iMet: A graphical user interface software tool to merge metabolic networks

Reza Mohammadi a,∗, Javad Zahirib, Mohammad Javad Niroomandc

a Malek Ashtar University of Technology, Iran
b Bioinformatics and Computational Omics Lab (BioCOOL), Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran
c Learning Intelligent Systems Lab, School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran

ARTICLE INFO

Keywords:

Bioinformatics
Biotechnology
Microbiology
Systems biology

ABSTRACT

Nowadays, studying microorganisms has become faster and deeper than the last decades, thanks to the modeling of genome-scale metabolic networks. Completed genome sequencing projects of microorganisms and annotating these sequences have provided a worthwhile platform for reconstructing and modeling genome-scale metabolic networks. The genome-scale metabolic network reconstruction is a laborious and time-consuming task which needs an extensive study and search in different types of databases. Furthermore, it also requires an iterative process of creating and curating the obtained network, particularly with experimental methods. Hence, different types of reconstructions and models of a targeted microorganism can be found with different qualities, as the goal and need of researchers differ.

Due to these circumstances, scientists have to continue with only one of the reconstructed metabolic networks of each microorganism and ignore the rest in their in silico works. It is clear that having a tool which merges different metabolic networks of a single organism can be a useful and effective way to study them with minimal cost and time. To meet this need, we have developed iMet, the standalone graphical user interface (GUI) software tool to merge multiple reconstructed metabolic networks of microorganisms. As a case study, we merged three reconstructed metabolic networks of a cyanobacterium using iMet, and then all of them (including the new merged one) became modeled. The results of our evaluations including Flux Balance Analysis (FBA), revealed enhancing metabolic network coverage as well as increasing yield of desired products in the new obtained model.

1. Introduction

The shifting attitude in biology from classical molecular biology to the systems biology has led to dramatic and profound changes in the study of living organisms. With the completion of multiple genome sequencing projects in recent decades, our insight into the understanding of genotype-phenotype relationships has been increasingly developed (Lewis et al., 2012; Mardinoglu and Nielsen, 2015). The availability of whole genome sequences has enabled scientists to simulate appropriate models of metabolic networks so they can predict and manage the behavior of microorganisms with higher efficiency (Feist et al., 2009; Henry et al., 2006). Genome-scale metabolic network reconstructions are biochemical, genetic, and genomic (BiGG) structured knowledge bases that can be represented mathematically. Mathematical modeling of reconstructions can help scientists to find the right and effective targets of metabolic engineering (Erdrich et al., 2014; Fleming and Thiele, 2011; Fleming et al., 2009; Schellenberger et al., 2010; Schellenberger et al., 2011).

Genome-scale metabolic network reconstruction is a very laborious and time-consuming work, which in some cases requires teamworking of experts from several months to a few years (Chindelevitch et al., 2012; Dias et al., 2015). Genome-scale metabolic models are available for more than 100 organisms (Zhang et al., 2017). Likewise, for almost all of these organisms, there are multiple versions of metabolic network reconstructions. For example, one strain of cyanobacteria, the Synechocystis sp. PCC6803 has eleven different reconstructed metabolic networks developed by various research groups (Erdrich et al., 2014; Joshi et al., 2017; Knoop et al., 2013; Nogales et al., 2012; Saha et al., 2012; Yu et al., 2013). Unfortunately, the published metabolic networks have been presented inconsistently with substantially different levels of annotations (Swainston et al., 2011). Also, all of the metabolic network reconstructions have varying degrees of completeness (Pitkänen et al., 2011). In a recent study, Ravikrishnan and Raman critically examined about 100 different metabolic networks (Ravikrishnan and Raman, 2015). Their study shows a significant inconsistency in published metabolic networks that makes it very difficult to compare and use them. The
substantial lack of consistency and compatibility between metabolic networks highlights the urgent need for the development of an appropriate tool that can compare and merge them. Such a tool should produce enhanced metabolic network reconstructions without the need for more experimental tests and without detracting from the potential values of primary networks.

In this study, we developed a standalone, easy-to-use, graphical user interface (GUI) software tool iMet that can be used to integrate metabolic networks effectively. The iMet has been extensively tested on most common operating systems including Linux, Windows, and Macintosh. It is not web based and no other auxiliary software needed to run it. Using this tool, metabolic networks can be loaded, edited and integrated. Finally, the integrated network is given out in the standard SBML (Systems Biology Markup Language (Hucka et al., 2003)) file format and will be ready for more analysis. It should be mentioned that we developed the initial version of this software in 2016 (Mohammadi et al., 2016), but in the current version, we added many other features and capabilities.

