ELSEVIER

Predicting genetic engineering targets with Elementary Flux Mode Analysis: a review of four current methods

Research Paper

David E. Ruckerbauer[1,2], Christian Jungreuthmayer[1,2] and Jürgen Zanghellini[1,2]

[1] Austrian Centre of Industrial Biotechnology, Muthgasse 11, A1190 Vienna, Austria
[2] Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria

Elementary flux modes (EFMs) are a well-established tool in metabolic modeling. EFMs are minimal, feasible, steady state pathways through a metabolic network. They are used in various approaches to predict targets for genetic interventions in order to increase production of a molecule of interest via a host cell. Here we give an introduction to the concept of EFMs, present an overview of four methods which use EFMs in order to predict engineering targets and lastly use a toy model and a small-scale metabolic model to demonstrate and compare the capabilities of these methods.

Introduction

Microorganisms as production hosts

Currently a wide variety of microorganisms are used as cell factories for the sustainable production of chemical commodities, ranging from complex molecules such as therapeutic proteins [1] to flavors and fragrances [2] to bulk chemicals [3,4] such as amino acids for animal feed. However, most microorganisms that natively produce these substances did not evolve to allow for economical production of these compounds. Some microorganisms even lack the anabolic pathways to synthesize the desired products at all. In both cases microorganisms have to be modified to increase their productivities and yields and reach economically competitive levels. To optimize productivity and/or yield one has to change the cell's environment, the cell itself or both. Changing the growth media or the culture conditions and introducing new genes and deleting others are examples for different optimization strategies [5–7].

In essence, two approaches to strain optimization can be identified: targeted and untargeted optimization. The latter uses methods such as random mutagenesis or directed evolution. By increasing the mutation rate of an organism or by exposing cells to (increasing) stress, favorable strains can be found by screening for cells with the desired characteristics [8]. The advantage of these methods is that little knowledge about the organism is required, but this is in part counteracted by the amount of screening that is required to identify favorable cells.

Targeted approaches are based at least on the partial knowledge of an annotated genome. This knowledge is particularly useful when non-native production hosts are selected. These hosts might be useful due to their favorable tolerance to specific environments, or their ease of handling or because of patent issues, requiring alternative hosts. In comparison to untargeted approaches, targeted approaches may save time and money because large-scale screening becomes unnecessary and undesired side effects could be avoided. The question now is how to identify these targets for genetic modifications. Some engineering targets might be obvious, like, for instance overexpressing an enzyme that is directly responsible for the production of the product of interest. Generally speaking, however, in a metabolic network of thousands of genes, metabolites and reactions, the optimal engineering strategy will not be obvious. Arguably, the most successful methods for identifying engineering targets *in silico* are based on constraints based reconstruction and analysis (COBRA) [9].

Constraints based reconstruction and analysis (COBRA)

A (genome-scale) metabolic reconstruction is a collection of (all) biochemical reactions in an organism. The stoichiometric matrix represents the reconstruction by an ordered collection of the stoichiometric coefficients of all reactions. Its knowledge is key for all COBRA-methods, as it enables the analysis of all feasible,

Corresponding author:. Zanghellini, J. (juergen.zanghellini@boku.ac.at)

http://dx.doi.org/10.1016/j.nbt.2015.03.017
1871-6784/

Please cite this article in press as: Ruckerbauer, D.E. et al., Predicting genetic engineering targets with Elementary Flux Mode Analysis: a review of four current methods, New Biotechnol. (2015), http://dx.doi.org/10.1016/j.nbt.2015.03.017

steady state metabolic processes that can take place in the cell [10,11]. In essence, the stoichiometric matrix together with the steady state assumption allows identifying all flux distributions that conserve the total mass. However, not all of these solutions may be biologically relevant. By adding additional constraints, COBRA-methods try to identify and characterize only the relevant solutions. For instance, by integrating the thermodynamic knowledge about the reversibility of reactions the solution space can be further restricted.

In the last two decades a multitude of COBRA-methods have been developed. All of them fall into two major groups: biased and unbiased methods [12]. Biased COBRA-methods use a biologically motivated optimization principle for their analysis. The difficulty with biased COBRA methods is to justify the choice of the optimization objective. Probably the best-known method is flux balance analysis (FBA), which often utilizes maximization of growth to analyze or predict phenotypes [13]. Although maximization of growth has been shown to be an excellent proxy to find biologically relevant flux distributions in wild type organism, the method becomes less reliable in mutants. The reason for that is that evolutionary pressure selects for the fittest strains. However, after mutation of a strain, it needs time to adapt to the new situation and these mutants are therefore often characterized by suboptimal flux distributions until adapted. This becomes particularly worrisome for the rational design of cell factories where many genetic alterations are applied simultaneously [14].

