High-dimensional Pareto Surfaces in the Genetic Design of Escherichia Coli

Costanza J (1), Zammataro L (2), Liò P (3), Nicosia G (1)

(1) Department of Mathematics and Computer Science, University of Catania, Italy
(2) Department of Translational Medicine, University of Milan, Italy
(3) Computer Laboratory University of Cambridge, UK

Motivation
In this work, we present a novel multi-objective optimization algorithm that computes Pareto optimal tradeoff surfaces for identifying genetic manipulations leading to targeted overand under-productions. These surfaces provide key information of the phase space of the outcome of best genetic design strategies, i.e., the result of genetic knockouts. Our algorithm performs the simultaneous optimization of multiple cellular functions (i.e., multiple objectives), while minimizing the knockout cost; it also selects those genetic designs with greater in silico production of desired metabolites. Knockouts are modeled in terms of gene sets that can affect one or more reduced reactions using gene-protein reaction mapping.

Methods
The genome-scale FBA model of E. coli, iAF1260 consists of three parts. From m metabolites and n reactions, we form an mxn stoichiometric matrix S; whose ij-th element Sij is the stoichiometric coefficient of metabolite i in reaction j. The vector of flux values v; whose j th element vj is the ux though reaction j, are constrained by a lower bound vector a and an upper bound vector b. Finally, the linear objective is formed by multiplying the axes by an objective vector f; whose j th element fj is the weight of reaction j in the biological objective (Biomass). To allow the algorithms to function at the genetic level, we used gene protein-reaction (GPR) mappings. GPR mappings define how certain genetic manipulations affect reactions in the metabolic network. For a set of L genetic manipulations, the GPR mappings is summarized with an Lxn matrix G; where the lj th element Glj of G is 1 if the lth genetic manipulation maps onto reaction j and is 0 otherwise. To knockout the genes we declared the knockout vector y, whose l th element yl is equal to 1 if the gene involved in manipulation l is knocked out and 0 otherwise.

Results
As matter of comparison we report the following: GDLS algorithm, OptFlux algorithm and OptKnock algorithm. GDLS algorithm performs a single-objective optimization; it optimizes the synthetic objective function Acetate, obtaining 15.914 mmolh-1gDW-1 with knockout cost (kcost) equal to 15. For the second synthetic objective function, Succinate, GDLS obtains 9.727 mmolh-1gDW-1 with kcost= 26. The biomass is constant:
OptFlux algorithm uses two meta-heuristics which obtain the following results: OptFlux with Evolutionary Algorithm reaches Acetate = 15.138 mmolh⁻¹gDW⁻¹ and Succinate = 9.874 mmolh⁻¹gDW⁻¹, while OptFlux with Simulated Annealing obtains Acetate = 15.219 mmolh⁻¹gDW⁻¹ and Succinate = 10.007 mmolh⁻¹gDW⁻¹. The designed algorithm performs a multi-objective optimization obtaining the following results. Acetate 21.901 mmolh⁻¹gDW⁻¹, Succinate 12.720 mmolh⁻¹gDW⁻¹ and Biomass 0.050 h⁻¹ turning off a single gene b0918 (kdsB), that is, kcost = 1. The algorithm discovers hundreds of non-dominated solutions with above Acetate and Succinate values but different kcosts; with kcost = 1 the algorithm discovered two distinct genetic design strategies (gene b0918, and gene b3867, hemN), with kcost = 2 there are 8 genetic design strategies, with kcost = 3 it is possible to use 11 distinct genetic design strategies and so on. The study of genes and reactions of E. Coli has involved inferring 16 Pareto trade-offs in anaerobic conditions. The main purpose is to find the genetic design that maximizes the desired fluxes (more than one) and biomass, with a lower kcost. We also performed a four-objective optimization, to maximize acetate, succinate, biomass and simultaneously to minimize kcost. Anaerobic conditions were simulated as zero oxygen and 10 mmolh⁻¹ available glucose; aerobic condition as 5 and 10 mmolh⁻¹ available oxygen and glucose. We computed: Succinate vs. Biomass and Acetate vs. Biomass trade-off. In aerobic conditions (10 mmol/h oxygen), we compared the designed algorithm against GDLS. For Acetate, GDLS found 23.1522 mmolh⁻¹gDW⁻¹ with kcost = 4, while for Succinate 9.2704 mmolh⁻¹gDW⁻¹ with kcost = 13. The designed algorithm obtains the following results Acetate = 27.0799 mmolh⁻¹gDW⁻¹, Succinate = 15.8250 mmolh⁻¹gDW⁻¹, with kcost = 6. One of the goals of the present research work is to use the Pareto optimal solutions of E. Coli in order to produce useful metabolites and effective drugs. The algorithm scales effectively as the size of the metabolic system and the number of genetic manipulations increase. We clearly outperform the GDLS heuristic, OptFlux, OptKnock and other heuristics, search methods, global and local optimization algorithms. Moreover, the results obtained show that the multi-objective approach is very suitable for the genetic design strategies (GDS) discovering. To our knowledge, this is the first study on multi-objective optimization for the GDS problem and in the characterizing of biological pathway in terms of Pareto optimal fronts.

Contact email
costanza@dmi.unict.it