the concentrations of ingredients of the media. For example, according to the data supplied, the mMCB media contained 6.5 g/L of bacteriological peptone, and various other concentrations of other ingredients. The concentrations of metabolites in *silico* media are scaled to match the proportions of the media ingredients they originate from. With the final set of information compiled, the result is a collection of metabolites, their molar masses, and their concentration within the media based on literature definitions, which is sufficient to calculate their estimated flux lower bound in metabolic exchange reactions in the bacteria given some other biological data as outlined in the previous section (*Methods, Flux Constraints*). The Python notebooks used for the FBA pipeline can be found in the acknowledgements section.

Results

F. prausnitzii's niche environment, the gut lumen, is characterised by low oxygen tension and high concentration of nutrients such as vitamins (Duncan *et al*., 2002). Likely through fermentation of glucose or other undigested (by enzymes) carbon sources, *F. prausnitzii* produces butyrate, one of the Short Chain Fatty Acids (SCFAs) of the intestine (other organic acids end-products of microbial fermentation are acetate and propionate). Butyrate is an energy source for the colonic epithelium and is beneficial for gut health, by reducing inflammation, oxidative stress and carcinomic risk in the colon (Hamer *et al*., 2007), (Louis and Flint, 2009). The availability of acetate in the culture medium has been previously shown to stimulate *F. prausnitzii* growth (Duncan *et al*., 2002). Other crucial elements for the growth of *F. prausnitzii* are B vitamins, and in particular riboflavin, which is commonly present in mammalian guts. Riboflavin is used by Faecalibacterium as an electron shuttle to grow in oxic-anoxic interfaces, such as the gut mucosa where low oxygen diffuses from epithelial cells (Khan *et al*. 2012).

By simulating in-silico the growth of the bacterium on the three media, we hoped to obtain values in the same order of magnitude as found in vitro and similar relative differences. We also expected our simulation to reflect the elements we have gathered from literature: consumption of acetate if present, production of butyrate, essentiality of vitamins for growth.

During the definition of the metabolite composition, results were hampered by undefined medium components such as peptones, liver and yeast extracts and meat extracts. Additionally, the fact that BiGG IDs are not unique and those in the medium need to correspond to the metabolite defined in the model sometimes caused inconsistencies.

Inspecting the metabolites used in the exchange reactions, some essential metabolites were determined based on an estimation of their presence in the available medium. For each metabolite used in the model's exchange reactions, one metabolite was removed from the medium completely, the upper and lower flux boundaries were set to -10 and 1000 respectively for the remaining metabolites, and growth was calculated. This process was repeated for every available metabolite, resulting in a list of metabolites that limited growth, based on growth in its absence (mean growth for all estimations: 9.924965). Using their objective functions as a guide, these metabolites are essential for the growth of *F.prausnitzii* and should be present in all media as growth is near-zero in their absence:

Metabolite	ID	Objective Function
Niacin	nac(e)	7.88E-13
L-Tryptophan	trp_L(e)	6.21E-15
Potassium	k(e)	3.82E-15
Cobalt	cobalt2(e)	3.60E-15
Copper	cu2(e)	3.60E-15
Folate	fol(e)	1.34E-15
L-Serine	ser_L(e)	-2.87E-15
Chlorine	cl(e)	-2.94E-15
Biotin	btn(e)	-7.17E-15
Calcium	ca2(e)	-1.53E-14
Pantothenate	pnto_R(e)	-8.22E-14
Sulphate	so4(e)	-8.60E-14
Pyridoxal	pydx(e)	-1.57E-13
Magnesium	mg2(e)	-1.71E-13

Table 1: Limiting metabolites for F.prausnitzii growth

After the bottom-up selection of metabolites, first preliminary results were obtained, including the selected metabolites and setting their fluxes to the arbitrary value of -10. The biomass functions for all media were unrealistically high. We therefore proceeded to define specific lower bounds to the fluxes for each

of the metabolites, applying assumptions and approximations mentioned in *"Flux Constraints"*.

The in-silico simulation shows that glucose is a limiting metabolite for all three media, and riboflavin is for mMCB and RCM. Final estimations of growth rate per minute are *2.17E-08* for mMCB, *1.85E-07* for mGAM, and *2.27E-05* for RCM. These growth rates are very small compared to the in vitro experimental results. Indeed, experimental results from the wet lab show an estimated maximum growth in the range *[0.00463, 0.00542]* for mMCB, *[0.00352, 0.00409]* for mGAM, *[0.00462, 0.00468]* for RCM. In silico, RCM is predicted to promote faster growth than mGAM, as expected, while the relative value for mMCB is off.

Discussion

The results that were obtained using the in-silico model are too low compared to the in-vitro growth curves. These discrepancies can be explained by not completely defined media. We would like to stress that knowing the exact composition of the medium is critical to predict growth rate in-silico. Among the biggest factors of uncertainty are B-Vitamins of which many were required for growth and in the current solution riboflavin was growth limiting for all three media. In order to improve on our model the concentration of B-vitamins in undefined medium components such as yeast/meat/liver extract could be investigated. Related to the additional large sources of error is the estimation of unknown metabolite concentration that were found to be limiting growth if excluded from the medium, these components got a flux of -0.001 assigned to still allow for growth but not overestimate their availability.

