RE: Revisions to the manuscript *A compact model of Escherichia coli core and biosynthetic metabolism* (round 2)

Dear Editor

We would like to thank you for the valuable feedback. We have revised our work to address the additional comments resulting from the first round of reviews. With the following, we would like to summarise the changes since the previous version.

Sincerely,

Marco Corrao
on behalf of all the authors

Answers to review comments

Figure 3: The caption (and text) refers to the EcoCyc database, but panel A refers to the Biocyc database

While the vast majority of Biocyc annotations in the model are in the ECOLI namespace (and hence map to EcoCyc), there are two that are in the META namespace and hence map to the broader MetaCyc database (also part of the BioCyc ecosystem). The caption for Figure 3A was therefore corrected by replacing *EcoCyc* with *Biocyc.* Further, the presence of the two reactions mapping to the MetaCyc namespace was made explicit in the text (see the answer to the next comment).

In your article you wrote: "Notably, the annotations pointing to EcoCyc, the main knowledgebase for E. coli, are complete: every metabolic reaction in iCH360 is uniquely mapped to a corresponding EcoCyc ID (with the exception of three reactions, for which a match on the database could only be found for a different choice of redox cofactor, namely NADH instead of NADPH)." ===> This can be misunderstood, because if it misses three reactions, it is not complete. Would "nearly complete" be a better? Or are those three reactions mapped to the reaction with the different cofactor (then using another term than "with the exception" could clarify this).

We agree that this aspect was not very clear in the previous version of the manuscript and that the phrase *nearly complete* would be more correct. In the revised manuscript, this paragraph has been corrected and extended as follows:

Notably, the annotations pointing to the BioCyc knowledgebase [22], are nearly complete: Out of 321 enzymatic reactions in the model, 317 are mapped to Biocyc with a single ID (for the remaining four unannotated reactions, all involved in the biosynthesis of unsaturated fatty acids, a match in the database could not be found for the specific use of NADPH as a redox cofactor). Further, nearly all of these Biocyc annotations (315 /317) are in the ECOLI namespace and therefore point to the organism-specific EcoCyc database, a widely used and extensive reference for E. coli molecular biology [25, 26]. The remaining two reactions map instead to the broader MetaCyc database, also part of the the BioCyc ecosystem, via the META namespace.

Note that, during the review process, we realised that the unannotated enzymatic reactions are actually four, rather than three (even though the distinction is mostly formal, since these reactions all represent the same biochemical transformation, but performed on hydrocarbons with different chain length) , and corrected this accordingly.

In the new version you have added a nice extension of the production envelops and also a short explanation in the Methods. You state you use cobraPy, and in the legend of Fig. 2 you state you limit the glucose uptake flux and maximize the production of different compounds. I

assume you fixed the growth rate and then did the pFBA to maximize the compound production flux, but a short sentence in the Methods could clarify this.

We have now extended the production envelopes section of the Methods (Section 4.3) to briefly describe the algorithm used by COBRApy to generate the production envelopes, and the constraints we imposed on the model for this analysis. The revised Section is reported below:

All production envelopes shown in the main text and Supplementary Information were generated using the built-in production envelopes tools from the cobra.flux analysis.phenotype_phase_plane module in the COBRApy package [17]. Briefly, the algorithm first computes the maximum and minimum production rates of the metabolite of interest, given the existing constraints in the model (including the specified bounds on the uptake of the carbon source and oxygen). The interval between the maximum and minimum achievable production rates is then discretised into an equally spaced grid of points. For each point in the interval, the production rate is fixed and the model's objective (here, the growth rate) is sequentially maximised and minimised, thus generating the boundary of the production envelope. All prediction envelopes were computed by specifying an upper bound on the uptake of the carbon source of 10 mmol/gDW/h, and blocking oxygen uptake for the anaerobic scenario. For comparisons with ECC and ECC2, the maintenance requirement (lower bound on the ATPM reaction) of these two models was set to the same value used in iML1515 and iCH360 (6.86 mmol/gDW/h).

The use of in text citations could enhance the readability. For example, you write:"obtained via competitive fitness assays, from [25]." & "turnover numbers from [29]." => It might be nicer to use an in text citation here mentioning the authors/year (and the number)

The style of in-text citation has been updated as suggested (for example, the revised text now reads: "[...] ,*obtained via competitive fitness assays, from Price et al. (2018) [27*]")

Also the references could benefit from an improved style, e.g. removing the superfluous information (e.g. visiting date, double URLs and publisher information for journals), and mentioning all authors instead of just the first.

Information about URLs, visiting date, and publisher has now been removed from the bibliography. We thank the editor for noting and pointing out these redundancies.