On the other hand, unbiased methods aim to characterize the full available solution space in terms of some elementary pathways. One especially promising approach is elementary flux mode (EFM) analysis [15,16]. In contrast to FBA, which yields a single solution that optimizes a particular objective, EFM analysis captures the full metabolic capabilities of the entire network.

Elementary flux modes (EFMs)

An EFM is a feasible and minimal steady state pathway that obeys all reversibility constraints [15,16]. Here, minimal means that if one EFM-supporting reaction is removed, then the whole pathway is disabled. As a consequence of the minimality condition, EFMs are also unique up to a scalar factor. In a properly (mass) balanced network, EFMs can either be loops inside the network or connect two or more external metabolites. Usually EFM-loops are removed from the complete set of EFMs, as they are thermodynamically infeasible.

The power of EFM analysis sits in the ability to describe every feasible flux distribution by a linear combination of EFMs using only non-negative scalar factors [17]. In other words, the complete set of EFMs spans the feasible flux space of a model, i.e. the set of all EFMs describes all metabolic capabilities of the organism. EFMs can therefore be thought of as basic minimal functional units. For instance, by analyzing the EFMs one is able to single out the pathways with the highest (lowest) feasible yields.

The bottleneck when working with EFMs is calculating them, as it is both time consuming and computationally demanding [18]. Currently, only small or medium scale models are used, which generally comprise of the central carbon metabolism with the addition of a few selected pathways of interest. Yet even such small networks containing a hundred reactions may give rise to several million EFMs [19].

There are several tools available to calculate all EFMs or at least a subset of them in small to medium-scale metabolic models [20–28]. We specifically single out efmtool [20], which is, to the best of our knowledge, the fastest tool currently available [29]. In the following we will assume that a complete EFM analysis of the network is possible and discuss and analyze four different methods that utilize EFM analysis for the prediction of genetic intervention targets that turn a wildtype organism into an optimized cell factory.

Methods for EFM analysis

EFM-based strain design methods

EFM analysis allows a complete characterization of the metabolic capabilities of an organism.

Thus, if all the EFMs are known, the question of strain design becomes a question of finding the 'best' intervention strategy which keeps desired network states (i.e. EFMs) and disables unwanted EFMs in order to canalize all available resources toward the product of interest.

In the following we will use a toy model (Fig. 1) introduced by Klamt et al. [30] to illustrate the basic ideas of four EFM-based strain design methods. The model consists of eight internal metabolites (S, B, C, D, E, F, G, P) and 14 reactions (R1-R12, Pex and biomass production). Furthermore there are four external metabolites, which are not in steady-state: a substrate (S external), a side product (F external), a product (P external) and biomass, the production of which is given by the equation $4C + D + 4E + G \rightarrow Biomass$. This toy model yields 16 EFMs, which are listed in Table 1. Note that the EFMs were normalized with respect to substrate uptake, R1, and that a loop, consisting of R6 and R7, has been omitted.

The toy model is available as an sbml-file in the supplementary materials, S1.

As a way to obtain a graphical representation of EFMs we plot a 2D-projection of the EFMs onto a plane and use two fluxes of interest, such as product secretion and biomass production, as coordinates (see e.g. Fig. 2). In this plot every EFM (or group of EFMs with the same flux through the depicted reactions) is represented by a circle. The color-coding represents the efficiency of the EFM, which we define as the product of the normalized biomass secretion and the normalized product secretion [31]. Since every linear combination (with positive coefficients) of EFMs is again a feasible flux-state, the space of all feasible flux distributions is given by the envelope around all EFMs, i.e. a polygon where all EFMs are either on one of the edges or inside.

It is important to note three things: (i) since every multiple of an EFM is again an EFM, they are normalized, usually with respect to carbon uptake, (ii) loops are removed, because they are thermodynamically infeasible and (iii) the null-vector, which formally is an EFM, is omitted as well, since it is biologically irrelevant.

It would not make sense to include the EFM consisting of zeros only into the polygon, as no point between this particular EFM and any other EFM could be reached by a linear combination of the (normalized) EFMs.

This space of possible flux distributions is sometimes called the phenotypic space.

We will use these phenotypic space plots and analyze the changes in the distribution of (operational) EFMs upon different metabolic interventions on the EFMs. We refer to a specific phenotypic space as a 'strain-design', because different metabolic

Please cite this article in press as: Ruckerbauer, D.E. et al., Predicting genetic engineering targets with Elementary Flux Mode Analysis: a review of four current methods, New Biotechnol. (2015), http://dx.doi.org/10.1016/j.nbt.2015.03.017