Another source of uncertainty next to the medium composition is in the calculation of the flux values. The exact dry-weight of the bacterium was not known (refer to the methods section for details), possibly we were overestimating the weight of *F. prausnitzii* resulting in underestimating the fluxes. The weight estimate could be improved by measuring the weight of the bacteria in addition to counting the number of the cells in in-vitro experiments.

Surprisingly, we detected consumption of acetate for the mGAM medium, which does not contain acetate. Upon investigating it was found that acetate was produced in all 3 media. Looking into the chain of reactions that lead to acetate being produced can improve our model by excluding the metabolite linked to the exchange reaction that is at the root of the production of acetate.

Additional assumptions that increase the uncertainty of the outcome are the assumption of steady-state of the metabolites of the bacterium, this steady state assumption would be justified when using a chemostat instead of a batch culture experiment to compare the in-silico results to.

Conclusion

In conclusion, the model might be more or less similar to reality but without a more complete medium definition we were not able to validate this. Further research needs to be done to improve the environment definition and to be able to judge the quality of the model.

Acknowledgements

We would like to thank Clémence Joseph for her guidance and patience and kindness throughout the project and both her and Karoline Faust for the review of this manuscript.

Supplements

The Python code for the FBA analysis can be found here: https://github.com/TimBlokker/Flux_Balance_F_Prausnitzii

References

Barry G. Hall, Hande Acar, Anna Nandipati, Miriam Barlow, Growth Rates Made Easy, Molecular Biology and Evolution, Volume 31, Issue 1, January 2014, Pages 232–238, https://doi.org/10.1093/molbev/mst187

Duncan SH, Hold GL, Harmsen HJM, Stewart CS, Flint HJ. Growth requirements and fermentation products of Fusobacterium prausnitzii, and a proposal to reclassify it as Faecalibacterium prausnitzii gen. nov., comb. nov. Int J Syst Evol Microbiol. 2002 Nov;52(Pt 6):2141-2146. doi: 10.1099/00207713-52-6-2141. PMID: 12508881.

Ebrahim, A., Lerman, J.A., Palsson, B.O. et al. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Syst Biol 7, 74 (2013). https://doi.org/10.1186/1752-0509-7-74

Hamer HM, Jonkers D, Venema K, Vanhoutvin S, Troost FJ, Brummer RJ. Review article: the role of butyrate on colonic function. Aliment Pharmacol Ther. 2008 Jan 15;27(2):104-19. doi: 10.1111/j.1365-2036.2007.03562.x. Epub 2007 Oct 25. PMID: 17973645.

Heinken, A., Khan, M. T., Paglia, G., Rodionov, D. A., Harmsen, H. J. M., & Thiele, I. (2014). Functional metabolic map of Faecalibacterium prausnitzii, a beneficial human gut microbe. *Journal of Bacteriology*, *196*(18), 3289–3302. https://doi.org/10.1128/JB.01780-14

Khan MT, Browne WR, van Dijl JM, Harmsen HJ. How can Faecalibacterium prausnitzii employ riboflavin for extracellular electron transfer? Antioxid Redox Signal. 2012 Nov 15;17(10):1433-40. doi: 10.1089/ars.2012.4701. Epub 2012 Jun 25. PMID: 22607129.

Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol Lett. 2009 May;294(1):1-8. doi: 10.1111/j.1574-6968.2009.01514.x. Epub 2009 Feb 13. PMID: 19222573.

Loferer-Krößbacher, M., Klima, J., & Psenner, R. (1998). Determination of bacterial cell dry mass by transmission electron microscopy and densitometric image analysis. *Applied and Environmental Microbiology*, *64*(2), 688–694. https://doi.org/10.1128/aem.64.2.688-694.1998

Marinos, G., Kaleta, C., & Waschina, S. (2020). Defining the nutritional input for genome-scale metabolic models: A roadmap. *PLoS ONE*, *15*(8), 1–17. https://doi.org/10.1371/journal.pone.02

Miquel, S., Martín, R., Rossi, O., Bermúdez-Humarán, L. G., Chatel, J. M., Sokol, H., Thomas, M., Wells, J. M., & Langella, P. (2013). Faecalibacterium prausnitzii and human intestinal health. Current Opinion in Microbiology, 16(3), 255–261. https://doi.org/10.1016/j.mib.2013.06.003

Orth, Jeffrey D., Ines Thiele, B. Ø. P. (2010). What is flux balance? *Nat. Biotechnol.*, *28*(3), 245–248. https://doi.org/10.1038/nbt.1614

Kevin D'hoe, Stefan Vet et. al.(2018). Integrated culturing, modeling and transcriptomics uncovers complex interactions and emergent behavior in a three-species synthetic gut community. DOI: 10.7554/eLife.37090