

Peer review of *Designing Microbial Communities to Maximize the Thermodynamic Driving Force for the Production of Chemicals*

The authors present an interesting algorithm for the computational design of single-species communities that maximise the thermodynamic driving force of producing a certain target metabolite. The paper is written in a clear and educational style, which makes it pleasant to read. The algorithm is potentially important, and has novelty for both the modelling of microbial communities and the design of bio-based production. Although I encountered some issues that I think should be resolved, I think the manuscript is fit for publication in Plos Computational Biology after some major revisions.

My main criticism is related to the strong and non-negligible assumptions that the ASTHERISC algorithm is based on. I understand that making these assumptions is necessary: the modelling of microbial communities is difficult, and some assumptions are needed to make computational problems feasible. However, I think the reader should be guided much more in interpreting the consequences of these assumptions. My remarks, ordered per section, can be found below.

Remarks Introduction

The example described in Figure 1

Although I appreciate the attempt to introduce the concept of DoL in a very basic way, the example model is a bit too basic for my taste. At a first glance, I am left with the idea that the assumptions that are made are necessary to draw the conclusions that the authors want to make. I think this is not the impression that the authors want to leave, because their argument is much more general. Also, the statement "Hence, it is unlikely that a DoL strategy will be advantageous with respect to the metabolic burden ..." is repeated a couple of times in the introduction, and I think it is not entirely correct. See my explanation below on how this could be nuanced.

The two assumptions that I find too specific are:

- the saturation functions only depend on the substrate of the reaction, rather than also on the product.
- the same amount of enzymes E1 and E2 is needed for a specific flux

I know that the authors already state that the second assumption does not affect the conclusions, but the reader is still left wondering what the effect of this assumption might be.

I would suggest two edits:

1. I think it is fairly straightforward to write a small mathematical proof that DoL does not give an advantage when the metabolite concentrations are unchanged, i.e., the saturation functions remain the same. This proof can be added to the SI, and can be referred to in this Introduction. I have added an outline for the proof below.

Adding this proof would also allow for a more exact statement. The authors now state: *"Hence, it is unlikely that a DoL strategy will be advantageous with respect to the metabolic burden ... "* But this should actually be something like: *"Hence, it is unlikely that a DoL strategy will be advantageous with respect to the metabolic burden, as long as the kinetic advantage of having different metabolite concentrations in the different strains do not outweigh the added costs of metabolite transport."* The statement is a bit less strong, but definitely more correct. Also, adjusting this statement would allow the authors to give a preview on the rest of their work, because adding a growth advantage by having different metabolite concentrations is exactly what they are going to do, albeit focusing on thermodynamics and not on kinetics.

2. In this Introduction I would at least remove the first assumption, and just make the saturation functions depend on all metabolite concentrations. This does not change the reasoning at all. Personally, I would also skip the second assumption, but I realize that this is a matter of taste. Generalising this here might complicate matters. It is also not really necessary as long as you can refer to the general proof in the SI.

An idea for the proof:

I wanted to write a quick outline for the proof, but it got a bit more involved than I expected. I hope the proof below is useful for the authors. Please check everything carefully, because I did not.

One could start by taking a general reaction vector, v , that leads to some production of P . We can multiply v by a constant and still have the production pathway, and the production rate of P is proportional to the reaction rates in v . Let \tilde{v} now be rescaled such that $\tilde{v}_P = 1$. The enzyme costs of this production pathway can be expressed in terms of some (arbitrary) saturation function $f_i(S, M, P, \dots) =: f_i(x)$ for each reaction i . The enzyme costs for having $v_P = 1$ are then

$$\sum_i e_i = \sum_i \tilde{v}_i / (k_{\text{cat},i} f_i(x)).$$

So, given a limited amount of resources ($\sum_i e_i \leq C$) we can get a maximal production flux of

$$v_P = \frac{C}{\sum_i \tilde{v}_i / (k_{\text{cat},i} f_i(x))}.$$

Note however that this is the production flux per gram biomass. The total biomass may also be constrained by $B \leq C_2$. The absolute production flux for a single strain is thus

$$J_P = C_2 \frac{C}{\sum_i \tilde{v}_i / (k_{\text{cat},i} f_i(x))}.$$

Now, we can divide the reaction pathway over two strains. One could for example take a v^1 and v^2 , such that $v^1 + v^2 = v$, where one should note that the metabolites should be transported in between, and this should be written down more carefully than I did here. Again, let \tilde{v}_1 and \tilde{v}_2 be normalised such that $\tilde{v}_1 + \tilde{v}_2$ produces $v_P = 1$. Again, we have an enzyme constraint per strain, so we get maximal fluxes per gram biomass of

