J Ind Microbiol Biotechnol (2015) 42:1401–1414
DOI 10.1007/s10295-015-1663-0

SYSTEMS BIOTECHNOLOGY

In silico model-driven cofactor engineering strategies for improving the overall NADP(H) turnover in microbial cell factories

Meiyappan Lakshmanan[1] · Kai Yu[2,3] · Lokanand Koduru[2] · Dong-Yup Lee[1,2,3]

Received: 23 February 2015 / Accepted: 23 July 2015 / Published online: 8 August 2015

Abstract Optimizing the overall NADPH turnover is one of the key challenges in various value-added biochemical syntheses. In this work, we first analyzed the NADPH regeneration potentials of common cell factories, including *Escherichia coli*, *Saccharomyces cerevisiae*, *Bacillus subtilis*, and *Pichia pastoris* across multiple environmental conditions and determined *E. coli* and glycerol as the best microbial chassis and most suitable carbon source, respectively. In addition, we identified optimal cofactor specificity engineering (CSE) enzyme targets, whose cofactors when switched from NAD(H) to NADP(H) improve the overall NADP(H) turnover. Among several enzyme targets, glyceraldehyde-3-phosphate dehydrogenase was recognized as a global candidate since its CSE improved the NADP(H) regeneration under most of the conditions examined. Finally, by analyzing the protein structures of all CSE enzyme targets via homology modeling, we established that the replacement of conserved glutamate or aspartate with serine in the loop region could change the cofactor dependence from NAD(H) to NADP(H).

Keywords NADPH · Metabolic engineering · Cofactor specificity engineering (CSE) · Flux balance analysis (FBA) · Cofactor modification analysis (CMA)

Introduction

The use of microbes for the production of chiral pharmaceutical intermediates and several other plant-based natural products is a promising alternative to the traditional chemical synthesis routes. Such compounds are usually complex molecules with multiple stereocenters and chiral functional groups, requiring extensive amounts of NADP(H) as a specific cofactor for their biosynthesis via highly specific oxidoreductase family of enzymes, such as the cytochrome P450s [6, 48]. Therefore, it is a prerequisite for the microbial cell factories to sufficiently regenerate the required NADPH to achieve high yields and large titers of such complex molecules. However, it is well known that most of the commonly used industrial microbes, such as *Escherichia coli* and *Bacillus subtilis* naturally possess low NADP(H) turnover rates when compared to NAD(H) [14]. This is mainly due to the type of reactions they are involved in cellular metabolism: NAD(H) is generally associated with the high flux pathways, i.e., glycolysis, fermentation, and oxidative phosphorylation whereas NADP(H) is linked with moderately low flux routes, such as pentose phosphate pathway (PPP), transdehydrogenase enzymes, and amino acid and nucleotide biosynthetic pathways. As such, it is essential for the microbial expression hosts to possess high NADP(H) turnover in the metabolic backbone for the efficient synthesis of these natural products.

Electronic supplementary material The online version of this article (doi:10.1007/s10295-015-1663-0) contains supplementary material, which is available to authorized users.

✉ Dong-Yup Lee
cheld@nus.edu.sg

[1] Bioprocessing Technology Institute, Agency for Science, Technology and Research (A*STAR), 20 Biopolis Way, #06-01, Centros, Singapore 138668, Singapore

[2] Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117585, Singapore

[3] NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), Life Sciences Institute, National University of Singapore, 28 Medical Drive, Singapore 117456, Singapore

