OPINION PAPER Evolutionary Constraint-Based Formulation Requires New Bi-level Solving Techniques

Marko Budinich(✉), Jérémie Bourdon, Abdelhalim Larhlimi,
and Damien Eveillard

LINA, UMR 6241 CNRS, EMN, Université de Nantes,
2 rue de la Houssinière, Nantes, France
`marko.budinich@univ-nantes.fr`

Abstract. Constraint Based Methods had been successfully used to simulate genome-scale metabolic behaviors over a range of experimental conditions. In most applications, environmental constraints are parameterized, and the use of metabolic reactions and corresponding genes is the direct consequence of the tuning of these parameters.

However, in evolutionary studies, the problem is different: one knows the relative importance of reactions and one seeks environmental conditions that could explain such a biological fitness.

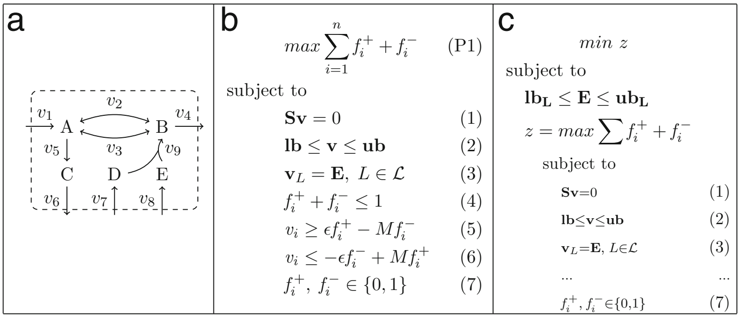
This study details this modeling paradigm change and discuss a putative formalization of such a biological problem in the form of a Mixed Integer Bi-level Linear Problem (MIBLP). Unfortunately, solving a MIBLP is difficult, paving the way for the need of further constraint based method developments for understanding evolutionary processes.

Constraint Based Methods (CBMs) are considered as efficient approaches to predict phenotypic responses and explore the structure of genome-scale networks of a variety of organisms [1,2]. For instance, they tackle effects of genetic mutations (resp. gene deletions [3,4] and gene insertion [5]) on metabolic behaviors, whereas complementary analysis focused on gene transfers [6], gene dispensability [7] or nutrient adaptation [8]. Similarly, high-throughput sequencing allows today to compare lineages and biological studies to infer evolutionary patterns [9], paving the way to bridge evolutionary studies and CBMs.

From an evolutionary viewpoint, environment exerts or relaxes pressure in biological systems. Thus, in front of detrimental or beneficial environments, organisms adapt themselves by gaining or loosing functions [10,11]. Those knowledge being available nowadays, it is of great interest to decipher the environmental conditions that maximize lineage evolution, pointing conditions that could lead to metabolic reaction losses [12].

When CBM is applied in evolutionary contexts, environment usually is first parameterized and its effect is then studied and interpreted *via* a range of simulations [6,13]. Herein, instead of standard approaches, we propose to focus on selecting environmental conditions that make most reactions unable to carry

O. Roux and J. Bourdon (Eds.): CMSB 2015, LNBI 9308, pp. 279–281, 2015.
DOI: 10.1007/978-3-319-23401-4_23



$$\begin{aligned} &max \sum_{i=1}^{n} f_i^+ + f_i^- && \text{(P1)} \\ &\text{subject to} && \\ &\mathbf{Sv} = 0 && (1) \\ &\mathbf{lb} \leq \mathbf{v} \leq \mathbf{ub} && (2) \\ &\mathbf{v}_L = \mathbf{E},\ L \in \mathcal{L} && (3) \\ &f_i^+ + f_i^- \leq 1 && (4) \\ &v_i \geq \epsilon f_i^+ - M f_i^- && (5) \\ &v_i \leq -\epsilon f_i^- + M f_i^+ && (6) \\ &f_i^+,\ f_i^- \in \{0,1\} && (7) \end{aligned}$$

$$\begin{aligned} &min\ z && \\ &\text{subject to} && \\ &\mathbf{lb_L} \leq \mathbf{E} \leq \mathbf{ub_L} && \\ &z = max \sum f_i^+ + f_i^- && \\ &\text{subject to} && \\ &\mathbf{Sv}{=}0 && (1) \\ &\mathbf{lb}{\leq}\mathbf{v}{\leq}\mathbf{ub} && (2) \\ &\mathbf{v}_L{=}\mathbf{E},\ L{\in}\mathcal{L} && (3) \\ &\ldots && \ldots \\ &f_i^+, f_i^- {\in}\{0,1\} && (7) \end{aligned}$$

Fig. 1. Evolutionary problem formulation. Considering a putative metabolic network (a), we assume the production of metabolite B as a fitness proxy. If A is the only substrate in a particular environment, we expect that genes coding for v_7, v_8 and v_9 disappear upon evolution. b) The inner Problem (P1) identify blocked reactions, i.e., those that can not carry a non-zero flux under steady-state conditions. A variation of (P1) is used in [14,15]. c) A mixed integer bi-level linear problem seeking for an environmental setting (i.e., defined values for environmental variables in $\mathcal{L}$, see text) $\mathbf{E}$ that maximizes the number of blocked reactions.

fluxes (see Fig. 1a). Indeed, recent evolutionary studies hypothesize that such blocked reactions are likely to be lost as functions due to evolution [12].