2. Materials and method

2.1. Algorithm

Fig. 1 shows a schematic workflow of the algorithm. We can summarize the iMet software algorithm in the seven main steps below:

Step 1. Validation of inputs and information extraction

In the first step, the validity of the input metabolic networks is evaluated for the SBML file format. Then, the details of the information about metabolic networks are extracted and displayed in an easily editable format. The main information extracted about genes, metabolites, and reactions includes the following: Gene ID, E.C. (Enzyme Commission) Number (Bairoch, 2000), KEGG ID (Kanehisa and Goto, 2000), UniProt ID (The UniProt Consortium, 2011), CAS (Chemical Abstracts Service) Number, Formula, and ChEBI (Chemical Entities of Biological Interest) ID (Degtyarenko et al., 2008).

Step 2. Curating and enhancing of input models

After parsing the input metabolic networks, the users can edit inconsistent information about each metabolite and/or reaction. In addition, iMet uses KEGG database that has rich information about genes, metabolites, and reactions to enhance the quality of input networks by adding more information. This step can considerably enhance the process of networks integration. However, users can ignore this step and go straight to the next step if they want to.

Step 3. Detection of similar metabolites

In this step, a similarity score is assigned to each pair of metabolites across two input metabolic networks. The main contributors to the similarity score of metabolites are similarities based on Name, KEGG ID, Protein ID, ChEBI ID, CAS Number, and Compartment.

Step 4. Editing the similarities of metabolites

After detecting similar metabolites, there is a capability to revise the assigned similarity scores by the user in a simple manner.

Step 5. Detection of similar reactions

In this step, similarity scores for each pair of reactions across two networks are computed. This score is computed according to the similarities of different features of reactions, including reactants, products, and the total number of them, stoichiometry coefficients of similar metabolites, and reversibility.

Step 6. Editing the similarities of reactions

At this point, users can either accept or reject the reactions that are considered as similar or identical, in a similar way as step 4.

Step 7. Integrating the input models

Finally, iMet integrates input metabolic networks using the detected similar or identical reactions as bridges. The obtained integrated metabolic network can be saved as a standard SBML file format for more analysis.

2.2. Implementation

iMet is a standalone software tool to reconcile metabolic networks where metabolic networks can be loaded, edited and integrated. This tool has been implemented using the Java programming language as a cross-platform tool with a user-friendly graphical interface. Fig. 2 shows the main parts of the graphical user interface of the iMet. This tool can be used on each operating system with a Java runtime environment (JRE). iMet has been extensively tested on the most common operating systems including Linux, Windows, and Macintosh.

iMet accompanied with sample files is available in the supplementary file 1 at the following web address: “https://github.com/rmohamadi/iMet-Samples”. In addition, the source code of the iMet is available in the supplementary file 2 in the zip file format at this web address: “https://github.com/rmohamadi/iMet_Source-Code”.

2.3. Main features

iMet is an efficient standalone software tool with a user-friendly GUI to integrate genome-scale metabolic network reconstructions. This tool has some distinguishing features that have summarized in Table 1.

3. Results

3.1. Case study

Use of fossil fuels in the past decades has led to the accumulation of contaminants in natural ecosystems (Atsumi et al., 2009; Machado and Atsumi, 2012). Cyanobacteria are photosynthetic microorganisms that can take atmospheric Carbon dioxide (CO2) and produce organic compounds like third generation biofuels (Ducat et al., 2011; Zhou and Li, 2010). In addition, they can produce oxygen during photosynthesis (Quintana et al., 2011). Hence, this group of microorganisms can be a good model to produce biofuels and reduce the pollution of ecosystems (Lin and Liao, 2012; Lindberg et al., 2010; Quintana et al., 2011; Rosgaard et al., 2012).

As a case study, we tried to integrate several metabolic network reconstructions of a cyanobacterium, Synechocystis sp. PCC 6803, using iMet. In this study, we took three different metabolic network reconstructions of this microorganism in SBML file format from three different study sources: a. Knoop, H., et al. (2010) (Knoop et al., 2010), b. Knoop, H., et al. (2013) (Knoop et al., 2013), and c. Erdrich, P., et al. (2014) (Erdrich et al., 2014). Then, by the help of iMet, we merged these reconstructions to create a new reconstruction named iMCyn1. The SBML files of these reconstructions (three primary models and the iMCyn1) can be found in the supplementary file 3 in the zipped format at “https://github.com/rmohamadi/Models-Analyses-Instruction”. The .mat files of these reconstructions are also included in the supplementary file 3.