Recognizing the importance of amplifying the overall NADPH bioavailability, several techniques has been proposed for the same [25]. In one of the earlier and simpler work, San et al. used different carbon substrates to utilize diverse metabolic pathways, and thus manipulating intracellular cofactor levels [42]. Later, specific genetic engineering strategies were proposed to forcefully re-direct the metabolic fluxes via desired pathways to improve the required cofactor regeneration. For example, overexpression of glucose-6-phosphate dehydrogenase (*zwf*) and/or gluconate-P dehydrogenase (*gnd*) [24, 30, 44] and deletion of phosphoglucose isomerase (*pgi*) [1, 5] and phosphofructokinase (*pfk*) in *Escherichia coli* increased the carbon flux through PPP, and thus augmented the overall NADP(H) turnover. Other strategies, such as overexpression of NAD kinase [29] and the endogenous nucleotide transhydrogenase (*PntA*/*UdhA*) [2, 22] enzymes, are also proposed to improve the intracellular NADP(H) levels at the expense of NAD(H). As an alternative, engineering the cofactor specificity of a particular enzyme from NAD(H) to NADP(H) was also shown to increase the NADP(H) production rates at the expense of NAD(H) [33].

In silico model-driven approaches have been increasingly employed to guide metabolic engineering in the past decade [23, 26, 34]. Notably, most of these studies utilized the simple and extensible steady-state modeling framework, constraint-based modeling (CBM), which can simulate the cellular phenotype in terms of metabolic fluxes with just the information of reaction stoichiometry and mass-balance from the underlying metabolic models [3, 28]. As a result, genome-scale metabolic models (GEMs) have now been reconstructed for more than 100 organisms across all domains of life [36]. Furthermore, the availability of various conveniently accessible CBM software applications along with the GEMs has greatly accelerated the in silico phenotype prediction and rational design of cellular systems [21]. In cofactor regards, Chin et al. first utilized the CBM approach to investigate the possible metabolic sources and their contributions toward overall NADP(H) synthesis in various xylitol overproducing mutants of *E. coli* [7]. Later, Ahn et al. analyzed the differences in the NADP(H) turnover rates between the wild-type *E. coli* and its *pgi*-mutant to account for the enhanced shikimate production [1]. Further, Ghosh et al. explored the effect of altering the cofactor specificity of various enzymes among the non-native xylose catabolic pathways of *Saccharomyces cerevisiae* upon its cellular growth and ethanol production [15]. Following such initial successful applications of CBM to analyze redox balancing, novel in silico methods were also proposed to improve the regeneration of desired cofactor. Cipher for Evolutionary Design was the first algorithm proposed in this manner to unravel valid gene deletion targets which can improve the intracellular NADP(H) levels [5]. Two other in silico methods, cofactor modification analysis (CMA) [11, 20] and OptSwap [18, 19], were also proposed to identify relevant cofactor specificity engineering (CSE) targets which can augment the yield of several native and non-native products in *E. coli* and *S. cerevisiae* by improving the overall redox balance. Although these model-driven studies provided valuable insights regarding intracellular cofactor regeneration, still a comprehensive in silico analysis of how metabolic pathway organization can influence redox balance in different microbes is not available yet. Accessibility to such information will help experimentalists to rationally select the preferred microbial chassis and design the culture media appropriately for the NADPH–dependent product of interest.

In this work, we first utilize the previously proposed metabolite-centric approach, known as flux-sum analysis [9], to quantify the overall NADPH turnover rates in four of the commonly used industrial microbial hosts, *E. coli*, *S. cerevisiae*, *B. subtilis*, and *Pichia pastoris*, under aerobic and anaerobic conditions using the corresponding constraint-based GEMs. Subsequently, we employ the CMA to identify the optimal enzyme targets in each of the organism whose cofactor engineering can further improve the overall NADP(H) turnover. Finally, we identified the cofactor binding sites in these target enzymes by analyzing protein sequence and structural data, and propose mutation strategies that may potentially switch their cofactor specificity from NADH to NADPH.

