Bioinformatics, 35(3), 2019, 535–537
doi: 10.1093/bioinformatics/bty656
Advance Access Publication Date: 20 July 2018
Applications Note

Systems biology

gMCS: fast computation of genetic minimal cut sets in large networks

Iñigo Apaolaza [1], Luis Vitores Valcarcel[1,2] and Francisco J. Planes[1,*]

[1]Department of Biomedical Engineering and Sciences, Tecnun School of Engineering, University of Navarra, 20018 San Sebastián, Spain and [2]Centro de Investigación Médica Aplicada (CIMA), University of Navarra, Programa de Oncohematología, 31008 Pamplona, Spain

*To whom correspondence should be addressed.

Associate Editor: Bonnie Berger

Received on April 23, 2018; revised on July 11, 2018; editorial decision on July 18, 2018; accepted on July 19, 2018

Abstract

ABSTRACT: Motivation: The identification of minimal gene knockout strategies to engineer metabolic systems constitutes one of the most relevant applications of the COnstraint-Based Reconstruction and Analysis (COBRA) framework. In the last years, the minimal cut sets (MCSs) approach has emerged as a promising tool to carry out this task. However, MCSs define reaction knockout strategies, which are not necessarily transformed into feasible strategies at the gene level.

Results: We present a more general, easy-to-use and efficient computational implementation of a previously published algorithm to calculate MCSs to the gene level (*gMCSs*). Our tool was compared with existing methods in order to calculate essential genes and synthetic lethals in metabolic networks of different complexity, showing a significant reduction in model size and computation time.

Availability and implementation: *gMCS* is publicly and freely available under GNU license in the COBRA toolbox (https://github.com/opencobra/cobratoolbox/tree/master/src/analysis/gMCS).

Contact: fplanes@tecnun.es

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Algorithms and tools for minimal cut set (MCS) computation were originally developed for reaction knockouts (Kamp and Klamt, 2014; Tobalina *et al.*, 2016), i.e. MCSs define minimal sets of reactions whose removal would render the functioning of a given metabolic task impossible. However, we recently observed in a cancer study that MCSs are not directly converted into feasible and minimal knockout strategies at the gene level (Apaolaza *et al.*, 2017) due to the fact that the Gene-Protein-Reaction (GPR) rules, typically provided in genome-scale metabolic reconstructions, are not one-to-one. To overcome this issue, we adapted the underlying mathematical methods to obtain minimal gene knockout interventions, termed genetic Minimal Cut Sets (gMCSs) (Apaolaza *et al.*, 2017). This extension was focused on the application under study (cancer research), lacking a more general, easy-to-use and efficient approach to calculate gMCSs. To address this gap, we here present a novel computational tool, called *gMCS*, which demonstrates a superior computational performance over our previous approach and existing tools in the literature (Kamp and Klamt, 2017; Machado *et al.*, 2016).

2 Materials and methods

gMCS is integrated in the COBRA toolbox (Schellenberger *et al.*, 2011), a popular MATLAB software suite for the reconstruction and analysis of metabolic networks, enabling the calculation of gMCSs in a natural environment for the Systems Biology community. The main function, *calculateGeneMCS*, requires as compulsory input data: (i) the metabolic model structure (under COBRA Toolbox format); (ii) the identifier of the gene knockout constraints matrix, *G*; (iii) number of solutions to be computed and (iv) maximum length of solutions. The *G* matrix is an important feature in our approach whose calculation is not trivial. *gMCS* provides an efficient function for this task (*buildGMatrix*), which substantially

reduced the computation time presented in Apaolaza *et al.* (2017) (see Section 3). We also amended the mathematical formulation to deal with an issue of dependencies between rows of the *G* matrix, removing a possible source of false positives and inefficiencies. A full description of the final mathematical optimization model and algorithms for the *G* matrix construction can be found in Supplementary Appendix S1.

Note that *gMCS* runs on different solvers available in the COBRA Toolbox. However, fast performance is only available for IBM Ilog Cplex, where we implemented the "warm-start" strategy through the *Populate* function (Kamp and Klamt, 2014). A detailed tutorial of *gMCS* can be found in in the COBRA Toolbox.

3 Results

We calculated single, double and triple gene deletions which result to be lethal for growth in six different metabolic networks, i.e. *E. coli* core (Orth *et al.*, 2010), iAF1260 (Feist *et al.*, 2007), Recon1 (Duarte *et al.*, 2007), Recon2.v04 (Thiele *et al.*, 2013), Recon2.2 (Swainston *et al.*, 2016) and Recon3D_3.01 (Brunk *et al.*, 2018). We compared the results of *gMCS* with the method described in Kamp and Klamt (2017), here denoted as *vKK*. The *vKK* approach constitutes the latest implementation of the strategy presented in Machado *et al.* (2016), which extends the metabolic network to account for GPR rules (*extended metabolic networks*). We carried out our own implementation of the *vKK* approach and checked that CNA (Kamp *et al.* 2017) and our *vKK* implementation had comparable computation times to calculate MCSs. For the different networks considered, the same set of solutions using *vKK* and *gMCS* were obtained (see Supplementary Material SI). The computations were carried out on a 64 bit Intel Xeon E5-1620 v2 at 2.70 GHz (4 cores) and 16 GB of RAM.