After merging the networks and creating iMCyn1, we have tried to model all reconstructions (included iMCyn1) to produce five biofuels by adding some reactions to the networks. These reactions added to
propanol, isobutanol, 3-methyl-1-butanol, and 2-methyl-1-butanol metabolic pathways to provide the possibility of producing them. The production pathway of ethanol was complete and there was no need to add any other reaction. We also added excretion reactions to all biofuels and imposed several constraints on models (Knoop et al., 2013). We then performed the Flux Balance Analysis (FBA) (Orth, Thiele and Palsson, 2010; Schellenberger et al., 2011) to compare the new merged model, iMCyn1, with the previously reported models. The instructions of these three steps (integrating metabolic network reconstructions by iMet, making new models by adding reactions and constraints and running FBA in the MATLAB) are available at this web address: “https://github.com/rmohamadi/Models-Analyses-Instruction”.

Fig. 1. Schematic workflow of the iMet algorithm.
Fig. 2. The main parts of the graphical user interface of the iMet. A) Start page of the iMet; B) Validation of inputs and information extraction; C) curating the input models; D) Collecting information from KEGG; E) Similar metabolites detection; F) Editing the similarities of metabolites; G) Detection of similar reactions; H) Integrating the input models.
Fig. 2. (continued).
Fig. 2. (continued).
Fig. 2. (continued).
Biofuels could be seen by merging model 1 into model 2 (Knoop et al., 2010) into the others was useful too. An increment in the significant improvement in FBA results of model was failed, merging it into model 2 and model 1 caused a significant about genes or gene IDs. Despite the fact that analyzing the third Model (Erdrich et al., 2014) as it does not have any information about reactions and metabolites as well as an increment in the yields of desired products in iMCyn1, in comparison with primary models. For example, the flux of some reactions and metabolites increased from 3.27E-01 mmol/gDW/h to 5.66E-01 mmol/gDW/h in iMCyn1 (Please refer to the results of analyses in the supplementary file 3).

4. Discussion and conclusion

The results of analyses showed an enhancement in the coverage of reactions and metabolites with their related flux values included in the supplementary file 3.

Table 1: Distinguished features of the iMet software tool.

| Features                        | A brief description                                                                 |
|---------------------------------|-------------------------------------------------------------------------------------|
| Cross-platform                  | iMet can be run on the most common operating systems including Linux, Windows, and Macintosh. |
| Graphical user interface        | iMet provides a user-friendly GUI (graphical user interface) that can be used easily by biologists with no significant computer science knowledge. |
| SBML high compatibility         | Currently, SBML files are defined in three levels and each level has multiple versions. iMet supports all SBML file formats, which enables users to integrate metabolic networks that published by various research groups around the world at different times. |
| Employing the supervision of expert users | Users can edit the input models (that displayed in an appropriate format) in a mouse-driven fashion. In addition, users can supervise the iMet by accepting or rejecting the detected similar metabolites or reactions and revising the similarity scores. |
| Collecting information from online resources | iMet can collect information about reactions and metabolites, such as KEGG IDs; CHEBI IDs; CAS Numbers and Metabolites’ alternative names, from the KEGG database. The obtained information can greatly influence the process of detecting similar metabolites and reactions and so can enhance the correctness of the integration steps. |

To date, few approaches have not been completed yet, have been made on merging metabolic networks of organisms. This study is an endeavor to make a further step forward in this area, by using a progressive approach. The results of the case study show that the additional information which is added by iMet in the process of merging could improve the functionality of the output network without any reduction in the previous functionalities.

Furthermore, the iMet can detect different types of inconsistencies and correct them. These inconsistencies can be in identifiers, names, standard identifiers, stoichiometric coefficients, and reversibility of reactions. The design of the iMet algorithm, allows us to correct most of these inconsistencies. Another part of the inconsistencies, including the mass balance and the atoms, can be corrected by the user at each step of the merging. In general, by information enrichment of various databases and the greater the degree of monitoring in each phase of integration, the probability of such errors will be reduced.

The production of futile cycles or type II pathways during reconstruction and modeling of metabolic networks may occur most often by automated methods. These futile cycles can result in infeasible and implausible production fluxes of ATP, reducing equivalents, or other substrates (Fritzemeier et al., 2017). However, since at each step of the iMet algorithm there is a manual monitoring and supervising phase, iMet can then be useful in removing and correcting such cycles or pathways.

