Filling gaps in metabolism using hypothetical reactions

Citation for the original published paper (version of record):
Li, F. (2022). Filling gaps in metabolism using hypothetical reactions. Proceedings of the National Academy of Sciences of the United States of America, 119(49).
http://dx.doi.org/10.1073/pnas.2217400119

PNAS COMMENTARY

Filling gaps in metabolism using hypothetical reactions

Feiran Li[a,1]

Metabolism is crucial for all living cells since it provides energy and molecules for all biological functions. Systematically understanding metabolism is therefore important both in medical research and in synthetic biology where it can help to better engineer cells. Over the last decade, researchers have built genome-scale metabolic models (GEMs) to systematically simulate the complete known metabolism of the organism of interest. However, there remain many knowledge gaps in these models such as unannotated and misannotated genes, promiscuous enzymes, unknown reactions and pathways, and underground metabolism. Gaining insight into the unexplored metabolism is essential because a detailed understanding of cellular functions drives biomedical applications such as drug-targeting strategies. On the other hand, exploring the knowledge gaps in metabolism is also necessary for synthetic biology. For example, knowing the complete synthetic pathways enables efficient design of cell factories to produce chemicals and pharmaceuticals, especially valuable secondary metabolites (1). Thus, in order to systematically identify and reconcile the metabolic gaps at a genome scale, Vayena et al. (2) propose a computational gap-filling workflow, Network Integrated Computational Explorer for Gap Annotation of Metabolism (NICEgame).

Vayena et al. (2) applied this gap-filling workflow NICEgame to identify and reconcile the knowledge gaps in the latest *Escherichia coli* GEM iML1515 (3). They compared the model prediction and experimental phenotype of *E. coli* single-gene knockouts in glucose minimal media and identified metabolic gaps for 148 false gene essentiality predictions linked to 152 reactions. The workflow NICEgame can propose alternative reaction sets as gap-filling solutions to reconcile the false essential gene predictions.

Earlier gap-filling methods rely on biochemical reaction databases or published GEMs as reaction pools for gap-filling (4). Solutions suggested by those algorithms are limited within the scope of known biochemical reactions. There may exist only a unique solution for filling the same gap, leading to much more identical biased metabolic networks among diverse organisms, especially for those organisms which are poorly annotated and contain a large part of knowledge gaps. The NICEgame workflow relies on a much more extensive reaction database, ATLAS of Biochemistry (5), consisting of known and broader hypothetical reactions built from mechanistic understandings of enzyme function mechanisms, which provides more possibilities for the knowledge gaps and enables the identification of new biochemical capabilities and enzyme functions, ensuring more knowledge gaps to be filled and more gap-filling solutions to be found.

In the case study of *E. coli*, Vayena et al. (2) identified that the average number of solutions per rescued reaction is 252.5 when using ATLAS as the reaction pool versus 2.3 when using the KEGG reaction database, a resource covering known biochemical reactions. The comparison was performed by constraining both reaction pools within the scope of *E. coli* and yeast metabolites. Moreover, 53 of the total identified 152 false essential reactions were reconciled with thermodynamically feasible gap-filling solutions when using the KEGG reaction database as the gap-filling reaction pool. In comparison, 93 of 152 false essential reaction gaps can be rescued using the subset of ATLAS. Besides that, compared with the earlier gap-filling methods, NICEgame outputs alternative subsets, allowing users to evaluate those subsets based on biological domain knowledge (Fig. 1).

With more subsets being proposed for each targeted gap, the question remains of which one should be added to the model. Vayena et al. (2) adopted a scoring system to rank the reaction subsets by considering the thermodynamic feasibility and the minimum impact on the model.

The introduction of longer paths, new metabolites, and novel enzyme functions (when the third level EC number does not exist in the original GEM) was panelized.

Annotating genes for the proposed reactions for gap-filling is extremely useful, which drives further experiments to identify novel metabolic discoveries. In the NICEgame workflow, Vayena et al. (2) adopted BridgIT (6), a previous tool developed in the Hatzimanikatis group, to identify the enzymes associated with reactions for gap-filling. Reactions annotated with enzymes of higher BridgIT confidence scores were favorable.