In Figure 1, the computation time was divided into two different steps: the time required for the preparation of the optimization model (*preparation time*), which involves the building of the *G* matrix in *gMCS* and the construction of extended networks in *vKK*, and the time required to enumerate solutions (*solving time*). Overall, it can be observed that the preparation time is smaller in *vKK*, since *G* matrix requires additional steps (see Supplementary Appendix S1). However, with respect to Apaolaza *et al.* (2017), the improvement in *G* matrix construction is clear. In that work, it took us 20 min to calculate *G* in Recon2.04, only including lowly expressed genes in the cancer samples under study (30%–50% of the total number of genes) as available knockouts. Here, we reached *G* in Recon2.04 in 136 seconds with all genes available for knockout.

The construction of *G* matrix is compensated in the computation time to calculate solutions. The solving time of *gMCS* is substantially lower for essential genes and synthetic lethals of order 2, becoming the differences more extreme for higher knockout strategies, as seen in Figure 1 for synthetic lethals of order 3. These significant differences are due to the increase in model dimension in the case of *vKK* (Supplementary Table S1). The most remarkable case is iAF1260, which has the highest number of synthetic lethals up to order 3, leading to an increase of computation time in both approaches, but dramatically in the case of *vKK*. Finally, we checked that *gMCS* is capable of enumerating a large number of solutions in complex networks, such as Recon2.04 or Recon3D_3.01, in a moderate computation time (see Supplementary Appendix S2 and Supplementary Material SII).

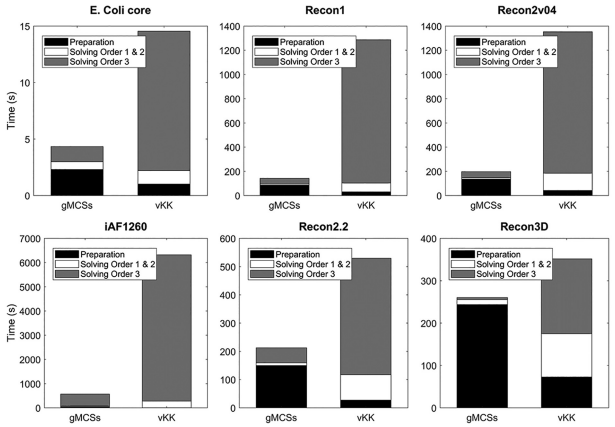

Fig. 1. Side-by-side comparison to calculate minimal gene knockout interventions. For each metabolic model under study, **(a)** E. coli core, **(b)** iAF1260, **(c)** Recon1, **(d)** Recon2.2, **(e)** Recon2.v04 and **(f)** Recon3_3.01, preparation (black coloring) and solving time (in seconds) for synthetic lethals of order 1 and 2 (white) and synthetic lethals of order 3 (grey) for our gMCSs approach and the *vKK* approach are computed

4 Discussion

gMCS will facilitate the search of effective gene knockout interventions in different application areas in Systems Biology. As shown in Apaolaza *et al.* (2017), we can integrate gMCSs with –omics data in order to identify selective drug targets or response biomarkers in cancer, a major question in personalized medicine. The same approach can be applied to the other diseases, such as bacterial infections, where similar methods have been proven promising. *gMCS* will be of interest for researchers in the field of metabolic engineering, where strain optimization for the production of biocompounds of interest is an active research topic.

Acknowledgements

The authors thank Ronan Fleming and the COBRA Toolbox team (arXiv: 1710.04038) for their help in the development of gMCS.

Funding

I.A. was supported by a Basque Government predoctoral grant [PRE_2016_2_0044] and L.V.V. by Instituto de Salud Carlos III (ISCIII) [FI17/00297]. This work was supported by the Minister of Economy and Competitiveness of Spain [BIO2013-48933, BIO2016-77998-R].

Conflict of Interest: none declared.

References

Apaolaza,I. *et al.* (2017) An in-silico approach to predict and exploit synthetic lethality in cancer metabolism. *Nat. Commun.*, **8**,

Brunk,E. *et al.* (2018) Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.*, **36**, 272–281.

Duarte,N.C. *et al.* (2007) Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci.*, **104**, 1777–1782.

Feist,A.M. *et al.* (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.

Kamp,A. *et al.* (2017) Use of CellNetAnalyzer in biotechnology and metabolic engineering. *J. Biotechnol.*, **261**, 221–228.

Kamp,A.V. and Klamt,S. (2014) Enumeration of smallest intervention strategies in genome-scale metabolic networks. *PLoS Comput. Biol.*, **10**, e1003378.

Kamp,A.V.and Klamt,S. (2017) Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms. *Nat. Commun.*, **8**, 15956.

Machado,D. *et al.* (2016) Stoichiometric representation of gene–protein–reaction associations leverages constraint-based analysis from reaction to gene-level phenotype prediction. *PLoS Comput. Biol.*, **12**, e1005140.

Orth,J.D. *et al.* (2010) Reconstruction and use of microbial metabolic networks: the core Escherichia coli metabolic model as an educational guide. *EcoSal plus*, **4**.

Schellenberger,J. *et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0. *Nature Protocols*, **6**, 1290.

Swainston,N. *et al.* (2016) Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*, **12**,

Thiele,I. *et al.* (2013) A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.*, **31**, 419–425.

Tobalina,L. *et al.* (2016) Direct calculation of minimal cut sets involving a specific reaction knock-out. *Bioinformatics*, **32**, 2001–2007.