In silico model-guided identification of transcriptional regulator targets for efficient strain design

Lokanand Koduru[1], Meiyappan Lakshmanan[2] and Dong-Yup Lee[2,3*]

Abstract

Background: Cellular metabolism is tightly regulated by hard-wired multiple layers of biological processes to achieve robust and homeostatic states given the limited resources. As a result, even the most intuitive enzyme-centric metabolic engineering endeavours through the up-/down-regulation of multiple genes in biochemical pathways often deliver insignificant improvements in the product yield. In this regard, targeted engineering of transcriptional regulators (TRs) that control several metabolic functions in modular patterns is an interesting strategy. However, only a handful of in silico model-added techniques are available for identifying the TR manipulation candidates, thus limiting its strain design application.

Results: We developed hierarchical-Beneficial Regulatory Targeting (h-BeReTa) which employs a genome-scale metabolic model and transcriptional regulatory network (TRN) to identify the relevant TR targets suitable for strain improvement. We then applied this method to industrially relevant metabolites and cell factory hosts, *Escherichia coli* and *Corynebacterium glutamicum*. h-BeReTa suggested several promising TR targets, many of which have been validated through literature evidences. h-BeReTa considers the hierarchy of TRs in the TRN and also accounts for alternative metabolic pathways which may divert flux away from the product while identifying suitable metabolic fluxes, thereby performing superior in terms of global TR target identification.

Conclusions: In silico model-guided strain design framework, h-BeReTa, was presented for identifying transcriptional regulator targets. Its efficacy and applicability to microbial cell factories were successfully demonstrated via case studies involving two cell factory hosts, as such suggesting several intuitive targets for overproducing various value-added compounds.

Keywords: Model-guided strain design, Genome-scale metabolic model, Constraint-based flux analysis, Transcriptional regulator, Systems biology

Background

Currently, a variety of value-added products can be newly synthesized and overproduced in microbial expression hosts at near-commercial levels through various pathway modifications such as gene up-/down-regulation and deletion in a serial and/or iterative manner [1, 2]. However, identifying such metabolic engineering targets is not trivial; more often than not, even the most intuitive enzyme manipulations may not lead to desired level of product yields due to the inherent regulation and complexity of metabolism [3]. To circumvent this issue, manipulating the transcriptional regulators (TRs), which often globally regulate the expression levels of a group of genes within a same cellular module in the form of regulons, has been considered as a promising strategy. For example, by fine tuning the expression of *FadR*, the TR regulating a number of genes in fatty acid biosynthesis including *fabA*, *fabB* and *iclR*, in *Escherichia coli*, fatty acid titres

*Correspondence: dongyuplee@skku.edu
[3] School of Chemical Engineering, Sungkyunkwan University, 2066, Seobu-ro, Jangan-gu, Suwon, Gyeonggi-do 16419, Republic of Korea
Full list of author information is available at the end of the article

Koduru *et al. Microb Cell Fact*   (2018) 17:167

Page 2 of 12

could be enhanced up to 73% of the theoretical yield which was not achieved by overexpressing any of the metabolic gene combinations [4]. Similarly, the global TR, *cra*, was targeted to channel more carbon flux via phosphoenolpyruvate carboxylation and the glyoxylate pathway in *E. coli*, thereby improving succinate yields [5]. Another recent study showed that the combinatorial overexpression of metabolic genes, *galP* and *glk*, along with a TR, *TyrR*, which represses the expression of multiple ʟ-phenylalanine pathway genes in *E. coli*, enhanced the yield of this amino acid significantly [6]. However, despite such several success stories, one of the major challenges is to identify more efficient and reliable TR manipulation targets.

Constraint-based metabolic modeling (CBM) is a simple and widely used approach that requires only metabolic network stoichiometry and environmental constraints to describe the cellular phenotype from genotype, and thus can be readily exploited to characterize and predict cellular behaviours under perturbed conditions [7, 8]. In this regard, several algorithms based on CBM framework have been developed for finding relevant metabolic engineering targets towards the enhanced production [9–11]. While most of these algorithms can suggest various strain design strategies via gene knockout, upregulation and downregulation [9, 12], metabolite intensification/attenuation [13] and also cofactor balancing [14, 15], only a handful of them are related to TR manipulation targeting. OptORF is the first ever constraint-based method developed for TR targeting [16] using a previously developed combined metabolic/regulatory model [17] where the transcriptional-regulatory information is described via Boolean logic, i.e. 'on' and 'off' states of TR expression. A bi-level mixed-integer linear programming (MILP) based solution procedure was proposed to identify TR manipulation targets in *E. coli* for overproducing ethanol, isobutanol and 2-phenylethanol. Later, Vilaça et al. [18] used evolutionary algorithm and simulation annealing as the optimization algorithms to find TR candidates from the same combined metabolic-regulatory model. However, the use of these methods is severely limited since it assumes the transcriptional-regulatory responses to be binary which could be continuous. In order to address this critical issue, recently, Kim et al., developed Beneficial Regulator Targeting (BeReTa), on the basis of an unintegrated approach where each TR in the transcriptional regulatory network (TRN) is ranked for genetic manipulation, i.e. up-/down-regulation, based on a beneficial score [19]. A systematic procedure was proposed to combine the regulatory strength information from the TRN and the desired flux slopes that could overproduce the desired compound.

While the unintegrated approach presented in BeReTa could effectively identify several relevant TR candidates for up-/down-regulation compared to OptORF, it still suffers from certain limitations. Firstly, BeReTa does not consider the inherent hierarchical structure of TRN; unlike metabolic genes, TRs are known to operate in a regulatory cascade when certain global TRs regulate multiple downstream TRs, all of which in turn can modulate the expression of target genes [20, 21]. Here, it should be highlighted that the regulation of TR–TR-gene in TRNs are complex which at times can be circular and negate the overall effects in a counter-intuitive manner. Therefore, it is important to incorporate the hierarchical structure of TRN while identifying the TR candidates such that the engineered TR's effect is not masked by another higher order TR. Secondly, BeReTa only takes into account the positively correlated reactions while calculating flux slopes, ignoring the reactions that are negatively correlated to the desired product which may also serve as relevant gene manipulation, i.e. down-regulation, targets. Furthermore, it does not consider the presence of equivalent competing pathways in the product synthesis that also gives rise to the same yield of product.

In this work, we propose "hierarchical-Beneficial Regulatory Targeting" (h-BeReTa), which extends the BeReTa by addressing the abovementioned shortcomings for identifying efficient TR targets. Specifically, h-BeReTa utilizes a TRN with hierarchies of TR clearly defined and a metabolic model to identify target candidates. Moreover, it also account for the negatively correlated reactions with the product flux, in addition to the positively correlated reactions because the flux through these reactions need to be minimized to improve product synthesis. Here, we first describe the methodology of h-BeReTa, and then demonstrate its applicability by identifying promising TR manipulation targets for overproducing various compounds in *E. coli* and *C. glutamicum*. Finally, we compare the resulting targets obtained from h-BeReTa to its preceding methods and discuss their performance.

Methods

h-BeReTa algorithm

h-BeReTa aims to identify the relevant TR targets for up-/down-regulation to overproduce the desired product using an unintegrated approach which was previously proposed by BeReTa [19]. Initially, constraint-based flux analysis is used to identify the reactions that are both positively and negatively correlated with the desired product flux across the entire metabolic network. Subsequently, the algorithm identifies the corresponding TRs which modulate the expression of these reactions and the strength of their regulation. Finally, each TR is scored based on its regulatory strength, position in the

Koduru *et al. Microb Cell Fact*   (2018) 17:167

Page 3 of 12

TRN hierarchy and their association with product flux either in the positive or negative manner. The candidate genes with highest and lowest scores from the ranked list can then be chosen for their up- and down-regulations, respectively. The scoring procedure involves five key steps as summarized in Fig. 1.

