Highlight on software that predicts overexpression strategies for overproduction of metabolites

Wenfa Ng

Citizen scientist, Singapore, Email: ngwenfa771@hotmail.com

Abstract

By tuning the expression of target genes through a combination of overexpression and downregulation, desired molecules of interest or metabolites could be overproduced in selected microbial hosts. However, the challenge lies in identifying the genes and pathways whose expression level needs to be modulated to help achieve a desired overproduction of a specific metabolite or molecule of interest. One approach is through trial and error combinatorial experimentation where individual genes and pathways are modulated to understand their effect on the production of a target molecule. Although this approach could be aided by expert knowledge gained from metabolic engineering, it remains tedious and inefficient and might not arrive at a global optimum for the gene expression system. Mathematical modelling approaches for understanding metabolism at the whole cell level such as metabolic flux analysis (MFA) could provide steady state readouts of metabolic flux into different pathways, but large system of equations render the approach only solvable by computer. While MFA could help the metabolic engineer understand how modulation of expression level of a gene could impact on production of a target metabolite, the approach still relies on trial and error computer simulation to arrive at a possible approach for overexpressing particular sets of genes. Hence, a method is needed to predict the overexpression strategies (i.e., genes and pathways whose modulation is necessary) for overproducing a particular metabolite. Writing in *Metabolic Engineering Communications*, Wang and coworkers developed UP Finder, a MATLAB based COBRA extension that could identify overexpression strategies for targeted overproduction of metabolites. Using genome-scale metabolic model as input, specification of the target metabolite would lead to the software yielding a list of genes useful for gene expression modulation for overproducing a product. Further, the software ranks the list of genes in order of preference for overproduction of a particular molecule of interest. Concurrence of the set of genes identified for overexpression in lycopene and fatty acyl-ACP overproduction in *Escherichia coli* and *Synechocystis* sp. PCC6803 with those reported in the literature validated the utility of UP Finder. Collectively, software capable of identifying gene overexpression strategies in overproduction of specific metabolite fills an important gap in metabolic engineering research. Currently designed for steady state cellular analysis and metabolic optimization, future extensions of the software to analysing the dynamic state of cells would open up avenues for understanding the interconnection between gene expression dynamics and cellular physiology.

Keywords: overexpression, overproduction, genes, pathways, global optimization, metabolic flux analysis, software, constraints-based modelling, genome-scale metabolic models, steady state,

Subject areas: biochemistry, systems biology, biotechnology, microbiology, synthetic biology,

Metabolic engineering seeks to generate genetically engineered microbial strains capable of overproducing specific metabolites and target molecules. To do this, multiple genetic manipulations are typically performed at the plasmid and genome level where genes are inserted, deleted, or mutated to endow the cell with new functionality capable of directing carbon fluxes towards the pathway generating the molecule of interest. However, given the multitude of genes and pathways in the cell that cross-interact with each other through shared intermediates, deriving a strategy for overproducing a metabolite or target molecule is non-trivial, particularly from the perspective of ensuring good growth performance of the cell.

Typically, overproduction strategies involve the detailed elucidation of specific genes and pathway that require overexpression or reduced expression to help channel metabolic flux towards the target pathway and molecule. The oft-used approach in this endeavour would be through expert knowledge derived from the literature or personal experience that suggests genes and pathways whose alteration would help improve the production of desired metabolite or molecule. However, this approach might not yield global optimal solutions for improving pathway flux and metabolite concentration since enzymes and processes that drain metabolic fluxes away from the desired pathway and molecule are more difficult to identify from expert knowledge. Another problem with the expert knowledge approach lies in the inability to conduct a systematic optimization of the desired pathway and genes useful for improving overproduction of specific metabolites and molecules.

Thus, metabolic engineers turn to established methodologies such as flux balance analysis (FBA) and metabolic control analysis (MCA) to help identify strategies for improving overproduction of specific metabolites and product of interest. However, metabolic flux analysis consists of a set of mathematical techniques for solving coupled algebraic equations describing the flow of metabolic flux in different pathways. While the solution of such coupled sets of algebraic equations could be obtained through software such as MATLAB, software capable of solving the equations but without the need for programming skills would be desirable. Specifically, such user-friendly software or packages would expand the promulgation of the metabolic flux analysis technique in metabolic engineering studies.

One such software is UP Finder that implements the constraint-based modelling approach for identifying overexpression strategies for targeted overproduction of specific metabolites.[1] The software is a COBRA extension that works in the MATLAB environment. It takes a genome scale metabolic model as input and could output a list of genes that require either overexpression or downregulation for overproducing a target metabolite of interest. The outputs are quantitatively ranked to help the user identify preferable strategies for gene overexpression. Validation of the utility of the software through analysing the pathway for overproducing lycopene precursor in *Escherichia coli* and fatty acyl-ACP in the cyanobacterium *Synechocystis* sp. PCC6803 revealed that the target genes identified by the software for overexpression strategies agree closely with those reported in the literature.

Collectively, identification of overexpression strategies for overproducing specific metabolites or molecule of interest in metabolic engineering applications remain an important problem attracting much research attention. Expert knowledge or personal experience with the microbial chassis and embedded pathways remain a key source of information for identifying specific genes and pathway requiring modulation for improving the production of a downstream metabolite or target molecule. However, given the interconnected nature of metabolic pathways and the need for optimising metabolic flux through different nodes and pathways, software solutions for identifying genes and pathways requiring genetic modulation of expression levels offer a viable approach towards solving problems associated with identification of overexpression strategies. One solution available is the MATLAB-based COBRA constraint-based package UP Finder. Taking genome scale metabolic model of the organism as input as well as user identified metabolite of interest, the software is able to output a list of genes whose expression should be modulated to help increase the production of the target metabolite. More importantly, the identified genes requiring modulation of expression levels are quantitatively ranked to help the user select the genes and pathways whose expression can be most easily tuned. Future developments in software enabling the identification of overexpression strategies for targeted overproduction of specific metabolite or molecules of interest may provide information comparing the quantitative performance of modulating different metabolic pathways on the yield and titer of the target metabolite or molecule.

Conflicts of interest

The author declares no conflicts of interest.

Funding

No funding was used in this work.

References

1 Wang, X., Yu, L. & Chen, S. UP Finder: A COBRA toolbox extension for identifying gene overexpression strategies for targeted overproduction. *Metabolic Engineering Communications* **5**, 54-59, doi:https://doi.org/10.1016/j.meteno.2017.08.001 (2017).