Bioinformatics, 32(8), 2016, 1265–1266
doi: 10.1093/bioinformatics/btv742
Advance Access Publication Date: 18 December 2015
Applications Note

OXFORD

Systems biology

FluxModeCalculator: an efficient tool for large-scale flux mode computation

Jan Bert van Klinken[1,2]* and Ko Willems van Dijk[1,2,3]

[1]Department of Human Genetics, [2]Einthoven Laboratory for Experimental Vascular Medicine and [3]Department of Medicine, Division of Endocrinology, LUMC, Leiden, The Netherlands

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on 21 July 2015; revised on 11 December 2015; accepted on 14 December 2015

Abstract

Summary: Elementary flux mode (EFM) analysis is a powerful technique for determining the metabolic capacities and robustness of stoichiometric networks. Recently, several improvements have been made to the algorithm for enumerating the EFMs, making the study of large models possible. However, currently these tools require high performance workstations to perform large-scale EFM computations, thus limiting their applicability. We developed a more time and memory efficient implementation of the algorithm for EFM enumeration in MATLAB, called FluxModeCalculator, which enables large-scale EFM computation on ordinary desktop computers.

Availability and implementation: FluxModeCalculator is open source and freely available under the terms of the GNU General Public License v3.0 at http://www.lumc.nl/jan-bert-van-klinken

Contact: j.b.van_klinken@lumc.nl

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

Due to recent advances in genomic, transcriptomic and metabolomic technologies, it has become possible to reconstruct detailed metabolic maps of a wide range of organisms and tissue types. Elementary Flux Mode analysis is a powerful method for investigating the metabolic capacities of these networks, listing the complete set of physiological states that are feasible given mass balance and thermodynamic constraints. EFMs are defined as the minimal reaction sets that are able to operate at steady state (Schuster and Hilgetag, 1994) and provide a natural and unbiased definition of a metabolic pathway. EFM analysis has been applied for several different purposes, such as designing knockout strategies for bacteria (Trinh *et al.*, 2008), analyzing network robustness to perturbations (Stelling *et al.*, 2002), determining substrate cycles (Gebauer *et al.*, 2012), and investigating network flexibility in optimal growth states (Kelk *et al.*, 2012).

Formally, EFM analysis corresponds to enumerating the extreme rays of the polyhedral cone that is formed by the intersection of the right null space of the stoichiometric matrix and the positive orthant

$$\mathbf{N} \cdot \mathbf{v} = 0 \quad (1)$$

$$\mathbf{v} \geq 0 \quad (2)$$

where $\mathbf{v}$ represents the flux vector and $\mathbf{N}$ the m by n stoichiometric matrix, with N_{ij} the stoichiometric coeffcient of the ith compound in the jth reaction. Equation (1) represents the set of constraints imposed by the steady state assumption—i.e. internal metabolites are neither produced nor consumed—and (2) represents the thermodynamic constraint that the flux through irreversible reactions is either positive or zero. Importantly, in EFM analysis $\boldsymbol{N}$ represents the expanded stoichiometric matrix, where reversible reactions are split into two irreversible reactions in both directions.

Traditionally flux mode enumeration has been hampered by the computational complexity of the double description method (Motzkin *et al.*, 1953), which is the principal algorithm for calculating the extreme rays of a polyhedral cone. Consequently, this method has only been applied to relatively small metabolic networks. Due to recent advances in the implementation of the double description method, flux mode calculation has also been made possible in larger models (Jevremović and Boley, 2012; Terzer and Stelling, 2008). However, these tools are only able to perform large-scale flux mode computation on high performance workstations and clusters, which limits their usability. We developed a more efficient

implementation, called FluxModeCalculator, that enables large-scale EFM computation on ordinary desktop computers.

FluxModeCalculator includes the main improvements that have been previously made to the double description method, such as network compression, storing solved inequalities as bit patterns (Gagneur and Klamt, 2004), and using bit pattern trees for the combinatorial adjacency test and for candidate pruning (Terzer and Stelling, 2008). We further optimized the memory and time efficiency of the algorithm by vectorising the bit pattern tree and storing intermediate solutions on the hard drive. As a result, our implementation only requires $\#\text{fluxmodes} \cdot (4\lceil \frac{1}{32}\#\text{reactions}\rceil + 10)$ bytes of free memory to complete an iteration, plus ~1 GB for swapping data between the memory and hard drive. In addition, our implementation uses the OpenMP API to optimally exploit processor architectures with multiple cores. For details, see the Supplementary text.

A feature that we added to our tool is to check at the end of each iteration whether the resulting flux modes still fit in the memory; if this is no longer the case then the algorithm will terminate and return the intermediate solution. This allows the user to inspect the intermediate flux modes that satisfy all except the final constraints for gaining insight into the complexity of the model and detect the presence of hub metabolites or reactions. In addition, we combined this feature with the demand-based network subdivision strategy proposed by Hunt *et al.* (2014) to automatically subdivide the network for the remaining constraints and calculate the flux modes for the corresponding subnetworks.