$$\lambda^1 = \frac{C}{\sum_i \tilde{v}_i^1 / (k_{\text{cat},i} f_i(x))},$$

and

$$\lambda^2 = \frac{C}{\sum_i \tilde{v}_i^2 / (k_{\text{cat},i} f_i(x))},$$

where λ^j denotes the number of times that \tilde{v}_1 can run per gram biomass of strain 1. Since we need to add \tilde{v}^1 and \tilde{v}^2 in equal amounts to get the production of $v_P = 1$, one needs

$$B^1 \lambda^1 = B^2 \lambda^2,$$

with $B^1 + B^2 = C_2$. This means that $B^1/B^2 = \lambda^2/\lambda^1$, and inserting this in the total biomass constraint gives

$$C_2 = B^1 + B^2 = B^2 \frac{\lambda^2}{\lambda^1} + B^2 = B^2 \left(1 + \frac{\lambda^2}{\lambda^1}\right),$$

so that

$$B^2 = \frac{C_2}{\left(1 + \frac{\lambda^2}{\lambda^1}\right)}.$$

The total production is given by

$$J_P = B^1 \lambda^1 = B^2 \lambda^2 = \left(\frac{C_2}{\left(1 + \frac{\lambda^2}{\lambda^1}\right)} \right) (\lambda^2) = \frac{C_2}{1/\lambda^1 + 1/\lambda^2},$$

where we can insert the expression for the maximal possible λ^1 and λ^2 to get:

$$J_P = \frac{C_2 C}{\sum_i \tilde{v}_i^1 / (k_{\text{cat},i} f_i(x)) + \sum_i \tilde{v}_i^2 / (k_{\text{cat},i} f_i(x))} = \frac{C_2 C}{\sum_i \tilde{v}_i / (k_{\text{cat},i} f_i(x))},$$

which is exactly equal to the production that could be achieved with DoL.

Remarks Methods

The Methods-section is clear and covers all prerequisite knowledge without becoming excessively long or boring. The pseudocode is written very clearly, and will definitely help potential users of ASTHERISC. Several strong assumptions are made in the Methods section, and I feel the following should at least be justified a lot more:

Equation (3): $\bar{N}r = 0$

Of course, this seems like a standard assumption that is always made in stoichiometric modelling. However, in this paper the authors are considering the thermodynamic effects that in part depend on metabolite concentrations. The correct formula to use is $\bar{N}r = \mu c$, where μ is the growth rate, and c is the vector of metabolite concentrations. This dilution term implies that all metabolites should have some net production when the growth rate is positive. However, is this small net production still thermodynamically feasible in the solutions that ASTHERISC finds? ASTHERISC might just select a really high metabolite concentration, and it will detect no problem because there does not have to be any net production of this metabolite. I wonder if the authors foresee any problems, or if this issue is negligible. This might be important to add in the text. It is true that this issue is resolved by the following assumption that the growth rate is 0, but this might just be the issue: the dilution of metabolites is one of the reasons that I would say it is unreasonable to set the growth rates to 0.

The growth rate of all strains is 0

Now this is a rather strong assumption to make, and, as a reader, I really need to be convinced that this is justified. For example, if ASTHERISC assumes that biomass production is zero, how do I know that the algorithm doesn't choose a division of reactions that makes biomass production impossible for one of the strains? For example, it might be thermodynamically ideal to have a high concentration of ATP in one strain, and a low concentration in another. But can the strain with low

ATP still grow? Do the biomass-producing reactions not become thermodynamically infeasible? When an organism is unable to grow, is the chosen division of labor still useful for biotechnological applications?

My suggestion is that the authors explicitly state these type of problems. Finding all possible Divisions of Labors is still useful, if they can afterwards be tested for their feasibility. For example, testing if growth is still possible, and if the suggested metabolite exchanges are possible (as the authors do point out in the Discussion). I would like to see for which multi-strain solutions presented in the paper, growth at these metabolite concentrations is still possible.

All strains have identical fractions

Just like in the example in the Introduction, this assumption feels very limiting, but it might not be. Maybe the authors can justify this assumption, but this is not done at the moment. Now, the reader has to think about the consequences of this assumption, and I think that this is actually a task for the authors. My suggestion is that the authors could point out that they focus on the thermodynamic feasibility of a pathway, and not on resource-allocation. Maybe, the multiplication of all reaction rates of one strain is therefore allowed, and then the biomass fraction of the strain does not matter.

To get a feel for the consequences of this assumption, I suggest that the authors write down the full community model without assuming identical fractions (possibly in the SI). Then, the authors may look at the consequences of multiplying the biomass fraction of strain i with a scalar: $F_i \rightarrow \lambda_i F_i$. It will turn out that to keep satisfying the steady-state constraint, all reaction rates of strain i must then be divided by λ_i . It can then be seen where this will have consequences. I suspect the most important influence is on *satisfying the inhomogeneous flux bounds*. The authors may then explain why this is acceptable.