Step 1: Identification of reactions correlated with product flux (*nRAP*)

nRAP represents the extent of control a particular reaction flux could have on the product flux. Calculation of *nRAP* involves a constraint-based flux analysis formulation as shown below:

$$\text{Maximize } v_{target}$$
$$\text{Subject to : } \sum_n S_{mn} \cdot v_n = 0$$
$$\alpha_n \leq v_n \leq \beta_n$$
$$v_{biomass} \geq 0.5 \cdot v_{biomass}^{\max} \tag{P1}$$
$$v_r = v_{\min} + k \cdot (v_{\max} - v_{\min})$$
$$\forall k = 0, 0.1, 0.2 \dots 1 \text{ and } r \subset n$$

where, v_{target} represents the product forming flux, $v_{biomass}$ is flux through biomass forming reaction and $v_{biomass}^{\max}$ represents its maximum attainable value, $v_{\min}$ and $v_{\max}$ are minimum and maximum feasible flux values of reaction
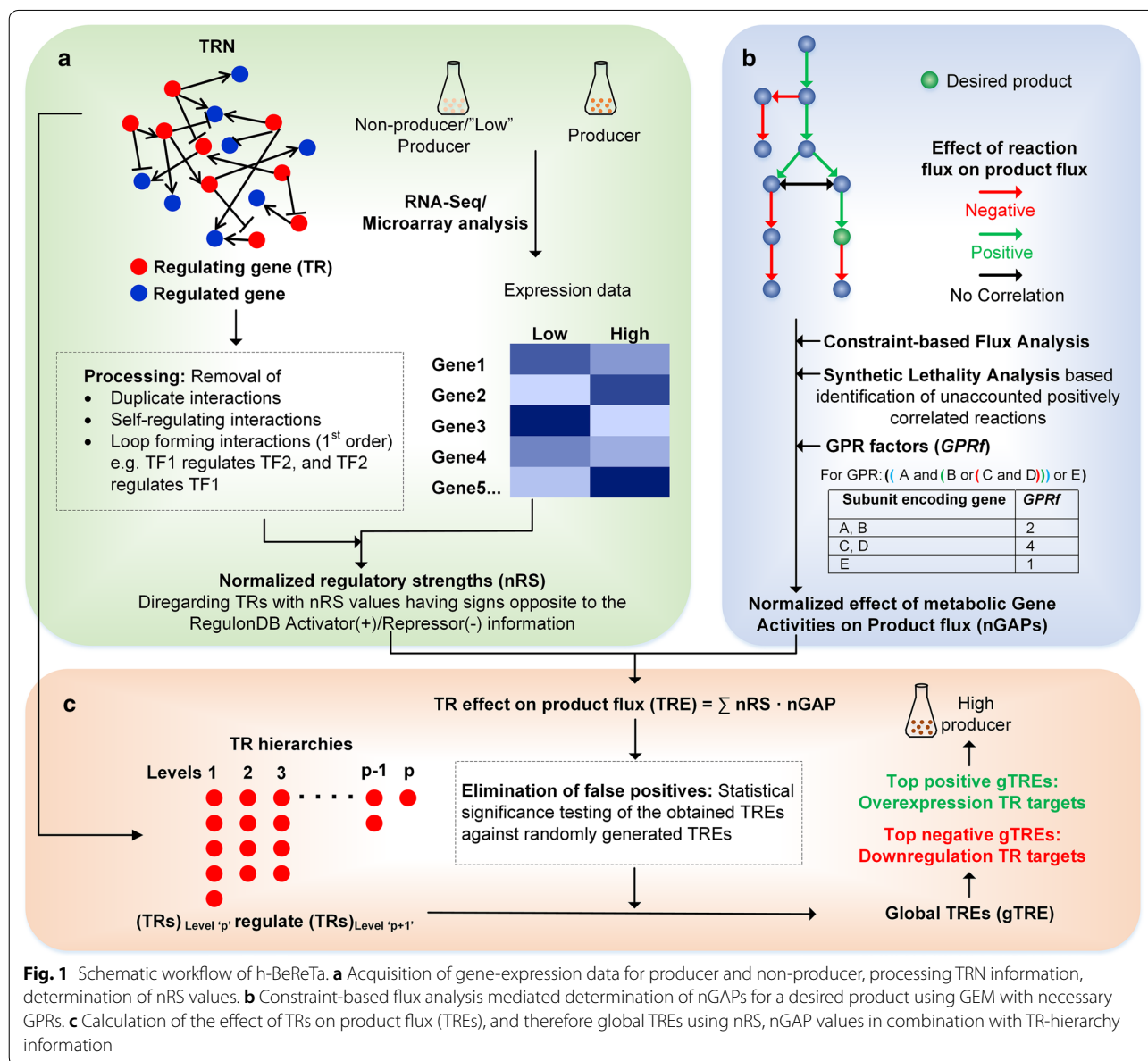
Fig. 1 Schematic workflow of h-BeReTa. **a** Acquisition of gene-expression data for producer and non-producer, processing TRN information, determination of nRS values. **b** Constraint-based flux analysis mediated determination of nGAPs for a desired product using GEM with necessary GPRs. **c** Calculation of the effect of TRs on product flux (TREs), and therefore global TREs using nRS, nGAP values in combination with TR-hierarchy information

Koduru *et al. Microb Cell Fact*   (2018) 17:167

Page 4 of 12

'r' (v_r), whose effect on v_{target} is to be determined. v_{min} and v_{max} are determined using flux variability analysis [22]. Once the linear programming (LP) problem shown above is solved for all gene-associated reactions, denoted by 'r', the normalized Reaction Activity on Product flux ($nRAP_r$) are calculated as slopes of linear plots for v_{target} versus k as shown previously [19]. Here, the fractional change 'k' is chosen instead of the absolute change in the reaction flux in order to avoid unrealistically large $nRAP$ values for reactions carrying small fluxes. $nRAP$ can take any real value depending upon the bounds used for model simulation; reactions potentially favouring product formation take a positive value and those that have potential negative impact on product flux take a negative value. Reactions which does not influence the product flux take a value, '0'. Here, it should be highlighted that the solutions obtained by solving optimization problem P1 will identify only the reactions which are present in the shortest path to the product will be ranked with positive or negative $nRAP$ values. However, there could be reactions that are part of alternative (or non-optimal) flux modes which could also be positively or negatively correlated to product formation (Additional file 1: Figure S1). Therefore, an additional optional step can be introduced in order to identify such reactions with appropriate $nRAP$ scores (Additional file 1).

Step 2: Identification of reactions correlated with product flux (*nRAP*)

Once the $nRAP$ is calculated including the alternate pathways, the normalized effect of metabolic Gene Activities on Product flux ($nGAP$) can be computed for each gene 'j' that are associated with reaction 'r' using gene-protein-reaction (GPR) association information available in the metabolic model using below equation:

$$nGAP_{jr} = \frac{nRAP_r}{GPRf_j} \qquad (1)$$

where, $GPRf_j$ is GPR factor of gene 'j', which distributes the weightage of each gene 'j' associated with the reaction 'r'. For example, if a reaction has GPR ((A and (B or (C and D)) or E), then, GPR_f of E is '1' as it can form a fully functional enzyme. On the other hand, GPR_f of A and B is '2' as each can constitute half of the multi-subunit enzyme complex; and GPR_f of C and D is '4', as they constitute quarter each.

Step 3: Calculation of normalized regulatory strength (nRS)

The regulatory strength (RS) represents the effect of TR expression on the expression of the downstream regulated gene. For a TR, 'i' regulating the metabolic gene 'j', the normalized regulatory strength (nRS_{ij}) can be calculated as below:

$$nRS_{ij} = \left(\frac{Gene_{j,prod} - Gene_{j,n\text{-}prod}}{TR_{i,prod} - TR_{i,n\text{-}prod}} \right) * \left(\frac{TR_{i,n\text{-}prod}}{Gene_{j,n\text{-}prod}} \right) \qquad (2)$$

where, $TR_{i,prod}$ and $TR_{i,n\text{-}prod}$ are the expression intensities of the TR 'i' at producer and non-producer product conditions, respectively. $Gene_{i,prod}$ and $Gene_{i,n\text{-}prod}$ are the expression intensities of the gene 'j' regulated by TR 'i' at producing and non-producing conditions. It should be noted that the "producer" and "non-producer" phenotypes can even be replaced by "slightly-better producer" and "producer" phenotypes, respectively. The absolute changes of expression intensities seldom determine the extent of the 'overall effect' of the change in gene expression. Hence, a normalization factor $\frac{TR_{i,n\text{-}prod}}{Gene_{j,n\text{-}prod}}$ was multiplied to the ratio of absolute changes in the gene expression levels to yield nRS_{ij}. Here, the normalization is important because the TRs with low expression values are bound to receive relatively higher regulatory strengths compared to those with high expression levels, although the actual extent of regulation of their respective genes could be similar. Furthermore, the activator/repressor information obtained from RegulonDB is used to eliminate the TRs with nRS values having sign opposite to its known functionality.