The results of analyses in our case study showed the performance of this software. These results show the superior performance of the iMet which can be profitable in other studies related to different types of organisms. In spite of the fact that monitoring and manual revising of similarity scores could lead to a better result by iMet, it is also capable to operate in a fully automated mode, which is really useful while facing to a huge network.

iMet is a standalone easy to use software application with a user-friendly GUI, that enables the integration of the metabolic networks in a mouse-driven fashion. In other words, there is no need for web access or any other auxiliary software to use it. However, web access might cause a better result. iMet satisfies the urgent need for an in-silico tool to reconcile existing incomplete metabolic networks into an enhanced metabolic network without any experimental efforts. iMet is freely available for academic and non-commercial use at “https://github.com/rmohamadi/iMet-Samples”.

Declarations

Author contribution statement

Reza Mohammadi: Conceived and designed the experiments;Performed the experiments; Analyzed and interpreted the data; Wrote the paper.
Javad Zahiri: Conceived and designed the experiments; Performed

Table 2: Comparison of previously reported models with our new model (iMCyn1).

| Models                          | N. of Metabolites | N. of Reactions | N. of Blocked Reactions | Ethanol Fluxes $^2$ | Propanol Fluxes | 3-methyl-1-butanol Fluxes | 2-methyl-1-butanol Fluxes | Isobutanol Fluxes |
|--------------------------------|-------------------|-----------------|-------------------------|---------------------|------------------|--------------------------|--------------------------|-------------------|
| 1. Knoop, H. et al. (Knoop et al., 2010) | 274              | 380             | 55                      | 0.225               | 0.0              | 2.15E-26                 | 0                        | 1.89E-29          |
| 2. Knoop, H. et al. (Knoop et al., 2013) | 556              | 759             | 55                      | 0.6465              | 0.431            | 0.2439                   | 0.2571                   | 0.3157            |
| 3. Erdrich, P. et al. (Erdrich et al., 2014) | 518              | 600             | ND$^3$                  | ND                  | ND               | ND                       | ND                       | ND                |
| 4. iMCyn1                          | 557              | 768             | 52                      | 0.6601              | 0.4400           | 0.2526                   | 0.2640                   | 0.5658            |

$^1$ In 2-methyl-1-butanol production pathway.
$^2$ mmol/gDW/h.
$^3$ Not Defined.
the experiments.
Mohammad Javad Nirooomand: Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Funding statement
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement
The authors declare no conflict of interest.

Additional information
Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2019.e01766.

Acknowledgements
The authors thank the Department of Bioscience and Biotechnology of Malek Ashtar University for providing this research facilities.

