Editorial office
PCI Math Comp Bio

December 12, 2024

Manuscript review: A compact model of *Escherichia coli* core and biosynthetic metabolism

Dear Prof. Meike Wortel, Dear Authors:

The paper by Corrao *et al* introduces *i*CH360, a medium-scale metabolic model of *Escherichia coli*, and evaluates its predictive capabilities. The authors highlight that genome-scale models offer broad insights, but have limitations that smaller models can address. However, existing small-scale *E. coli* models often rely on algorithmic reductions and require extensive manual curation. To bridge this gap, this study aimed to create a well-curated medium-scale model.

*i*CH360 integrates diverse datasets, including gene-protein-reaction annotations, protein complex composition, thermodynamics, and kinetic constants. The model demonstrates predictive accuracy comparable to the genome-scale model, *i*ML1515, while enabling elementary flux mode analysis, which is computationally infeasible for larger models. The study also applies thermodynamic and kinetic constraint methodologies, underscoring *i*CH360's potential for future applications.

The abstract presents a clear overview of the study. The introduction provides a good background on the developments made in modelling *E. coli* and provides a strong context for the introduction of the main motivations and aims for developing *i*CH360. The methods are comprehensive and provide a clear written description of the work completed. The assumptions made by the authors are clearly stated (in the methods, results or supplementary materials) and seemed sound. Mathematical notation was supported with a well written commentary. Furthermore, scripts available online provide a comprehensive guide on the actual implementation of the methods. The discussion brings everything together and provides a broad summary of the results, whilst highlighting potential extensions and applications of *i*CH360. Throughout the paper, all references seemed relevant and accurate.

The authors invested great effort into writing a clear commentary and guide for executing the scripts. I downloaded the Zenodo repository (`https://doi.org/10.5281/zenodo.13939696`, on 2024-11-28) and installed the required packages. I had to install a number of additional packages and enclose the scripts I used for environment setup. They may be of use for future users of the repository. I was able to run all scripts (with some minor changes, see below) and reproduce all results and plots presented in the paper. I did not run the `count_efs.m` script due to the MATLAB requirement and instead used the `efm_counts_unfiltered.csv` file included in the repository.

Some general points:

In my opinion, the findings could be reinforced in the abstract; that the predictions with *i*CH360 were in-line with *i*ML1515 and/or experimental data; that the smaller scale of the model enabled the application of more advanced methodology; and that *i*CH360 is *better* than previous small-scale models. Additionally, considering the paper focusses on the development of *i*CH360, it should be named in the abstract.

The approach to determine an equivalent biomass reaction is attractive and well explained. There are two biomass reactions in *i*ML1515: core and WT [1]. I assume that the WT biomass reaction was used but the authors should clarify this. Would the precursor fluxes change significantly when using

the other reaction? Additionally, how were energetic requirements accounted for in *i*CH360 and were the chosen values inherited from *i*ML1515? In addition, the stoichiometry of the equivalent biomass reaction should be made easily available as part of the supplementary material.

The authors highlight other small-scale models (ECC and ECC2) and include them in analysis of model properties. I believe further comparisons would be valuable in supporting *i*CH360. Is *i*CH360 better? How is it different? For example, simple predictions (FBA, phenotype phase planes) could be discussed. Whilst it is not a fair comparison for all carbon sources (only growth on glucose, glycerol, acetate and succinate were protected in ECC2 generation [2]), these comparisons could highlight the advantages of *i*CH360 being able to predict growth on a wider range of carbon sources.

Relatedly, the authors could further investigate the consistency of predictions between *i*CH360 and *i*ML1515. Whilst it does not necessarily validate the model predictions, the authors could compare fluxes (e.g. from pFBA) of reactions shared by both *i*CH360 and *i*ML1515 across the different carbon sources investigated.

Where possible, the sources/literature used for defining the different types of catalytic edges should be cited or made available in a supplementary file.

The authors approximate enzyme abundance by constructing an augmented matrix $\hat{E}$, allowing one to account for polypeptides that are part of additional enzymatic complexes which are not part of the model. How often are these polypeptides mapped out of the model? Based on the polypeptides and the in-model complexes they map to, are there certain types of complexes particularly affected by this and is this of interest?