In their paper published in this issue of PNAS, they proposed 77 new reactions associated with 35 *E. coli* genes to extend the latest *E. coli* GEM iML1515 to reconcile 47% of the 148 identified false essential gene predictions. Among these 35 genes, 33 were present in the original GEM iML1515. The added new reactions show the substrate or mechanism promiscuity of these 33 genes. Two new genes, ArcA and LacA, which were not part of the original reconstruction, have been assigned reactions and added to the model. In total, the added biochemistry reconciles metabolic gaps linked to the amino acid metabolism, cofactor metabolism, and biosynthesis of cell membrane peptidoglycans. The performance of the extended GEM of *E. coli*, iEcoMG1655, was validated on the gene essentiality experimental data on 15 carbon sources (3). iEcoMG1655 showed a 23.6% accuracy increase in gene essentiality predictions compared with the original GEM iML1515.

Author affiliations: [a]Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg SE-412 96, Sweden

Author contributions: F.L. wrote the paper.

The author declares no competing interest.

See companion article, "A workflow for annotating the knowledge gaps in metabolic reconstructions using known and hypothetical reactions," 10.1073/pnas.2211197119.

[1]Email: feiranl@chalmers.se.

Published November 29, 2022.

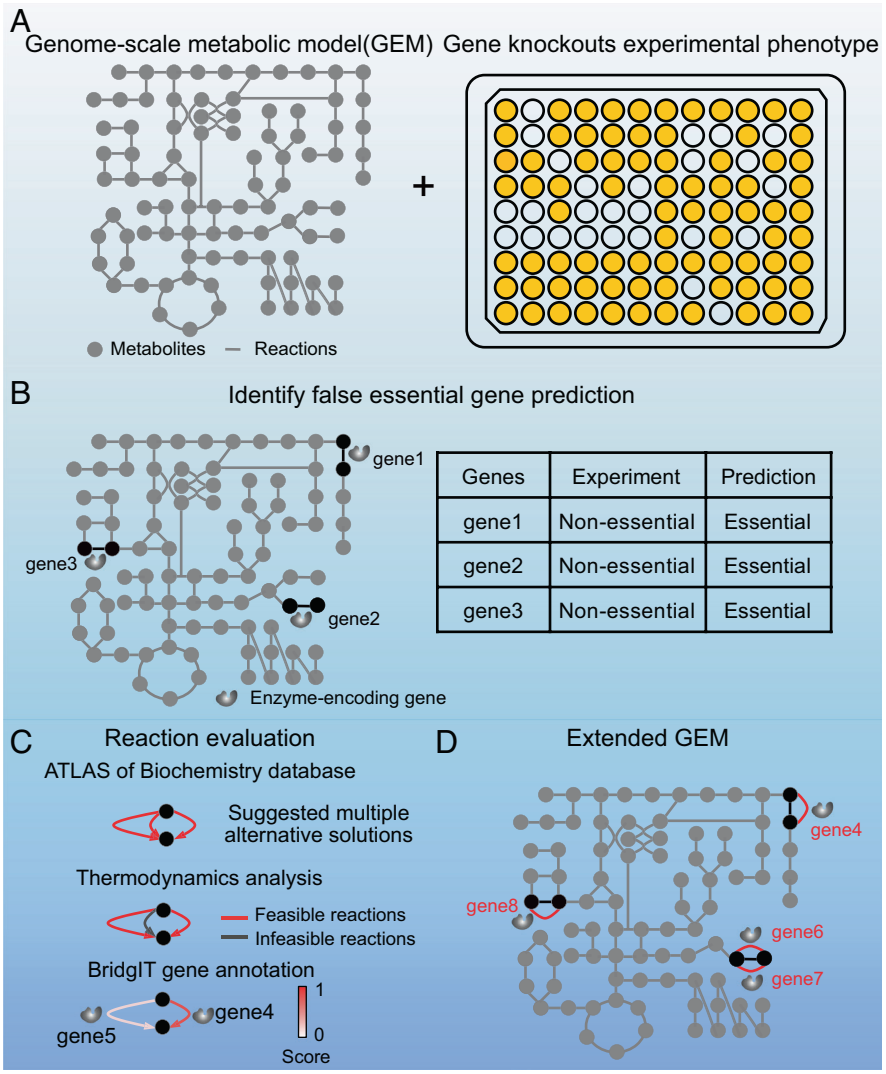

Fig. 1. NICEgame workflow for gap-filling using hypothetical reactions. (*A*) The input of the workflow is the GEM with the experimental gene knockout phenotype. (*B*) Identification of metabolic gaps. (*C*) The original GEM is merged with ATLAS, and an essentiality analysis is performed in the original and the expanded network to identify which gaps can be rescued. Alternative reaction sets are generated and evaluated. Gene associations for those proposed reactions are annotated with the BridgIT tool. Genes with higher BridgIT score were kept. (*D*) The extended curated GEM.

Note that metabolic phenotype data are very important to boost the gap-filling analysis and validate the performance. With the cheaper and faster methods for high-throughput phenotyping technologies and omics measurement, this workflow can be foreseen to generate more and better hypotheses to systematically identify and reconcile the metabolic gaps at the genome scale for other organisms.

"In order to systematically identify and reconcile the metabolic gaps at a genome-scale, Vayena et al. propose a computational gap-filling workflow, Network Integrated Computational Explorer for Gap Annotation of Metabolism (NICEgame)."