Step 4: TR effect on product flux (*TRE*)

The effect of each TR on product flux is calculated by combining the effect of normalized regulatory strength (nRS_{ij}) and its downstream metabolic gene activities on product flux ($nGAP$):

$$TRE_i = \sum_r nRS_{ij} \cdot nGAP_{jr} \qquad (3)$$

where, $nRS_{ij} \cdot nGAP_{jr}$ represents the effect of the TR 'i' on the product flux via the gene-associated reaction 'r'.

In order to ensure that the *TRE* scores are not affected by false positives/or result of random chances, h-BeReTa calculations were performed using $nGAP$ values derived from large sets (~1000) of $nRAP$ values that are randomly generated within the observed ranges. *TRE* scores for each TR were then obtained for the 1000 randomly generated $nGAP$ sets. Subsequently, the probability of the randomly generated *TRE* to fall in ±10% range of the actual *TRE* scores of the corresponding TR is calculated. TRs with this probability less than 0.05 (5%) are considered true positives and therefore carry forwarded to Step 5.

Step 5: Global TR effects based on hierarchies of TRN (*gTRE*)

Transcriptional regulation of metabolism includes not only the TR-metabolic gene interactions but also TR–TR interactions. Such interactions are effectively represented by

Koduru *et al. Microb Cell Fact*   (2018) 17:167

Page 5 of 12

TRN, which also provide information about the hierarchies of TR–TR interactions. In such hierarchies, the effect of a certain global TR on product flux (*gTRE*) would be the sum of its own effect on product flux (*TRE*) and *gTRE*s of different TRs in the immediate downstream level of the TR-hierarchy that it regulates. Hence, as depicted in Fig. 2, the calculation of *gTRE*s should be approached from the bottom-most level which only includes TRs that regulate "non-TR" genes to the top-most level which includes global TRs along with those TRs, which are not under the control of any known TRs. This effect can be calculated as follows:

$$gTRE_p = TRE_p + \sum_{p}^{h} nRS_{p(p+1)} \cdot gTRE_{p+1} \qquad (4)$$

where, '*p*' can take any value from 1 to '*h*' and is the level immediately upstream to '*p+1*' in the TR-hierarchy (Fig. 2), '*h*' is the total number of hierarchy levels and
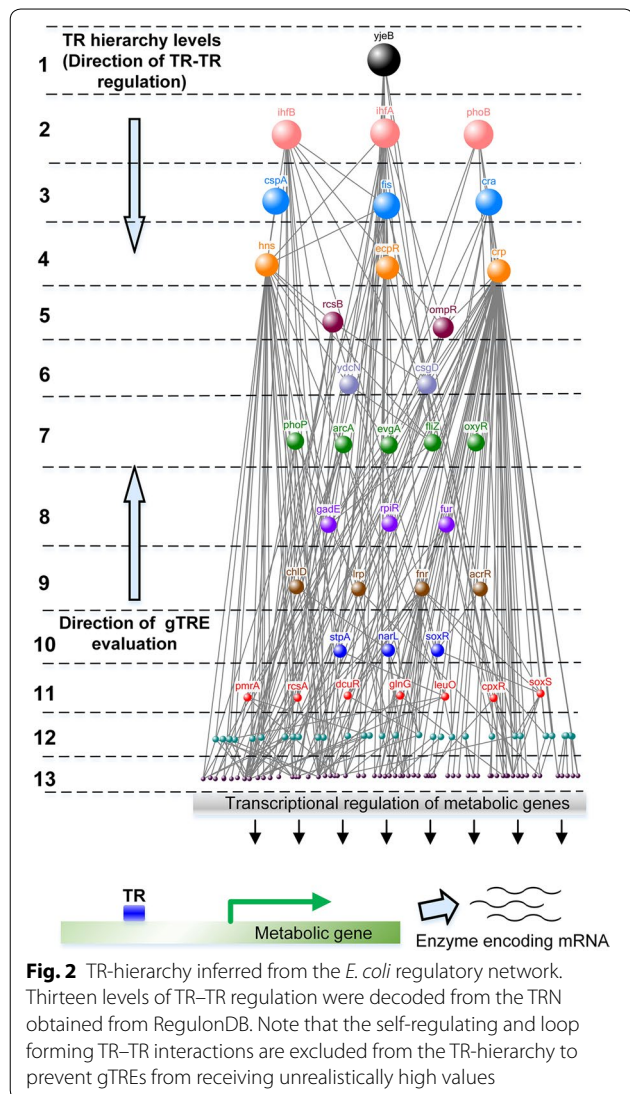
$nRS_{p(p+1)}$ is the normalized regulatory strength of TR at level '*p*' on TR at level '*p + 1*'. The summation term in the above expression exists only for those TRs that regulate other TRs and does not exist for TRs at level '*h*', i.e. the last downstream level in the hierarchy. The *gTRE*s of all TRs in the TRN, thus calculated, can be used to rank them as either overexpression or downregulation target depending on whether they receive high positive or high negative values, respectively.

In silico models and gene expression datasets

h-BeReTa require three inputs for its implementation: a genome-scale metabolic model (GEM), TRN along with the reconstructed TR-hierarchies and gene expression datasets of two reference strains.

Genome-scale metabolic models

The *i*JO1366 [23] and *i*AF1260 [24] GEMs were used to evaluate the *nGAP* values for the *E. coli* case studies, and the *i*CW773 [25] GEM was used for *C. glutamicum* case studies. All constraint-based simulations were performed using COBRA toolbox [26], implemented in MATLAB (http://www.mathworks.com) with Gurobi5 (http://www.gurobi.com) as the optimization solver. Note that FVA was performed by employing the FastLooplessFVA function, implemented in COBRA toolbox [26] which uses a fast sparsification algorithm to efficiently eliminate the thermodynamically infeasible loops [27].

Transcriptional regulatory networks

The TRN information of *E. coli* was downloaded from RegulonDB version 9.0 [28] including a total 4787 TR-gene interactions and 200 TRs. The TRN information of *C. glutamicum* was obtained from the Abasy Atlas database [29] which accounts for 3330 TR-gene interactions excluding self-regulators and 102 TRs. Here, it should be noted that the levels of TR regulation hierarchy were manually reconstructed from the RegulonDB TRN based on the TR–TR interaction relationships.

Gene expression datasets

Apart from a metabolic model and TRN, h-BeReTa requires two specific gene-expression datasets relevant to the desired phenotype, i.e. "producer vs. non-producer", for the identification of promising TR-manipulation targets. Such datasets can be obtained from the two different phases of a cell culture, e.g. growth vs. stationary phase, which shows differential transcriptional regulation. Alternatively, gene expression datasets obtained while comparing a wild-type to that of a transcriptional regulator engineered mutant can also be used for this purpose. Note that the gene expression datasets used

Fig. 2 TR-hierarchy inferred from the *E. coli* regulatory network. Thirteen levels of TR–TR regulation were decoded from the TRN obtained from RegulonDB. Note that the self-regulating and loop forming TR–TR interactions are excluded from the TR-hierarchy to prevent gTREs from receiving unrealistically high values