Methods

Constraints-based flux analysis

In this study, we utilized constraints-based flux analysis, also known as flux balance analysis, for computing the maximal cellular growth under varying environmental conditions [38]. The biomass reaction was maximized to quantify the internal flux distribution as described elsewhere [13]. Mathematically, the constraints-based flux analysis problem for biomass flux maximization subjected to stoichiometric and capacity constraints can be represented as

$$\begin{aligned} &\max\ v_{\text{biomass}} \\ &\text{s.t.} \sum_j S_{ij} v_j = 0 \quad \forall \text{ metabolite } i \\ &v_j^{\min} \le v_j \le v_j^{\max} \quad \forall \text{ reaction } j \end{aligned} \tag{1}$$

where S_{ij} is the stoichiometric matrix, v_j is the flux through reaction j, and, $v_j^{\min}$ and $v_j^{\max}$ are the possible lower and upper limits of flux v_j, respectively.

Flux variability analysis

Since constraints-based flux analysis is an optimization-based technique, it is often possible to have multiple flux values for the same objective. Therefore, we performed the flux variability analysis (FVA) [32] to identify all active fluxes and their possible ranges during maximal cell growth. The mathematical formulation of FVA problem can be represented using the same constraints of constraint-based flux analysis as follows:

$$\begin{aligned}
&\max/\min\ v_j \\
&\text{s.t.} \sum_j S_{ij} v_j = 0 \\
&\sum_j c_j v_j = Z_{obj} \\
&v_j^{\min} \le v_j \le v_j^{\max} \quad \text{for } j = 1, \ldots, n
\end{aligned} \tag{2}$$

where Z_{obj} denotes the value of objective calculated from problem (1) and n is the total number of fluxes. The upper range of fluxes is identified by maximizing the objective whereas the lower range is obtained by minimizing the same.

Flux-sum analysis

Since the constraint-based flux analysis simply indicates the rates of consumption/generation of cofactors, we utilized the previously developed concept of "flux-sum" (ϕ_i) to quantify the overall NADP(H) and NAD(H) turnover rates in the metabolic network [9, 17]. Under the steady-state assumption, the generation and consumption of any metabolite will be equal. Therefore, the flux-sum of metabolite i can be formulated as

$$\phi_i = 0.5 \sum_j \left| S_{ij} v_j \right|, \tag{3}$$

where the actual turnover rate of metabolite i is the half of absolute sum of consumption and generation rates. In this study, we compute both the basal cofactor flux-sum, i.e., the possible turnover rates of NADH and NADPH during maximal cellular growth, and the flux-sum maxima of NADPH, i.e., the maximum achievable flux-sum from the metabolic network.

Basal flux-sum

The basal flux-sum of cofactors, NADPH and NADH, was calculated from the "wild-type" flux distributions which are determined by solving the constraint-based flux analysis and FVA problems. Briefly, the procedure to compute the range of cofactor basal flux-sums is as follows. First, the problem (1) is solved to compute the maximal cell growth. Subsequently, problem (2) is solved to evaluate the maximal and minimal values of each flux in the model while the cells achieve the maximal cellular growth which is calculated from (1). Finally, the lower and upper range of cofactor flux-sums are calculated by substituting the solution obtained from problem (2) onto Eq. 3.

Flux-sum maxima

As mentioned earlier, the flux-sum of a NADPH simply provides its turnover rates in the metabolic network while the cell evolves toward a particular objective, typically growth maximization [9]. Besides basal flux-sum, another important metric to assess the NADPH regeneration ability of each microbe is the flux-sum maxima, i.e., maximum possible NADPH turnover rates. It should be noted that although cells may never achieve this level of NADPH during cell culture, its evaluation will provide us more insights about the flexibility of the metabolic network to synthesize NADPH. It was calculated using the previously proposed mixed-integer optimization (MIP) problem [9] which can be mathematically represented as follows:

$$\begin{aligned}
&\max \phi_{\text{NADPH}} = 0.5 \sum_j \left(f^+_{\text{NADPH},j} + f^-_{\text{NADPH},j} \right) \\
&\text{s.t.} \sum_j S_{ij} v_j = 0 \quad \forall \text{ metabolite } i \\
&S_{ij} v_j = f^+_{ij} - f^-_{ij} \\
&f^+_{ij} \le 1000\, I^+_{ij} \\
&f^-_{ij} \le 1000\, I^-_{ij} \\
&I^+_{ij} + I^-_{ij} = 1 \\
&v_j^{\min} \le v_j \le v_j^{\max} \quad \forall \text{ reaction } j \\
&I^+_{ij} \in \{0,1\},\ I^-_{ij} \in \{0,1\}, f^+_{ij} \ge 0, f^-_{ij} \ge 0.
\end{aligned}$$