Formalization of the previous statements leads to an optimization problem as shown in Fig. 1b. Constraints in (1) and (2) are mass balance and boundary conditions. Equations in (3) represent environmental variables as a subset of reaction fluxes indexed by $\mathcal{L}$.

To identify blocked reactions, we introduce for each reaction i two binary variables f_i^+ and f_i^- (resp. forward and reverse flux) in (7). Constraints in (4), (5) and (6) guarantee that a reaction i is blocked if and only if $f_i^+ = f_i^- = 0$. By M (resp. ϵ), we denote a large (resp. small) number. Given an environmental setting $\mathbf{E}$, maximizing $\sum f_i^+ + f_i^-$ identifying all blocked reactions.

As a next step in our study, we propose to use the Mixed Integer Bi-level Linear Problem (MIBLP) shown in Fig. 1c in order to select an environmental setting $\mathbf{E}$ that maximizes the number of blocked reactions. The main difference with other bi-level approaches is the focus on controlling metabolic networks using only environmental variables and not genetic manipulations [16].

Unfortunately, despite several tentatives [17,18], no general solution is available for this type of problem [19], emphasizing the need for an *ad-hoc* algorithm implementation to solve this new evolutionary problem. Furthermore, for the sake of generalization, any method that handle this type of bi-level program, will lead to theoretical and practical advances in system biology.

From an evolutionary viewpoint, we expect that solving this problem will pinpoint the environmental conditions that are responsible for the specification of lineages or microbial strains. This question is particularly vivid considering drastic environmental condition changes that are expected in a near future.

References

1. Bordbar, A., Monk, J.M., King, Z.A., Palsson, B.O.: Constraint-based models predict metabolic and associated cellular functions. Nat. Rev. Genet. **15**, 107–120 (2014)
2. Lewis, N.E., Nagarajan, H., Palsson, B.O.: Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat. Rev. Microbiol. **10**, 291–305 (2012)
3. Burgard, A.P., Pharkya, P., Maranas, C.D.: Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. **84**, 647–657 (2003)
4. Tepper, N., Shlomi, T.: Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics **26**, 536–543 (2010)
5. Larhlimi, A., Basler, G., Grimbs, S., Selbig, J., Nikoloski, Z.: Stoichiometric capacitance reveals the theoretical capabilities of metabolic networks. Bioinformatics **28**, i502–i508 (2012)
6. Pál, C., Papp, B., Lercher, M.J.: Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. Nat. Genet. **37**, 1372–1375 (2005)
7. Papp, B., Pál, C., Hurst, L.D.: Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature **429**, 661–664 (2004)
8. Ibarra, R.U., Edwards, J.S., Palsson, B.O.: Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature **420**, 186–189 (2002)
9. Koonin, E.V.: The Logic of Chance: The Nature and Origin of Biological Evolution. FT Press, New Jersey (2011)
10. Van Valen, L.: A new evolutionary law. Evol. Theory **1**, 1–30 (1973)
11. Van Valen, L.: Molecular evolution as predicted by natural selection. J. Mol. Evol. **3**, 89–101 (1974)
12. Morris, J.J., Lenski, R.E., Zinser, E.R.: The black queen hypothesis: evolution of dependencies through adaptive gene loss. MBio **3**(2), e00036-12 (2012)
13. Yang, H., Roth, C.M., Ierapetritou, M.G.: A rational design approach for amino acid supplementation in hepatocyte culture. Biotechnol. Bioeng. **103**, 1176–1191 (2009)
14. de Figueiredo, L.F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J.E., Schuster, S., Planes, F.J.: Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics **25**, 3158–3165 (2009)
15. Goldstein, Y.A.B., Bockmayr, A.: A lattice-theoretic framework for metabolic pathway analysis. In: Gupta, A., Henzinger, T.A. (eds.) CMSB 2013. LNCS, vol. 8130, pp. 178–191. Springer, Heidelberg (2013)
16. Chowdhury, A., Zomorrodi, A.R., Maranas, C.D.: Bilevel optimization techniques in computational strain design. Comput. Chem. Eng. **72**, 363–372 (2015)
17. Saharidis, G.K., Ierapetritou, M.G.: Resolution method for mixed integer bi-level linear problems based on decomposition technique. J. Glob. Optim. **44**, 29–51 (2008)
18. Xu, P., Wang, L.: An exact algorithm for the bilevel mixed integer linear programming problem under three simplifying assumptions. Comput. Oper. Res. **41**, 309–318 (2014)
19. Saharidis, G.K.D., Conejo, A.J., Kozanidis, G.: Exact solution methodologies for linear and (mixed) integer bilevel programming. In: Talbi, E.-G. (ed.) Metaheuristics for Bi-level Optimization. SCi, vol. 482, pp. 221–245. Springer, Heidelberg (2013)