Finally, the authors could expand on their interpretation of the results from the thermodynamic-based analysis.

Overall, this is a solid piece of work, that is well-written and with good accompanying code. With the additional validations mentioned, I would agree with the authors that "*i*CH360 holds the potential to become a reference metabolic model for *E. coli* ".

References

[1] Jonathan M. Monk et al. "iML1515, a Knowledgebase That Computes Escherichia Coli Traits". In: *Nature Biotechnology* 35.10 (Oct. 2017), pp. 904–908. DOI: 10.1038/nbt.3956. URL: https://www.nature.com/articles/nbt.3956.

[2] Oliver Hädicke and Steffen Klamt. "EColiCore2: A Reference Network Model of the Central Metabolism of Escherichia Coli and Relationships to Its Genome-Scale Parent Model". In: *Scientific Reports* 7.1 (Jan. 3, 2017), p. 39647. DOI: 10.1038/srep39647. URL: https://www.nature.com/articles/srep39647.

A number of minor issues found when running the scripts:

1. `primary_seondary_counting.Rmd` was empty. I assume it is an unused file.
2. In `enzyme_allocation_predictions.ipynb`, the path is written as `/manuscript_figures/` but should be `/Manuscript_Figures/`.
3. The path to the directory `Knowledge_Graph` was written as `Knowledge_graph` in `generate_model_tables.ipynb`, `catalytic_disruption_analysis.ipynb` and `estimate_enzyme_abundances_from_pp_counts.ipynb`.
4. `Flux-force-efficacy_vs_measured_enzyme_abundance.R` and `MDF_PTA.ipynb` initially failed due to usage of an unavailable file `../../Analysis/PTA/out/pta_reactions_data.csv`. Replacing the path with `../../Analysis/PTA/out/pta_fluxes.csv` fixed this issue and allowed for results and figures to be produced.
5. For `compute_efm_cost_yield.ipynb` to work, I had to create a directory `mkdir Analysis/EFM_growth_yield_screening/out`.

Environment setup

```
conda create -n ich360
conda activate ich360
conda config --add channels conda-forge
conda config --set channel_priority strict

export GRB_LICENSE_FILE=/path/to/gurobi/11.0.1/gurobi.lic
export GUROBI_HOME=/path/to/gurobi/11.0.1/
export PATH="${PATH}:${GUROBI_HOME}/bin"
export LD_LIBRARY_PATH="${LD_LIBRARY_PATH}:${GUROBI_HOME}/lib"

conda install -c conda-forge -c gurobi python==3.9.20 cobra==0.29.0 \
    numpy==1.24.1 scipy==1.10.1 pandas==1.5.3 matplotlib==3.7.1 \
    seaborn==0.12.2 networkx==3.0 tqdm==4.65.0 requests==2.28.2 \
    casadi==3.6.3 cvxpy==1.5.2 equilibrator-api==0.4.7 nb_conda_kernels==2.5.1 \
    notebook==7.1.3 pyvis==0.3.1 r-base==4.3.3 r-dplyr==1.1.4 r-ggplot2==3.5.1 \
    r-reshape2==1.4.4 r-ggsci==3.2.0 r-ggpubr==0.6.0 rpy2==3.5.11 gurobi==11.0.1 \
    adjusttext==1.3.0 r-svglite==2.1.3

pip install equilibrator-assets==0.4.1 efmtool==0.2.1 enkie==0.1.3 \
    straindesign==1.13 pta==0.6.0
```

In addition to setting up the conda environment, there were a number of additional steps needed. I include them here for reference.

1. For first time use of enkie and to run `pta.ipynb`, I needed to create a directory for it in my cache `mkdir /lisc/user/coltman/.cache/enkie`. This issue was previously reported https://gitlab.com/csb.ethz/enkie/-/issues/1
2. Due to the versions fixed by the package requirements, there were some issues with the setup of equilibrator. I had to manually save the files from the following repos (https://zenodo.org/records/4128543, https://zenodo.org/records/4013789, https://zenodo.org/records/4010930) to my `~/.cache/equilibrator` in order to run `drg0_estimation.ipynb`