The flux-sum term is nonlinear due to the modulus operator (Eq. 3). Therefore, its direct imposition into any optimization formulation will result in a nonlinear problem. In order to avoid such instance, we introduced additional constraints to represent it in linear integer form as proposed previously [9]. Accordingly, f^+_{ij} and f^-_{ij}, are the two new positive variables which refer to the generation and consumption components of metabolite i due to reaction j, respectively. I^+_{ij} and I^-_{ij} are the two binary variables which serve as switches to turn the generation and consumption components on and off such that only one of the components is active, effected by the $I^+_{ij} + I^-_{ij} = 1$ constraint.

Cofactor modification analysis (CMA)

To identify the optimal CSE candidate which can improve the overall NADP(H) regeneration, we implemented CMA

as described in our previous work [20], except, the outer objective as ϕ_{NADPH} instead of v_{product}. Mathematically, the bi-level MINLP optimization problem specific to the CMA can be represented as follows:

$$\max \phi_{\mathrm{NADPH}} = 0.5 \sum_j \left| S_{\mathrm{NADPH},j} v_j \right|$$

$$\text{s.t.} \begin{bmatrix} \max v_{\mathrm{biomass}} \\ \text{s.t.} \sum_j \left(S_{ij} v_j + S_{ij}^{\mathrm{cMod}} v_j^{\mathrm{cMod}} \right) = 0 \,\forall \text{ metabolite } i \\ v_{\mathrm{biomass}} \geq v_{\mathrm{biomass}}^{\min} \\ \left(1 - y_j^{\mathrm{cMod}}\right) \cdot v_j^{\min} \leq v_j \leq \left(1 - y_j^{\mathrm{cMod}}\right) \cdot v_j^{\max} \\ y_j^{\mathrm{cMod}} \cdot v_j^{\min} \leq v_j^{\mathrm{cMod}} \leq y_j^{\mathrm{cMod}} \cdot v_j^{\max} \\ y_j^{\mathrm{cMod}} = \{0, 1\} \,\forall \text{ reaction } j \end{bmatrix},$$

$$\sum_j y_j^{\mathrm{cMod}} \leq k$$

where S_{ij}^{cMod} is the cofactor modified stoichiometric matrix where the coefficients are same as S_{ij}, except the reactions which involve either NAD(H) or NADP(H). These reactions are swapped for cofactors in the S_{ij}^{cMod} matrix such that $S_{\mathrm{NAD(H)},j} = S_{\mathrm{NADP(H)},j}^{\mathrm{cMod}}$ and $S_{\mathrm{NADP(H)},j} = S_{\mathrm{NAD(H)},j}^{\mathrm{cMod}} \cdot v_j^{\mathrm{cMod}}$ is the flux through the cofactor-modified reaction and $v_{\mathrm{biomass}}^{\min}$ is the minimum amount of biomass that needs to be produced. The binary variable y_j^{cMod} ensures that the cofactor-associated reactions are allowed to carry flux either with its original or swapped cofactor but not both. The number of cofactor switches allowed in a particular simulation is controlled by the number k which is fixed at 1 for all simulations in this work. The bi-level MINLP problem was reformulated as a single-level MINLP problem using the primal dual transformation as described previously [20].