Besides that, NICEgame suggested over 7,000 novel and known reactions during the gap-filling process, among which around 6,000 reactions were determined as thermodynamically feasible and were annotated with *E. coli* genes. Previously, there were several attempts to explore the *E. coli* underground metabolism either deriving from experimental reports (7) or by probing isoenzyme candidates from omics data of gene knockouts (8). In this study, NICEgame, as a high-throughput approach, suggested 6,118 reactions associated with 590 candidate promiscuous enzyme-encoding genes in the *E. coli* genome, which validates the capability of NICEgame in systematically exploring the vast unknown underground metabolism of an organism.

Taken together, the NICEgame workflow, which consists of the extensive known and hypothetical reaction database ATLAS of Biochemistry and the gene annotation tool BridgIT, has shown great potential for exploring metabolic knowledge gaps, gene–protein–reaction associations, and enzyme or substrate promiscuity.

ACKNOWLEDGMENTS. My research is supported by the European Union's Horizon 2020 research and innovation program [No. 686070]. I thank Jens Nielsen for his valuable advice during the preparation.

1. M. R. Andersen *et al.*, Accurate prediction of secondary metabolite gene clusters in filamentous fungi. *Proc. Natl. Acad. Sci. U.S.A.* **110**, E99–E107 (2013).
2. E. Vayena *et al.*, A workflow for annotating the knowledge gaps in metabolic reconstructions using known and hypothetical reactions. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2211197119 (2022).
3. J. M. Monk *et al.*, iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* **35**, 904–908 (2017).
4. S. Pan, J. L. Reed, Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries. *Curr. Opin. Biotechnol.* **51**, 103–108 (2018).
5. J. Hafner, H. MohammadiPeyhani, A. Sveshnikova, A. Scheidegger, V. Hatzimanikatis, Updated ATLAS of biochemistry with new metabolites and improved enzyme prediction power. *ACS Synth. Biol.* **9**, 1479–1482 (2020).
6. N. Hadadi, H. MohammadiPeyhani, L. Miskovic, M. Seijo, V. Hatzimanikatis, Enzyme annotation for orphan and novel reactions using knowledge of substrate reactive sites. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 7298–7307 (2019).
7. R. A. Notebaart *et al.*, Network-level architecture and the evolutionary potential of underground metabolism. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11762–11767 (2014).
8. G. I. Guzmán *et al.*, Model-driven discovery of underground metabolic functions in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 929–934 (2015).