Probably, the authors can argue that leaving this flexibility of choosing the biomass fractions out only takes away flexibility from the model. So, something that is thermodynamically feasible in this constrained case, will also be feasible in the general case. This seems to be true, as long as the authors do not consider kinetics and limited cellular resources.

Remarks Results

The results of applying ASTHERISC to example models are presented. The models are well-chosen to illustrate the workings of the model, and the results are exposed in a clear and un-assuming fashion. The example of the community solution serves its purpose: it illustrates the principle underlying ASTHERISC, and its complexity shows why these results cannot be obtained without a computational toolbox.

Page 15

"Reactions for which no $\Delta_r G'^0$ could be found or calculated (93 reactions (18%) in EColiCore2 and 646 (24%) in iML1515) were blocked in the model (rate fixed to zero)" There are several things I'm wondering about after reading this comment. Why is the chosen solution of blocking these reactions the most reasonable? What are the consequences? Is there any particular property of these reactions which explains why the standard Gibbs free energy cannot be found or calculated?

Bottom page 15

"Interestingly, the highest MDF improvements occur when high product yields are demanded. This indicates that low-yield pathways in the single strain might have high MDFs which are more difficult to improve by a community solution." I agree that this is an interesting note.

Exemplary community solution

The authors state that "kdo8p is of interest as its synthase (KDOPS) has been evaluated as a potential antibiotics target". This does not sufficiently explain to me why kdo8p is of interest. How can the over-production of kdo8p be used to inhibit its synthase? It is not that I do not believe that the authors are correct, I just need a little more explanation.

I don't understand why the concentration of dha is decreased in *ecoli2*. The authors state that it is used to facilitate the (reverse) PGI, G6PDH2r and PGL reactions, but I don't see this. dha seems to only be used as a substrate in *ecoli2*, so why would it not be favourable to maximise its concentration, so that g3p-concentration can be decreased even more. Similarly, one would expect dha to be decreased in *ecoli1*.

Remarks Discussion

The discussion is clear and addresses some important points. Especially the paragraph starting with "The application of our algorithm to a metabolic core and a genome-scale model" is very important. Some additional questions that I had with the eventual application of the algorithm:

- I agree that reactions can be knocked-out, but will the optimisation of metabolite concentrations happen automatically? The genetically altered organism has not been evolved to use its new reaction set efficiently, so how do we know that the metabolite concentrations that make the production pathway thermodynamically feasible will be selected?
- Is there a way to add the energy expenditure needed for exchanging metabolites into the algorithm? I guess, maybe naively, that when a metabolite is exported from a low concentration, and imported to a high concentration, at least one of the transport steps needs to use energy. This means that it should at least be thermodynamically feasible for this strain to produce energy.
- Is there a way in ASTHERISC to select the metabolites that may be exchanged by the algorithm? This does not seem hard to implement.

In addition, I would like a discussion of the kinetic analog of this thermodynamic problem. Except for making pathways thermodynamically feasible, Division of Labor could also help with the efficient usage of cellular resources. Indeed, as the authors show in the Introduction, not when the metabolite concentrations are unchanged, but what if these concentrations are changed. Can this problem also be addressed? If not, why not? Or is it just less important according to the authors?

Minor comments

Figure 1 I think Figure 1 is rather messy. A reader that did not understand the explanation in the text, will probably not be aided by this figure a lot. This is a matter of taste though, not of right and wrong.

Caption Figure 2 Add comma: "would follow, indicating"

After Equation 5 Sentence starting with "Furthermore, the biomass fractions" is pretty hard to read.

Equation 9 It is not immediately clear whether the i -th column is taken before or after the transpose

Equation 10 I fail to see why the notation is different between Equations (9) and (10): $\Delta_r G'_i \rightarrow \Delta G_i$ and $\hat{N} \rightarrow N$. These might be typos. If not, more explanation might be needed.

Page 11: "A specified minimal fraction of this maximum yield must be reached in all subsequent community model calculations" Can this fraction also be chosen larger than 1? Or is the maximum possible yield determined without taking into account the thermodynamic feasibility? If so, please indicate this in the text.

Page 14 Typo in: "had to be separated in to strains"

Page 15 Typo: "For each of the producible target metabolite"

Page 17 "Some of these dominant MDF values occurred in all analyzed models." So, there are MDF values that appear often in the core- and in the genome-scale model? If so, it might be interesting to explain a bit more why this is.

The use of "exemplary" It might be that I'm wrong here, but it seems that the word 'exemplary' is used meaning that something is an example, but my interpretation of this word is that it is "a perfect example". It therefore seems a bit like bragging to me when this word is used.

Exemplary community solution "In the community solution found by ASTHERISC, the MDF value can be increased by 26 % up to 2.87 kJ/mol reaching a yield of 0.733 mol/mol" It confuses me that this yield is here repeated, it seems like this yield is different than the yield of the single-strain model. But this was exactly the yield that was requested, right? So, the reported MDF value for the single-strain model also attains this yield?