Genome-scale models and in silico simulation settings

The *i*JR904 [40], *i*YO844 [37], *i*MM904 [35], and *i*PP668 [10] metabolic models were used for analyzing the cofactor balancing in *E. coli*, *B. subtilis*, *S. cerevisiae,* and *P. pastoris*, respectively. First, the type III pathways [39] in all models were identified and one reaction in each of the loops was constrained to zero for eliminating the internal cycles during flux analysis computations. For all simulations, the carbon uptake rate was fixed at 100 C-mmol g^{-1} DCW h^{-1}. The oxygen uptake rate was set to either zero or 20 mmol g^{-1} DCW h^{-1}, corresponding to anaerobic and aerobic conditions, respectively. It should be noted that although some organisms, e.g., *S. cerevisiae* may have much lower oxygen uptake rates than 20 mmol g^{-1} DCW h^{-1} [12], we used the same to make a fair comparison with others. The exchange fluxes of essential nutrients, such as NH_3, phosphate, sulfite, H_2O, Fe^{2+}, Mg, and H^+ were left unconstrained during simulations based on the default settings of corresponding GEM (see Table S1–S4 in the Supporting Information for a list of unconstrained exchange reactions in each model). The minimum level of target biomass in CMA simulations was set at the value obtained during wild-type flux analysis simulations. Note that all simulations were performed in GAMS IDE software where the LP problems corresponding to constraint-based flux analysis and MIP problems related to flux-sum analysis were solved using CPLEX solver, and the MINLP problems specific to CMA were solved using LINDO solver.

Analysis of cofactor-binding sites in target enzymes

The protein sequences for the target enzymes were retrieved from the Uniprot Database (http://www.uniprot.org/). For enzymes with experimentally determined structures, the structural information was retrieved from the Protein Data Bank (PDB) and their NADH-binding sites were readily identified in the structures. For enzymes without available structural data, 3D structural models were built by using their protein sequences as queries for homology modeling. HHpred [46] was used to identify structural templates from PDB based on profile–profile alignments in the algorithm. A structural model was then built by the MODELLER program that is integrated into the HHpred server. The NADH cofactors were manually positioned in the models in a similar geometry and orientation observed in the binding sites of respective templates or their close homologs. Pairwise and multiple sequence alignments were performed by using Clustal Omega [45] with default parameters and observed for conserved sequences. The loop region in each enzyme that determines the cofactor preference was identified from the analysis of both sequence and structural information.

Results

NADP(H) turnover capacity of wild-type strains

The intracellular cofactor pools are known to be greatly influenced by the carbon nutrient sources and the oxygenation conditions of the culture medium [41]. Therefore, we first quantified the basal flux-sums of NADP(H) and NAD(H), i.e., the possible turnover rate ranges in wild type, in all the four organisms while growing on three of the most frequently used carbon sources for industrial fermentation, i.e., glucose, xylose, and glycerol, under both aerobic and anaerobic conditions. Figure 1 shows the basal NAD(H) and NADP(H) flux-sums and the corresponding in silico growth rates of all the microbes examined.