References
Atsumi, S., Higashide, W., Liao, J.C., 2009. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Nat. Biotechnol. 27 (12), 1177–1180.
Bairoch, A., 2000. The ENZYME database in 2000. Nucleic Acids Res. 28 (1), 304–305.
Chindelevitch, L., Stanley, S., Hung, D., Regev, A., Berger, B., 2012. ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 36, D344–D350. Database issue.
Ducat, D.C., Way, J.C., Silver, P.A., 2011. Engineering cyanobacteria to generate high-value products. Trends Biotechnol. 29 (2), 95–103.
Erdrich, P., Knoop, H., Steuer, R., Klamt, S., 2014. Cyanobacterial biofuel biosynthesis: new insights and strain design strategies revealed by computational modeling. Microb. Cell Factories 13 (1), 128.
Fleming, R.M., Thiele, I., 2011. von Bertalanffy 1.0: a COBRA toolbox extension to thermodynamically constrain metabolic models. Bioinformatics 27 (1), 142–143.
Fleming, R.M., Thiele, I., Nasheuer, H.P., 2009. Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to Escherichia coli. Biophys. Chem. 145 (2–3), 47–56.
Fritzemeier, C.J., Hartlieb, D., Szapanyos, B., Papp, B., Lercher, M.J., 2017. Erroneous energy-generating cycles in published genome-scale metabolic networks: identification and removal. PLoS Comput. Biol. 13 (4), e1005494.
Henry, C.S., Jankowski, M.D., Broadbelt, L.J., Hatzimanikatis, V., 2006. Genome-scale thermodynamic analysis of Escherichia coli metabolism. Biophys. J. 90 (4), 1453–1461.
Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., et al., 2003. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19 (4), 524–531.
Joshi, C.J., Peebles, C.A.M., Prasad, A., 2017. Modeling and analysis of flux distribution and bioprocess formation in Synechocystis sp. PCC 6803 using a new genome-scale metabolic reconstruction. Algal Research 27, 295–310.
Kanehsa, M., Goto, S., 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28 (1), 27–30.
Knoop, H., Grundel, M., Zilliges, Y., Lehmann, R., Hoffmann, S., Lockau, W., Steuer, R., 2013. Flux balance analysis of cyanobacterial metabolism: the metabolic network of Synechocystis sp. PCC 6803. PLoS Comput. Biol. 9 (6), e1003081.
Knoop, H., Zilliges, Y., Lockau, W., Steuer, R., 2010. The metabolic network of Synechocystis sp. PCC 6803: systematic properties of autotrophic growth. Plant Physiol. 154 (1), 410–422.
Lan, E.I., Liao, J.C., 2012. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. Proc. Natl. Acad. Sci. U. S. A. 109 (16), 6018–6022.
Lewis, N.E., Nagarajan, H., Palsson, B.O., 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat. Rev. Microbiol. 10 (4), 291–305.
Lindberg, P., Park, S., Melis, A., 2010. Engineering a platform for photosynthetic isopropanol production in cyanobacteria, using Synechocystis as the model organism. Metab. Eng. 12 (1), 70–79.
Machado, I.M., Atsumi, S., 2012. Cyanobacterial biofuel production. J. Biotechnol. 162 (1), 50–56.
Mardinoglu, A., Nielsen, J., 2015. New paradigms for metabolic modeling of human cells. Curr. Opin. Biotechnol. 34, 91–97.
Mohammad, R., Fallah Mehrabadi, J., Bidekori, G., Zahir, J., Javad Nirooomand, M., Masoudi-Nejad, A., 2016. A systems biology approach to reconcile metabolic network models with application to Synechocystis sp. PCC 6803 for biofuel production. Mol. Biosyst. 12 (8), 2552–2561.
Nogales, J., Gudmundsson, S., Knight, E.M., Palsson, B.O., Thiele, I., 2012. Detailing the optimality of photosynthesis in cyanobacteria through systems biology analysis. Proc. Natl. Acad. Sci. U. S. A. 109 (7), 2678–2683.
Orth, J.D., Thiele, I., Palsson, B.O., 2010. What is flux balance analysis? Nat. Biotechnol. 28 (3), 245–248.
Pitkänen, E., Joutlen, P., Hou, J., Syed, M.F., Blomberg, P., Kludan, J., et al., 2014. Comparative genome-scale reconstruction of gapless metabolic networks for present and ancestral species. PLoS Comput. Biol. 10 (2), e1003465.
Quintana, N., Van der Kooy, F., Van de Rhee, M.D., Voshol, G.P., Verpoorte, R., 2011. Renewable energy from Cyanobacteria: energy production optimization by metabolic pathway engineering. Appl. Microbiol. Biotechnol. 91 (3), 471–490.
Ravikrishnan, A., Raman, K., 2015. Critical assessment of genome-scale metabolic networks: the need for a unified standard. Briefings Bioinf. 16 (6).
Rozgaard, L., de Porcellinis, A.J., Jacobsen, J.H., Figgaard, N.U., Sakuragi, Y., 2012. Bioengineering of carbon fixation, biofuels, and biocatalysts in cyanobacteria and plants. J. Biotechnol. 162 (1), 134–147.
Saha, R., Versepaut, A.T., Berla, B.M., Mueller, T.J., Pakrasi, H.B., Maranas, C.D., 2012. Reconstruction and comparison of the metabolic potential of cyanobacteria: Cyanotherce sp. ATCC 51142 and Synechocystis sp. PCC 6803. PLoS One 7 (10), e48285.
Schellenberger, J., Park, J.O., Conrad, T.M., Palsson, B.O., 2010. BiGG: A Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. BMC Bioinf. 11, 213.
Schellenberger, J., Que, R., Fleming, R.M., Thiele, I., Orth, J.D., Feist, A.M., et al., 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Bioinf. 16 (6).
Swainston, N., Smallbone, K., Mendes, P., Kell, D., Paton, N., 2011. The SubMinaL Toolbox: automating steps in the reconstruction of metabolic networks. J. Integr. Biol. 8 (2), 186.
The UniProt Consortium, 2011. Ongoing and future developments at the universal Protein resource. Nucleic Acids Res. 39, D214–D219. Database issue.
Yu, Y., You, L., Liu, D., Hollinshead, W., Tang, Y.Y., Zhang, F., 2013. Development of Synechocystis sp. PCC 6803 as a phototrophic cell factory. Mar. Drugs 11 (8), 2894–2916.
Zhang, H., Ye, C., Xu, N., Chen, C., Chen, X., Yuan, F., et al., 2017. Reconstruction of a genome-scale metabolic network of komagataeibacter natacica RZ501 for cellulose production. Sci. Rep. 7 (1), 7911.
Zhou, J., Li, Y., 2010. Engineering cyanobacteria for fuels and chemicals production. Protein Cell 1 (3), 207–210.