Koduru *et al. Microb Cell Fact* (2018) 17:167

Page 6 of 12

are product-specific unlike BeReTa, which uses a general gene expression compendium for all products. The expression datasets for the case studies involving the production of tyrosine, acetate and fatty acids were downloaded using the GEO accessions provided in references cited for the respective case studies (see "Results").
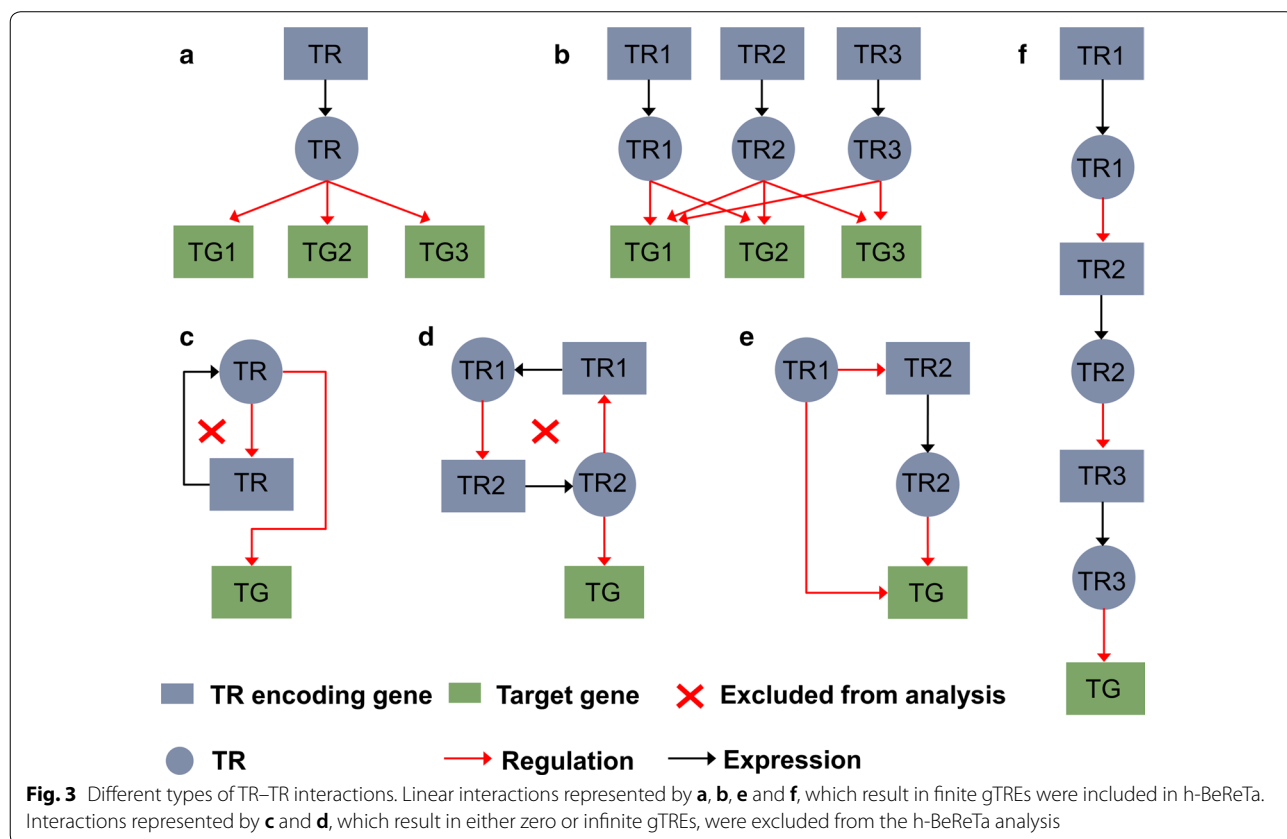
Results

Application of h-BeReTa to *Escherichia coli*

Escherichia coli is one of the well-studied microbes, and a commonly used cell factory for producing various value-added compounds due to the ease of gene manipulations with abundantly available genetic engineering tools. Hence, to take advantage of such valuable resources, we used *E. coli* to demonstrate the applicability of h-BeReTa (The Matlab code for h-BeReTa is provided in https://github.com/lokanandk/h-BeReTa). We used the RegulonDB information to manually retrieve several hierarchies from the *E. coli* TRN. A total of 13 levels of TR hierarchy were obtained as compared to the previously described five levels ([20, 21]; Fig. 2). It should be noted that the self-regulating and loop forming interactions are excluded from all the levels since the net effect of a TR–TR interaction causing negative feedback will be zero and that causing positive feedback will be infinite (Fig. 3). Such exclusion prevents the assignment of potentially very high or unrealistic values to *gTRE*s. Using the reconstructed hierarchical TRN, we then applied the h-BeReTa algorithm to identify best TR manipulation targets for the overproduction of five products, including acetate, tyrosine, fatty acids, lycopene and menaquinone. All TR targets thus obtained were comprehensively mined against the published literature to retrieve possible true or false positive evidences, if any.

Acetate

Overflow metabolism is a cellular process observed in *E. coli* during higher aerobic growth rates, characterized by wasteful energy dissipation in the form of secretion of the fermentative byproducts such as acetate. Excess accumulation of acetate by *E. coli* causes growth inhibition, thus affecting product yields as a consequence of loss of useful carbon. Previous studies have clearly demonstrated the role of transcriptional regulation in this metabolic phenomenon. Hence, the gene expression data for the growth of *E. coli* in a chemostat [30] was retrieved from GEO, and the h-BeReTa was applied to identify the TR manipulation targets. The top overexpression and downregulation targets are presented in Table 1. The effects of several TRs on acetate production were reported in literature, importantly many of which are consistent with the top targets. It has been observed that several of these

Fig. 3 Different types of TR–TR interactions. Linear interactions represented by **a**, **b**, **e** and **f**, which result in finite gTREs were included in h-BeReTa. Interactions represented by **c** and **d**, which result in either zero or infinite gTREs, were excluded from the h-BeReTa analysis

Koduru *et al. Microb Cell Fact*   (2018) 17:167

Page 7 of 12

TRs control acetate flux by directly regulating enzymes involved in glycolysis, TCA cycle and glyoxylate pathway [30–32].

Tyrosine

Tyrosine has been used for a wide range of industrial and pharmaceutical applications as dietary supplements and precursors for the synthesis of benzylisoquinoline alkaloids and polyketides. Several metabolic engineering strategies have been carried out to increase the production of tyrosine in *E. coli*. Here, we apply h-BeReTa for the overproduction of tyrosine using the expression data obtained from the mutagenesis libraries of the global transcription factors rpoA and rpoD using a high tyrosine-yielding engineered parental strain [33]. The constraint-based simulations for *nGAP* determination were performed on *i*JO1366 GEM with flux through the reaction catalysed by prephenate dehydratase (*pheA*) constrained to zero, in order to mimic the metabolic state of the engineered parental strain. Interestingly, many of the TR targets identified correspond to those regulating the pool of phosphoenolpyruvate (PEP), an early precursor for tyrosine biosynthesis (Table 1). This observation clearly indicates that despite undergoing sufficient modifications in the downstream module of tyrosine biosynthesis such as the deletion of repressor gene *tyrR*, deletion of *pheA* and overexpression of feedback resistant 3-deoxy-ᴅ-arabinoheptulosonate-7-phosphate synthase (*aroG*^fbr) and chorismate mutase/prephenate dehydrogenase (*tyrA*^fbr) [33], it still has some room for further improvement. Since the original dataset reported three different regulatory modifications (*rpoA14*, *rpoA27*, and *rpoD3*) [33], we further tested the consistency of h-BeReTa predictions across all three cases. Overall, we

could predict similar TR targets across all three cases using gene expression datasets which are obtained under different regulatory conditions, thus clearly indicating the robustness of h-BeReTa.

Fatty acids

Although bacterial hosts have been found to be a less appealing than yeasts for the industrial production of free fatty acids [34, 35], the tremendous potential of fatty acids and their derivatives for pharmaceutical and cosmetic applications and the ease to genetically manipulate have driven numerous engineering efforts in *E. coli*. The fatty acid metabolism in *E. coli* is extensively regulated at transcriptional level, and hence their overproduction would require significant interventions in the associated TRs [36]. Here, we use the expression data generated by one such study [4] to rank TR-manipulation targets for fatty acid overproduction. A synthetic reaction representing fatty acid biosynthesis was added to *i*JO1366 GEM to perform constraint-based simulations. h-BeReTa identified relevant TR overexpression and downregulation targets for fatty acid overproduction where at least three out of the four TR targets for fatty acid overproduction that were validated using experimental evidences either activate or repress fatty acid degradation (Table 1), suggesting the dominant role of β-oxidation in controlling fatty acid accumulation in *E. coli*.