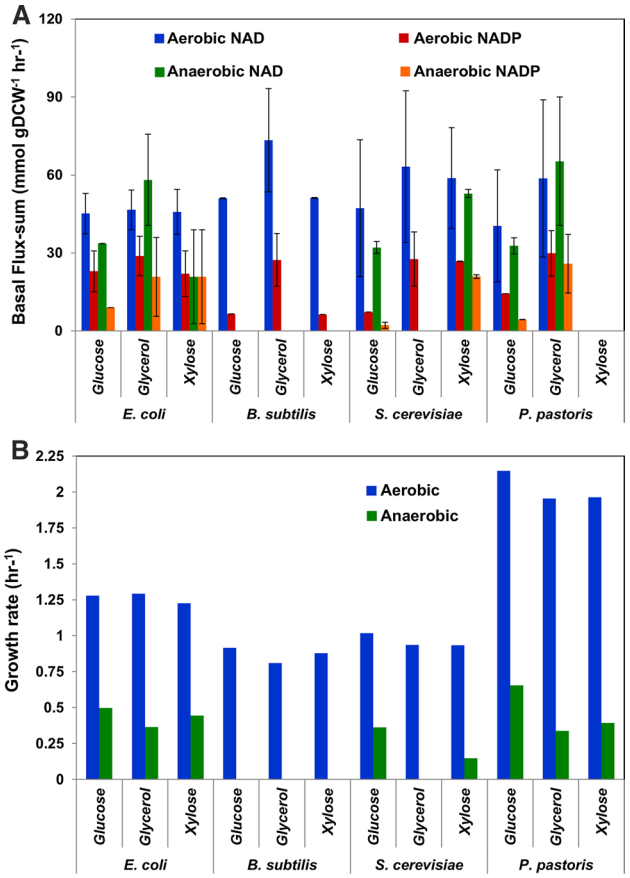


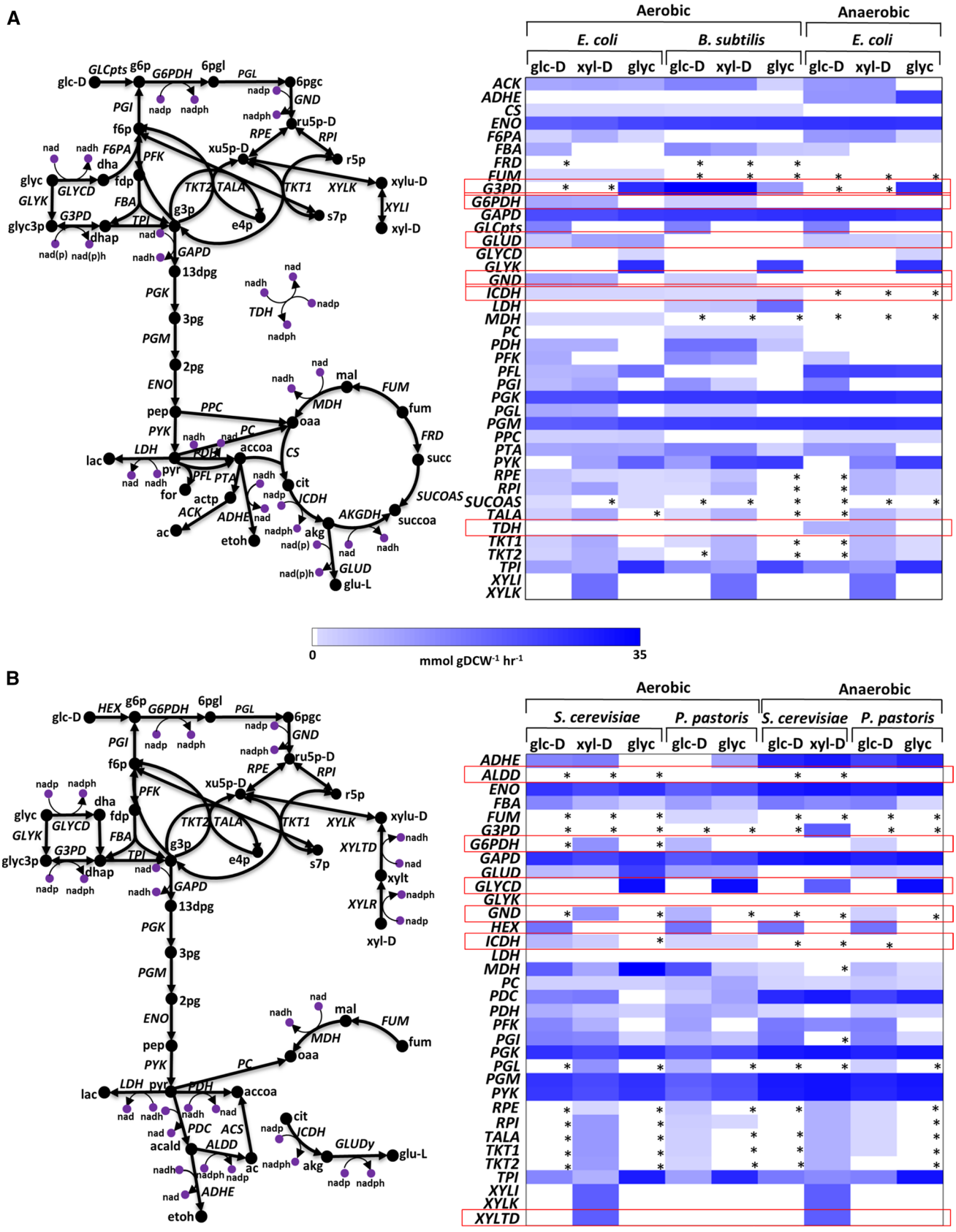
Fig. 1 Cofactor turnover rates (**a**) and growth rate (**b**) of the four microbial chassis under aerobic and anaerobic conditions. It should be noted that some organisms do not grow in certain conditions and thus the flux-sums cannot be computed. Following is the list of non-growth conditions for each organism: *B. subtilis*—anaerobic, *S. cerevisiae*—glycerol (anaerobic), *P. pastoris*—glycerol (both aerobic and anaerobic)