We compared the performance of FluxModeCalculator with that of existing tools on a set of stoichiometric matrices of varying size (Supplementary text). We found that our implementation is several orders of magnitude faster than metatool (von Kamp and Schuster, 2006), and 11 times faster than efmtool (Terzer and Stelling, 2008) on the largest model that efmtool could manage on our system (68M EFMs). In contrast, the largest set of EFMs that could be calculated with FluxModeCalculator in a single run was 10 times larger (679M EFMs). For smaller models the difference in performance was less pronounced; for instance, for models with 10^6 EFMs the reduction in computation time of FluxModeCalculator with respect to efmtool was between 4- and 6-fold.

To further validate our tool, we applied it to a set of metabolic models for which the EFMs had been calculated previously. Importantly, for the largest three models high performance workstations had to be used, whereas we were able to replicate these results on an ordinary desktop computer (see Table 1). Specifically, FluxModeCalculator required only 4 GB of free memory to calculate the 271M EFMs in the core metabolism of *E. coli*, whereas Gerstl *et al.* (2015) reported that efmtool needed 90 GB of free memory to perform the same computation.

Table 1. Metabolic models used in the validation study

Model	Reactions/ metabolites	EFMs	Computation time	Required memory
S.cerevisiae[a]	78/62	1 515 315	29.0	1 GB
S.cerevisiae[a]	83/63	68 868 602	43:44.9	2 GB
E. coli[b]	95/72	271 494 722	17:12:45.7	4 GB
P.tricornutum[c]	318/335	1 935 026 864	94:54:58.7	14 GB

[a]Jevremović and Boley (2012).
[b]Orth et al. (2009).
[c]Hunt et al. (2014). Computations were performed on a computer with an Intel 3.30 GHz i3 processor and 16 GB RAM.

For the final model it was not possible to determine the full set of EFMs with 14 GB of free memory, and FluxModeCalculator finished yielding 656M intermediate flux modes, leaving 2 reactions unconstrained. Subsequently, the model was subdivided iteratively using the demand-based network splitting approach, starting with the 2 remaining reactions. The total computation time was less than 4 days, in comparison to less than a month on a cluster with 50 nodes (Hunt *et al.*, 2014). Importantly, since we found 297 313 more EFMs than Hunt and coworkers, we validated our results by checking all EFMs for consistency and uniqueness and confirmed that no inconsistent or duplicate EFMs were present. Consequently, the set of EFMs computed by Hunt is incomplete, which is most probably due to numerical errors using floating point arithmetic.

In conclusion, we have developed FluxModeCalculator, a new tool for performing flux mode analysis in stoichiometric models. FluxModeCalculator is more efficient than existing tools and enables EFM computation in models with up to an order of magnitude more EFMs.

Funding

This work was supported by grants from the Netherlands Consortium for Systems Biology (NCSB), the Center of Medical Systems Biology (CMSB), the European Network for Genetic and Genomic Epidemiology (ENGAGE), the Biobanking and Biomolecular Resources Research Infrastructure (BBMRI) 2.0 and CardioVascular Research Netherlands (CVON-ENERGISE).

Conflict of Interest: none declared.

References

Gagneur,J. and Klamt,S. (2004) Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics*, **5**, 175.

Gebauer,J. *et al.* (2012) Detecting and investigating substrate cycles in a genome-scale human metabolic network. *FEBS J.*, **279**, 3192–3202.

Gerstl,M. *et al.* (2015) Metabolomics integrated elementary flux mode analysis in large metabolic networks. *Sci. Rep.*, **5**, 8930

Hunt,K. *et al.* (2014) Complete enumeration of elementary flux modes through scalable demand-based subnetwork definition. *Bioinformatics*, **30**, 1569–1578.

Jevremović,D. and Boley,D. (2012). Parallel computation of elementary flux modes in metabolic networks using global arrays. In: The 6th Conference on Partitioned Global Address Space Programming Models, Santa Barbara, CA.

Kelk,S. *et al.* (2012) Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Sci. Rep.*, **2**, 5801–5807.

Motzkin,T.S. *et al.* (1953). The double description method. In: Kuhn,H.W. and Tucker,A.W. (eds.) *Contributions to the Theory of Games II, volume 8 of Ann. of Math. Stud.*, Princeton University Press, Princeton, pp. 51–73.

Orth,J. *et al.* (2009). Reconstruction and use of microbial metabolic networks: The core escherichia coli metabolic model as an educational guide. In: Karp,P. (ed.) *EcoSal–Escherichia Coli and Salmonella: Cellular and Molecular Biology*, ASM Press, Washington, DC, pp. 56–99.

Schuster,S. and Hilgetag,C. (1994) On elementary flux modes in biochemical reaction systems at steady state. *J. Biol. Syst.*, **2**, 165–182.

Stelling,J. *et al.* (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature*, **420**, 190–193.

Terzer,M. and Stelling,J. (2008) Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics*, **24**, 2229–2235.

Trinh,C. *et al.* (2008) Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl. Environ. Microbiol.*, **74**, 3634–3643.

von Kamp,A. and Schuster,S. (2006) Metatool 5.0: fast and flexible elementary modes analysis. *Bioinformatics*, **22**, 1930–1931.