Lycopene

Lycopene is known to be an antioxidant and a potential cancer therapeutic agent, and thus, numerous attempts have been made to produce it using engineered *E. coli* as

Table 1 Top-five along with additional validated (if any) transcriptional regulator targets identified by h-BeReTa for overproducing various compounds in *E. coli* and its comparison previously existing methods

Product	Nature of target	h-BeReTa	BeReTa	OptORF
Acetate	Upregulation	nac, ihfA, tyrR, rpiR, fliZ	tdcA, tdcR, **argP** [52]	–
	Downregulation	*cra* [31, 59], *crp* [60], *oxyR* [17], *fur* [61], fnr, *phoB* [62], argP [52]	*cra* [31, 59]	–
Tyrosine	Upregulation	*soxR* [63, 64], *soxS* [63, 64], *fadR* [65], rcsB, arcA, *cra* [53]		–
	Downregulation	acrR, adiY, gadW, flhD, argP, *tyrR* [33]	**cra** [53], *tyrR* [33], trpR, yebP	–
Fatty acids	Upregulation	fis, fnr, rpiR, yjeB, torR, *ompR* [66], *fadR* [4], **pdhR** [54]	*fadR* [4]	–
	Downregulation	fur, arcA, rcsB, oxyR, cra, *crp* [67, 68]	*pdhR* [54], cra, yijC	–
Lycopene	Upregulation	arcA, gadX, fis, ihfA, soxS	soxS	–
	Downregulation	rcsB, gadE, rcsA, ihfB, cra	cra, yebP	–
Menaquinone	Upregulation	arcA, nadl, trpR, cra, atoC		–
	Downregulation	rob, fnr, creB, tdcR, tdcA	iclR	–
Ethanol	Upregulation	fur, oxyR, glpR, envY, dnaA, **arcA** [69]	fur	–
	Downregulation	rbsE, cpxR, *ihfB* [56], *fnr* [56], *crp* [56]	*pdhR* [70], **cra** [71], gntR, kdgR	arcA-pgi, fnr-gntR-pflB-tdcE-tpiA

TRs are provided in decreasing absolute *gTRE* trends. TR targets highlighted in italic letters are true positives (TP) and those underlined are false positives (FP)

Koduru *et al. Microb Cell Fact*   (2018) 17:167

Page 8 of 12

host [37]. Initially, it has been shown that lycopene can be produced in *E. coli* via mevalonate [38] and non-mevalonate pathways [39]. However, with an increased interest for lycopene, alternative strategies are being actively sought to further enhance its yields [37]. In this regard, one of the earlier study showed that a point mutation in the global regulator, cAMP receptor protein (CRP), resulted in significant improvements to lycopene yield in *E. coli*, indicating the potential of transcriptional regulator engineering approach for lycopene production [40]. Here, we used the gene expression data obtained from the study for an *E. coli* K12 strain capable of producing lycopene and its derivative harbouring the mutant *crp* gene to predict TR engineering targets. The h-BeReTa results for TRs targets potentially improving lycopene production are listed in Table 1. The identification of *soxS*, the TR part of the *soxRS* regulon involved in relieving oxidative stress [41], as an up-regulation target is consistent with the previous observations: measurable lycopene content decreased with increasing oxidative stress [42]. Further, it should be noted that since a major portion of lycopene biosynthesis overlaps with the canonical isoprenoid biosynthesis, the TR targets obtained here can be generalized for the production of other carotenoid metabolites in *E. coli*.

Menaquinone (vitamin K2)
Vitamin K2 or menaquinones is a group of molecules is essential for healthy arteries and bones whose deficiency in humans could result in osteoporosis, impairment in blood coagulation and cardiovascular disease [43]. The average intake of vitamin K among the adults in the United States has been estimated to be only about 70–90% of the recommended intake value [44], emphasizing the relevance of its large-scale production to pharmaceutical and food industries. The pathway of menaquinone biosynthesis, which partially overlaps with that of aromatic amino acid and isoprenoid biosynthesis is subjected to a high level of transcriptional regulation. In this regard, we use the gene expression data obtained for the wild type *E. coli* and a mutant strain accumulating higher menaquinone pool to predict potential TR targets for vitamin K2 overproduction (Table 1). The prediction of *trpR* as an upregulation target is interesting as it represses the aromatic amino acid biosynthesis which also competes for chorismate, a common precursor for both compounds.

Application of h-BeReTa to *Corynebacterium glutamicum*
In this work, we also applied h-BeReTa to *C. glutamicum,* an industrially important gram-positive bacterium and a representative host lesser studied compared to *E. coli*, in order to test its wider applicability. The most

comprehensive TRN of *C. glutamicum* available to date [29] was used to retrieve six levels of top–down TR hierarchy (Additional file 1: Table S3). We specifically applied h-BeReTa in *C. glutamicum* to identify the TR manipulation targets for glutamate, an amino acid which it naturally produces under several conditions, and lycopene.

Glutamate
Corynebacterium glutamicum is widely used for the production of several amino acids, especially glutamate. Recently, it has been shown that *C. glutamicum* can secrete glutamate in larger amounts when exposed to the antibiotic, ciprofloxacin [45]. The gene expression data obtained in this study was therefore used here to understand the transcriptional regulation and to identify TR candidates that potentially augment the glutamate production in *C. glutamicum*. The top overexpression and downregulation TR targets are presented in Table 2 where at least one TR target each among the each category has been already reported in literature. Importantly, the prediction of *glxR* and *ramA* as overexpression targets has direct implications in decreasing glutamate yield where *glxR* is a repressor of glutamine synthase [46] and malate synthase [47], and *ramA* is a repressor of malate synthase [48], both are key enzymes in glutamate biosynthetic pathway.

Lycopene
Recently, it was reported that overexpression of the housekeeping sigma factor, *sigA*, resulted in more reddish coloured cells compared to the control strain of *C. glutamicum*, indicating the overproduction of lycopene [49]. Hence, we used this gene expression data obtained to characterize the transcriptional regulation of *sigA* and to suggest other TR targets to improve lycopene production even further. Interestingly, two of the TR targets identified by h-BeReTa have been validated by the same study to either increase or decrease the lycopene yields when overexpressed in *C. glutamicum* (Table 2). Here, it should be highlighted that among all targets identified, *relA* is a promising target for lycopene production, as it

Table 2 Top-five along with additional validated (if any) transcriptional regulator targets identified by h-BeReTa for overproducing various compounds in *C. glutamicum*

Product	Nature of target	TRs
Glutamate	Upregulation	sigA, glxR, ramA, farR, *argR* [72]
	Downregulation	cg1861, sigH, lexA, sugR, sigB, *lldR* [73]
Lycopene	Upregulation	*sigH* [49], relA, nrdR, ramB, ltbR
	Downregulation	sugR, *sigB* [49], ripA, cg2544, znr

TRs are provided in decreasing absolute *gTRE* trends. TR targets highlighted in italic letters are true positives (TP) and those underlined are false positives (FP)

Koduru *et al. Microb Cell Fact* (2018) 17:167

Page 9 of 12

induces stringent response which is shown to be counter-acted by one of the enzymes (4-hydroxy-3-methylbut-2-enyl diphosphate reductase or lytB or ispH) involved in the flux limiting branch point of lycopene (isoprenoid) biosynthesis [50, 51].