From Fig. 1a, it can be observed that the NADP(H) turnover is generally lower than that of NAD(H) in all organisms under both aerobic and anaerobic conditions. Importantly, the NAD(H) regeneration rate is about two to five times than that of NADP(H) in aerobic conditions and as high as 20 in anaerobic conditions. It should be noted that these observations are in good agreement with the previous experimental study which also reported that NAD^+ pools are always higher than that of $NADP^+$ in several bacteria [14]. A simple reason for observing such high NAD(H)/NADP(H) ratios, especially under anaerobic conditions, is the differences in pathway utilization and the flux flow through them. Typically, the carbon flux is channeled through the fermentative pathways in the absence of oxidative phosphorylation, and thus the carbon available through the primary NADP(H) synthetic routes, pentose phosphate pathway (PPP), and tricarboxylic acid (TCA) cycle, will be very low under such conditions (Fig. 2). These observations highlight the importance of maintaining aerobic conditions to achieve a high intracellular NADP(H) turnover, in general.

Apart from oxygenation, the use of carbon sources also impacted the overall NADP(H) turnover rates, especially in yeasts, *S. cerevisiae* and *P. pastoris* (Fig. 1a). Moreover, *B. subtilis* and *S. cerevisiae* showed remarkably low NADPH turnover when compared to *E. coli* and *P. pastoris* (Fig. 1a). In order to get a better insight into such diverse observations, we examined the flux distribution through central metabolic pathways and assessed how they contribute to the differences in flux-sum. In general, six key enzymes, glycerol-3-phosphate dehydrogenase (G3PD), glucose-6-phosphate dehydrogenase (G6PDH), gluconate dehydrogenase (GND), glutamate dehydrogenase (GLUD), and isocitrate dehydrogenase (ICDH) are known to contribute for NADPH in bacteria. *B. subtilis* has a much lower NADPH flux-sum than *E. coli* possibly due to its low flux through the PPP enzymes, G6PDH and GND, and/or the different specificity of the G3PD and GLUD enzymes that are NAD^+ specific in *B. subtilis* whereas they are $NADP^+$ specific in *E. coli* (Fig. 2a). Notably, *S. cerevisiae* also has very poor NADPH basal flux-sum while growing on glucose (two-folds less than that of *P. pastoris*). A closer examination of internal flux distribution revealed that *S. cerevisiae* has remarkably low flux through the pentose phosphate pathway when compared to *P. pastoris* (~25 times lower) and channels most of the carbon flux into the fermentative pathways due to its unique respiro-fermentative metabolism [8]. However, both yeasts showed a very high NADPH flux-sum while growing on glycerol due to the use of the exact same metabolic route: NADP-dependent glycerol dehydrogenase (GLYCD) enzyme in the first few initial steps, where the carbon flux is usually be very high (Fig. 2b). Similarly, *E. coli* also had slightly higher NADPH flux-sums while growing on glycerol as it also utilizes the NADP-dependent glycerol-3-phosphate dehydrogenase (G3PD) for the breakdown of glycerol. Collectively, these observations highlight glycerol as the ideal carbon source for achieving high NADP(H) regeneration due to the presence of several NADP-dependent enzymes, particularly in the upstream, among its catabolic pathways.