Comparison of h-BeReTa with other TR-based approaches

In order to further evaluate the performance of h-BeReTa, we compared the TR targets with those obtained by BeReTa and OptORF for *E. coli* case studies. Initial comparison of the individual targets for various products from h-BeReTa and BeReTa showed a significant overlap among the resulting TRs due to the similarity in implementation (Table 1). However, h-BeReTa exclusively identified many global TRs such as *phoB*, *ihfA*, *ihfB*, *cra* and *fis* as top candidates ahead of other TRs which were commonly identified by both methods, suggesting the importance of accounting the TR-hierarchies. Moreover, the comparison also revealed a few key cases where the two methods resulted in contradicting gene manipulations for same TR targets. For example, *argP* was identified as down-regulation target for acetate production by h-BeReTa while it was predicted to be an up-regulation target by BeReTa. Similarly, h-BeReTa suggested up-regulation of *pdhR* and *cra* for improving fatty acid and tyrosine biosynthesis, respectively, whereas BeReTa predicted otherwise. To further understand why these algorithms suggested different targets for same compounds and to test their validity, we surveyed published literature. It was earlier shown that *argP* overexpression is negatively correlated with acetate production [52], confirming h-BeReTa predictions. Similarly, *cra* is shown to be a potential positive regulator of tyrosine biosynthesis through simultaneous activation of phosphoenolpyruvate synthase and repression of PTS system [53], both increasing the availability of PEP, and thus enhancing tyrosine biosynthesis. However, as *pdhR* is implicated in the repression of pyruvate dehydrogenase complex which is involved in the biosynthesis of acetyl-CoA [54], it could be a negative regulator of fatty acid biosynthesis as predicted by BeReTa. *pdhR* might therefore be a possible false positive TR target for fatty acid production predicted by h-BeReTa. We further compared h-BeReTa and BeReTa predictions through binary classification statistical tests on the basis of true positives (TP), false positives (FP) and false negatives (FN) (Additional file 1: Table S1). These results clearly demonstrate better prediction of h-BeReTa over BeReTa: it has higher sensitivity, precision and F1 score, and a low false discovery rate (Table 3). Here, it should be noted that True Negatives (TN) were not included in these tests due to the limited information available from literature sources. Finally, we also compared the results of h-BeReTa with those of OptORF for the case of ethanol production in *E. coli*. In

order to obtain TR targets for ethanol overproduction, the gene expression datasets obtained from a study involving the ethanologenic strain of *E. coli* K12 grown in glucose minimal medium and a synthetic hydrolysate medium was used [55]. All three methods predicted both certain common and unique TR targets (Table 1), where only h-BeReTa was able to predict several global regulators as top targets for ethanol production in good agreement with previous reports [56]. Moreover, the TR predictions by both h-BeReTa and BeReTa also to some extent depended on the version of *E. coli* genome-scale model used to evaluate the TR targets. An overall comparison of predicted targets that are validated through experimental evidences suggested that *i*AF1260 performed better than *i*JO1366 (data not shown). While *i*JO1366 yielded unrealistic *nRAP* values for h-BeReTa possibly due to the dubious feasible flux ranges obtained using FVA, the BeReTa results using *i*JO1366 were contradictory to those of *i*AF1260, cautioning the user to ensure the reliability of the *nRAP* values or flux slopes derived from the model by evaluating them on a case-by-case basis. Moreover, h-BeReTa predictions are largely affected by the completeness of the gene expression data. When random partial gene expression datasets were used in the acetate case study, h-BeReTa predicts inferior TR targets compared to those obtained from the original datasets (Additional file 1: Table S2).

Discussion

In this study, we introduced a new method, h-BeReTa, for identifying TRs which need to be up-/down-regulated for the overproduction of desired compounds. Unlike earlier methods, it accounts for the hierarchies of TRs in the regulatory cascade and also considers the reaction fluxes which compete with the product flux while identifying relevant TR candidates. h-BeReTa is able to identify efficient TR manipulation strategies as it is successfully demonstrated via several case studies of *E. coli* and *C. glutamicum* for overproducing various products

Table 3 Comparison of h-BeReTa and BeReTa through statistical binary classification tests

Measures	h-BeReTa	BeReTa
True positives (TP)	17	5
False positives (FP)	2	3
False negatives (FN)	10	10
FP/FN	0.2	0.3
Sensitivity or true positive rate (TPR), TP/(TP + FN)	0.63	0.33
Precision or positive predictive value (PPV), TP/(TP + FP)	0.893	0.625
False negative rate (FNR), 1-TPR	0.37	0.67
False discovery rate, 1-PPV	0.105	0.375
F1 score (0 = worst, 1 = best)	0.739	0.435

Koduru *et al. Microb Cell Fact* (2018) 17:167

Page 10 of 12

including acetate, tyrosine, fatty acids, menaquinone, and lycopene. Here, it is important to note that the validation of the TR target predictions was only based on those examples that are available from published literature and hence, many targets remain to be validated.

As mentioned earlier, h-BeReTa utilizes an unintegrated approach which treats the cellular metabolism and regulation as two modules in the framework and then combines them systematically as previously proposed in BeReTa algorithm. However, h-BeReTa still encompasses several differences at various levels of the formulation, thereby resulting in improved performance. First, the TR hierarchy information is newly incorporated into the framework, thus identifying TR targets with higher regulatory impact on the product formation. The importance of such considerations can be perceived from h-BeReTa results which exclusively include global TRs such as *phoB*, *ihfA*, *ihfB*, *cra* and *fis* as top candidates where several of them are experimentally validated in the literature (Table 1). In addition, the TR hierarchy can provide some clues regarding the potential outcomes of global TR targeting. For example, Fig. 2 shows that *fnr* is regulated by *ihfB* (positive), *ihfA* (positive) and *fur* (negative), which occupy the upper levels of the TR hierarchy. It is furthermore clear from Table 1 that the prediction of *fnr* as an experimentally validated downregulation target for ethanol production has been consistently translated to *ihfB* and *fur* as downregulation and overexpression targets, respectively. Another important difference between the two approaches is that h-BeReTa uses a different constraint-based flux analysis formulation in which it also takes into account reactions with negative *nRAP* scores, i.e., those reactions whose fluxes compete with product formation. The inclusion of reactions with negative *nRAP* scores is important because high-value products are often secondary metabolites which the cells does not produce naturally and experiences direct competition from a large part of the fluxes in the metabolic network which are associated with biomass precursor biosynthesis. Furthermore, the accounting of negatively correlated fluxes in h-BeReTa allows it to rank the global TRs accordingly, considering that it could regulate multiple genes in other parts of the metabolic network in addition to the product flux. In contrast, since BeReTa does not consider the negatively correlated reaction fluxes there could be a bias for TRs to be identified just by considering the positive beneficial scores calculated.

Although h-BeReTa is able to identify efficient TR targets consistently, one major limitation is the inability to predict the extent of changes to product yields as a function of TR manipulation which is mainly due to the unintegrated nature of the methodology. However, the actual increase in product yield might mainly depend on several contributing factors, including the degree of correlation between the mRNA and protein levels of the TR, nature of interaction, saturation kinetics between the TR and its regulatory targets, and the intracellular metabolite concentrations, which are generally ignored in CBM approaches. Therefore, further improvements in h-BeReTa predictions could be made possible by incorporating concepts such as the metabolite dilution [57] or molecular crowding constraints [58]. Incorporation of such additional constraints into constraint-based flux analysis could potentially improve the flux predictions and therefore yield more promising TR targets. Furthermore, the use of $\pm 10\%$ cut-off for assessing false positive *TRE* scores was arbitrary and can be subjected to scrutiny. However, with this cut-off range we observed a minimal rejection of true positive (literature validated) TR targets. Additionally, using more than one set of transcriptomic data representing the desired phenotype, i.e. producer and non-producer, to calculate the regulatory strength (*nRS*) values may increase the accuracy of TR candidate predictions. Alternatively, if no relevant datasets could be found for the desired phenotype, a general gene expression compendium can be used as it is in BeReTa.

Despite its limitations and scope for further improvements, the agreement of h-BeReTa predictions with experimental evidences from literature was substantial. Although the gene expression datasets used in this study for various case studies correspond to exponentially growing cells cultures, the method can also be readily extended to those of stationary phase cultures, provided an appropriate objective function is employed during the computation of *nRAP* scores. We believe that the less-stringent resource requirements and the computationally less-intensive methodology make h-BeReTa to be more readily employed in comparison to the existing methods for identifying non-intuitive TR targets, thereby advancing metabolic engineering applications.

Additional file

Additional file 1. Table S1. False-negative TRs for the E. coli case studies from literature evidences. **Figure S1.** Simplified toy network showing alternate routes positively correlated to product formation. **Table S2.** Randomly selected Partial (Half) gene expression datasets for acetate case study. **Table S3.** TR-hierarchy of *C. glutamicum*.

Koduru *et al. Microb Cell Fact*   (2018) 17:167

Page 11 of 12

Author details

[1] Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576, Singapore. [2] Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), 20 Biopolis Way, #06-01, Centros, Singapore 138668, Singapore. [3] School of Chemical Engineering, Sungkyunkwan University, 2066, Seobu-ro, Jangan-gu, Suwon, Gyeonggi-do 16419, Republic of Korea.

Acknowledgements

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analyzed during this study are included and/or cited appropriately in this published article and its additional file.

Consent for publication

All authors agreed to publish this article.

Ethics approval and consent to participate

Not applicable.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Keasling JD. Manufacturing molecules through metabolic engineering. Science. 2010;330:1355–8.
2. Nielsen LK. Metabolic engineering: from retrofitting to green field. Nat Chem Biol. 2011;7:408–9.
3. Lee SY, Kim HU. Systems strategies for developing industrial microbial strains. Nat Biotechnol 2015;33:1061
4. Zhang F, Ouellet M, Batth TS, Adams PD, Petzold CJ, Mukhopadhyay A, et al. Enhancing fatty acid production by the expression of the regulatory transcription factor FadR. Metab Eng. 2012;14:653–60.
5. Zhu LW, Xia ST, Wei LN, Li HM, Yuan ZP, Tang YJ. Enhancing succinic acid biosynthesis in *Escherichia coli* by engineering its global transcription factor, catabolite repressor/activator (Cra). Sci Rep. 2016;6:36526.
6. Liu Y, Xu Y, Ding D, Wen J, Zhu B, Zhang D. Genetic engineering of *Escherichia coli* to improve L-phenylalanine production. BMC Biotechnol. 2018;18:5.
7. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet. 2014;15:107–20. https://doi.org/10.1038/nrg3643.
8. Lakshmanan M, Koh G, Chung BKS, Lee D-Y. Software applications for flux balance analysis. Brief Bioinform. 2014;15:108–22. https://doi.org/10.1093/bib/bbs069.
9. Long MR, Ong WK, Reed JL. Computational methods in metabolic engineering for strain design. Curr Opin Biotechnol. 2015;34:135–41.
10. Maia P, Rocha M, Rocha I. In silico constraint-based strain optimization methods: the quest for optimal cell factories. Microbiol Mol Biol Rev. 2016;80:45–67. https://doi.org/10.1128/MMBR.00014-15.
11. Lakshmanan M, Lee N-R, Lee D-Y. Genome-scale metabolic modeling and in silico strain design of *Escherichia coli*. In: Systems biology. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA; 2017. p. 109–37.
12. Machado D, Herrgård M. Co-evolution of strain design methods based on flux balance and elementary mode analysis. Metab Eng Commun. 2015;2:85–92. https://doi.org/10.1016/j.meteno.2015.04.001.
13. Lakshmanan M, Kim TY, Chung BKS, Lee SY, Lee D-Y. Flux-sum analysis identifies metabolite targets for strain improvement. BMC Syst Biol. 2015;9:73.
14. King ZA, Feist AM. Optimizing cofactor specificity of oxidoreductase enzymes for the generation of microbial production strains—OptSwap. Ind Biotechnol. 2013;9:236–46.
15. Lakshmanan M, Chung BK-S, Liu C, Kim S-W, Lee D-Y. Cofactor modification analysis: a computational framework to identify cofactor specificity engineering targets for strain improvement. J Bioinform Comput Biol. 2013;11:1343006. https://doi.org/10.1142/S0219720013430063.
16. Kim J, Reed JL. OptORF: optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst Biol. 2010;4:53. https://doi.org/10.1186/1752-0509-4-53.
17. Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. Nature. 2004;429:92–6. https://doi.org/10.1038/nature02456.
18. Vilaça P, Rocha I, Rocha M. A computational tool for the simulation and optimization of microbial strains accounting integrated metabolic/regulatory information. BioSystems. 2011;103:435–41.
19. Kim M, Sun G, Lee DY, Kim BG. BeReTa: a systematic method for identifying target transcriptional regulators to enhance microbial production of chemicals. Bioinformatics. 2017;33:87–94.
20. Ma H-W, Buer J, Zeng A-P. Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory network revealed by a new top–down approach. BMC Bioinform. 2004;5:199.
21. Yu H, Gerstein M. Genomic analysis of the hierarchical structure of regulatory networks. Proc Natl Acad Sci USA. 2006;103:14724–31. https://doi.org/10.1073/pnas.0508637103.
22. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng. 2003;5:264–76.
23. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. Mol Syst Biol. 2014;7:535. https://doi.org/10.1038/msb.2011.65.
24. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol. 2007;3:121. https://doi.org/10.1038/msb4100155.
25. Zhang Y, Cai J, Shang X, Wang B, Liu S, Chai X, et al. A new genome-scale metabolic model of *Corynebacterium glutamicum* and its application. Biotechnol Biofuels. 2017;10:169. https://doi.org/10.1186/s13068-017-0856-3.
26. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat Protoc. 2011;6:1290–307.
27. Saa PA, Nielsen LK. Fast-SNP: a fast matrix pre-processing algorithm for efficient loopless flux optimization of metabolic models. Bioinformatics. 2016;32:3807–14. https://doi.org/10.1093/bioinformatics/btw555.
28. Gama-Castro S, Salgado H, Santos-Zavaleta A, Ledezma-Tejeida D, Muñiz-Rascado L, García-Sotelo JS, Alquicira-Hernández K, Irma Martínez-Flores I, Pannier L, et al. RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. Nucleic Acids Res. 2016;44:D133–43.
29. Ibarra-Arellano MA, Campos-González AI, Treviño-Quintanilla LG, Tauch A, Freyre-González JA. Abasy Atlas: a comprehensive inventory of systems, global network properties and systems-level elements across bacteria. Database. 2016;2016:baw089.
30. Valgepea K, Adamberg K, Nahku R, Lahtvee PJ, Arike L, Vilu R. Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. BMC Syst Biol. 2010;4:166.
31. Matsuoka Y, Shimizu K. Importance of understanding the main metabolic regulation in response to the specific pathway mutation for metabolic engineering of *Escherichia coli*. Comput Struct Biotechnol J. 2012;3:e201210018. https://doi.org/10.5936/csbj.201210018.
32. Wolfe AJ. The acetate switch. Microbiol Mol Biol Rev. 2005;69:12–50. https://doi.org/10.1128/MMBR.69.1.12-50.2005.

Koduru *et al. Microb Cell Fact* (2018) 17:167

Page 12 of 12

33. Santos CNS, Xiao W, Stephanopoulos G. Rational, combinatorial, and genomic approaches for engineering L-tyrosine production in *Escherichia coli*. Proc Natl Acad Sci. 2012;109:13538–43. https://doi.org/10.1073/pnas.1206346109.

34. Hong KK, Nielsen J. Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. Cell Mol Life Sci. 2012;69:2671–90.

35. Runguphan W, Keasling JD. Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. Metab Eng. 2014;21:103–13.

36. Tee TW, Chowdhury A, Maranas CD, Shanks JV. Systems metabolic engineering design: fatty acid production as an emerging case study. Biotechnol Bioeng. 2014;111:849–57.

37. Hernández-Almanza A, Montañez J, Martínez G, Aguilar-Jiménez A, Contreras-Esquivel JC, Aguilar CN. Lycopene: progress in microbial production. Trends Food Sci Technol. 2016;56:142–8.

38. Misawa N, Nakagawa M, Kobayashi K, Yamano S, Izawa Y, Nakamura K, et al. Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. J Bacteriol. 1990;172:6704–12.

39. Harker M, Bramley PM. Expression of prokaryotic 1-deoxy-D-xylulose-5-phosphatases in *Escherichia coli* increases carotenoid and ubiquinone biosynthesis. FEBS Lett. 1999;448:115–9.

40. Huang L, Pu Y, Yang X, Zhu X, Cai J, Xu Z. Engineering of global regulator cAMP receptor protein (CRP) in *Escherichia coli* for improved lycopene production. J Biotechnol. 2015;199:55–61.

41. Liochev SI, Benov L, Touati D, Fridovich I. Induction of the soxRS regulon of *Escherichia coli* by superoxide. J Biol Chem. 1999;274:9479–81.

42. Bongers M, Chrysanthopoulos PK, Behrendorff JBYH, Hodson MP, Vickers CE, Nielsen LK. Systems analysis of methylerythritol-phosphate pathway flux in *E. coli*: insights into the role of oxidative stress and the validity of lycopene as an isoprenoid reporter metabolite. Microb Cell Fact. 2015;14:193. https://doi.org/10.1186/s12934-015-0381-7.

43. Booth SL. Roles for vitamin K beyond coagulation. Annu Rev Nutr. 2009;29:89–110. https://doi.org/10.1146/annurev-nutr-080508-141217.

44. Bailey RL, Fulgoni VL, Keast DR, Dwyer JT. Examination of vitamin intakes among US adults by dietary supplement use. J Acad Nutr Diet. 2012;112:657–63.

45. Lubitz D, Wendisch VF. Ciprofloxacin triggered glutamate production by *Corynebacterium glutamicum*. BMC Microbiol. 2016;16:235. https://doi.org/10.1186/s12866-016-0857-6.

46. Kohl TA, Baumbach J, Jungwirth B, Pühler A, Tauch A. The GlxR regulon of the amino acid producer *Corynebacterium glutamicum*: in silico and in vitro detection of DNA binding sites of a global transcription regulator. J Biotechnol. 2008;135:340–50.

47. Kim HJ, Kim TH, Kim Y, Lee HS. Identification and characterization of glxR, a gene involved in regulation of glyoxylate bypass in *Corynebacterium glutamicum*. J Bacteriol. 2004;186:3453–60.

48. Auchter M, Cramer A, Hüser A, Rückert C, Emer D, Schwarz P, et al. RamA and RamB are global transcriptional regulators in *Corynebacterium glutamicum* and control genes for enzymes of the central metabolism. J Biotechnol. 2011;154:126–39.

49. Taniguchi H, Henke NA, Heider SAE, Wendisch VF. Overexpression of the primary sigma factor gene sigA improved carotenoid production by *Corynebacterium glutamicum*: application to production of β-carotene and the non-native linear C50 carotenoid bisanhydrobacterioruberin. Metab Eng Commun. 2017;4:1–11.

50. Brockmann-Gretza O, Kalinowski J. Global gene expression during stringent response in *Corynebacterium glutamicum* in presence and absence of the rel gene encoding (p)ppGpp synthase. BMC Genomics. 2006;7:230.

51. Gustafson CE, Kaul S, Ishiguro EE. Identification of the *Escherichia coli* lytB gene, which is involved in penicillin tolerance and control of the stringent response. J Bacteriol. 1993;175:1203–5.

52. Ginesy M, Belotserkovsky J, Enman J, Isaksson L, Rova U. Metabolic engineering of *Escherichia coli* for enhanced arginine biosynthesis. Microb Cell Fact. 2015;14:29.

53. Saier MH Jr, Ramseier TM. The catabolite repressor/activator (Cra) protein of enteric bacteria. J Bacteriol. 1996;178:3411.

54. Ogasawara H, Ishida Y, Yamada K, Yamamoto K, Ishihama A. PdhR (pyruvate dehydrogenase complex regulator) controls the respiratory electron transport system in *Escherichia coli*. J Bacteriol. 2007;189:5534–41.

55. Schwalbach MS, Keating DH, Tremaine M, Marner WD, Zhang Y, Bothfeld W, et al. Complex physiology and compound stress responses during fermentation of alkali-pretreated corn stover hydrolysate by an *Escherichia coli* ethanologen. Appl Environ Microbiol. 2012;78:3442–57.

56. Kargeti M, Venkatesh KV. Effect of global transcriptional regulators on anaerobic fermentative metabolism of *Escherichia coli*. Mol Biosyst. 2017;13:1388–98.

57. Benyamini T, Folger O, Ruppin E, Shlomi T. Flux balance analysis accounting for metabolite dilution. Genome Biol. 2010;11:R43. https://doi.org/10.1186/gb-2010-11-4-r43.

58. Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabási A-LA-L, et al. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. Proc Natl Acad Sci USA. 2007;104:12663–8. https://doi.org/10.1073/pnas.0609845104.

59. Yao R. Effect of cra gene mutation on the metabolism of *Escherichia coli* for a mixture of multiple carbon sources. Adv Biosci Biotechnol. 2013;4:477–86. https://doi.org/10.4236/abb.2013.43A063.

60. Yao R, Hirose Y, Sarkar D, Nakahigashi K, Ye Q, Shimizu K. Catabolic regulation analysis of *Escherichia coli* and its crp, mlc, mgsA, pgi and ptsG mutants. Microb Cell Fact. 2011;10:67. https://doi.org/10.1186/1475-2859-10-67.

61. Kumar R, Shimizu K. Transcriptional regulation of main metabolic pathways of cyoA, cydB, fnr, and fur gene knockout *Escherichia coli* in C-limited and N-limited aerobic continuous cultures. Microb Cell Fact. 2011;10:3.

62. Marzan LW, Shimizu K. Metabolic regulation of *Escherichia coli* and its phoB and phoR genes knockout mutants under phosphate and nitrogen limitations as well as at acidic condition. Microb Cell Fact. 2011;10:39.

63. Rodriguez A, Martnez JA, Flores N, Escalante A, Gosset G, Bolivar F. Engineering *Escherichia coli* to overproduce aromatic amino acids and derived compounds. Microb Cell Fact. 2014;13:126. https://doi.org/10.1186/s12934-014-0126-z.

64. Rungrassamee W, Liu X, Pomposiello PJ. Activation of glucose transport under oxidative stress in *Escherichia coli*. Arch Microbiol. 2008;190:41–9.

65. Peng L, Shimizu K. Effect of fadR gene knockout on the metabolism of *Escherichia coli* based on analyses of protein expressions, enzyme activities and intracellular metabolite concentrations. Enzyme Microb Technol. 2006;38:512–20.

66. Scheel RA. Deregulation of fatty acid transport in *Escherichia coli* for enhanced control of biodegradable plastic copolymer production. Honors Theses. State University of New York, College of Environmental Science and Forestry; 2014.

67. Pauli G, Ehring R, Overath P. Fatty acid degradation in *Escherichia coli*: requirement of cyclic adenosine monophosphate and cyclic adenosine monophosphate receptor protein for enzyme synthesis. J Bacteriol. 1974;117:1178–83.

68. Feng Y, Cronan JE. Crosstalk of *Escherichia coli* FadR with global regulators in expression of fatty acid transport genes. PLoS ONE. 2012;7:e46275.

69. Nikel PI, Pettinari MJ, Ramírez MC, Galvagno MA, Méndez BS. *Escherichia coli* arcA mutants: metabolic profile characterization of microaerobic cultures using glycerol as a carbon source. J Mol Microbiol Biotechnol. 2008;15:48–54.

70. Kim Y, Ingram LO, Shanmugam KT. Construction of an *Escherichia coli* K-12 mutant for homoethanologenic fermentation of glucose or xylose without foreign genes. Appl Environ Microbiol. 2007;73:1766–71.

71. Orencio-Trejo M, Flores N, Escalante A, Hernández-Chávez G, Bolívar F, Gosset G, et al. Metabolic regulation analysis of an ethanologenic *Escherichia coli* strain based on RT-PCR and enzymatic activities. Biotechnol Biofuels. 2008;1:8.

72. Hwang JH, Hwang GH, Cho JY. Effect of increased glutamate availability on L-ornithine production in *Corynebacterium glutamicum*. J Microbiol Biotechnol. 2008;18:704–10.

73. Supkulsutra T, Maeda T, Kumagai K, Wachi M. A role of the transcriptional regulator LldR (NCgl2814) in glutamate metabolism under biotin-limited conditions in *Corynebacterium glutamicum*. J Gen Appl Microbiol. 2013;59:207–14. https://doi.org/10.2323/jgam.59.207.