Maximum NADP(H) turnover capacity and its tradeoffs with cellular growth

In the previous section, we computed the basal NADPH flux-sum of all microbes, which just provides the range of achievable cofactor turnover rates while the cell evolves toward maximal growth [9]. Since constraint-based flux analysis is an optimization problem, the flux-sum computed from such analysis will be a function of the biomass components, and thus the NADPH balancing could be highly


A
Aerobic
Anaerobic
E. coli
B. subtilis
E. coli
glc-D xyl-D glyc glc-D xyl-D glyc glc-D xyl-D glyc
ACK ADHE CS ENO F6PA FBA FRD FUM G3PD G6PDH GAPD GLCpts GLUD GLYCD GLYK GND ICDH LDH MDH PC PDH PFK PFL PGI PGK PGL PGM PPC PTA PYK RPE RPI SUCOAS TALA TDH TKT1 TKT2 TPI XYLI XYLK
0
mmol gDCW-1 hr-1
35
B
Aerobic
Anaerobic
S. cerevisiae
P. pastoris
S. cerevisiae
P. pastoris
glc-D xyl-D glyc glc-D glyc glc-D xyl-D glc-D glyc
ADHE ALDD ENO FBA FUM G3PD G6PDH GAPD GLUD GLYCD GLYK GND HEX ICDH LDH MDH PC PDC PDH PFK PGI PGK PGL PGM PYK RPE RPI TALA TKT1 TKT2 TPI XYLI XYLK XYLTD

◀ **Fig. 2** Flux distribution across the central metabolic enzymes in bacteria (**a**) and yeasts (**b**). The reactions in red boxes correspond to enzymes which has NADP(H) as cofactor in one or more organisms and asterisks represent enzymes with very low, non-zero flux value. Enzyme abbreviations are as follows: *ACK* acetate kinase, *ADHE* alcohol dehydrogenase, *ALDD* aldehyde dehydrogenase, *CS* citrate synthase, *ENO* enolase, *F6PA* fructose-6-phosphate aldolase, *FBA* fructose-bisphosphate aldolase, *FRD* fumarate reductase, *FUM* fumarase, *G3PD* glycerol-3-phosphate dehydrogenase, *G6PDH* glucose-6-phosphate dehydrogenase, *GAPD* glyceraldehyde-3-phosphate dehydrogenase, *GLCpts* glucose transport via PEP:Pyr PTS, *GLUD* glutamate dehydrogenase, *GLYCD* glycerol dehydrogenase, *GLYK* glycerol kinase, *GND* gluconate dehydrogenase, *HEX* hexokinase, *ICDH* isocitrate dehydrogenase, *LDH* lactate dehydrogenase, *MDH* malate dehydrogenase, *PC* pyruvate carboxylase, *PDC* pyruvate decarboxylase, *PDH* pyruvate dehydrogenase, *PFK* phosphofructokinase, *PFL* pyruvate-formate lyase, *PGI* glucose-6-phosphate isomerase, *PGK* phosphoglycerate kinase, *PGL* 6-phosphogluconolactonase, *PGM* phosphoglycerate mutase, *PPC* phosphoenolpyruvate carboxylase, *PTA* phosphotransacetylase, *PYK* pyruvate kinase, *RPE* ribulose-5-phosphate epimerase, *RPI* ribose-5-phosphate isomerase, *SUCOAS* succinyl-CoA synthase, *TALA* transaldolase, *TDH* transdehydrogenase, *TKT* transketolase, *TPI* triose phosphate isomerase, *XYLI* xylose isomerase, *XYLK* xylulokinase, *XYLR* xylose reductase, *XYLTD* xylitol dehydrogenase

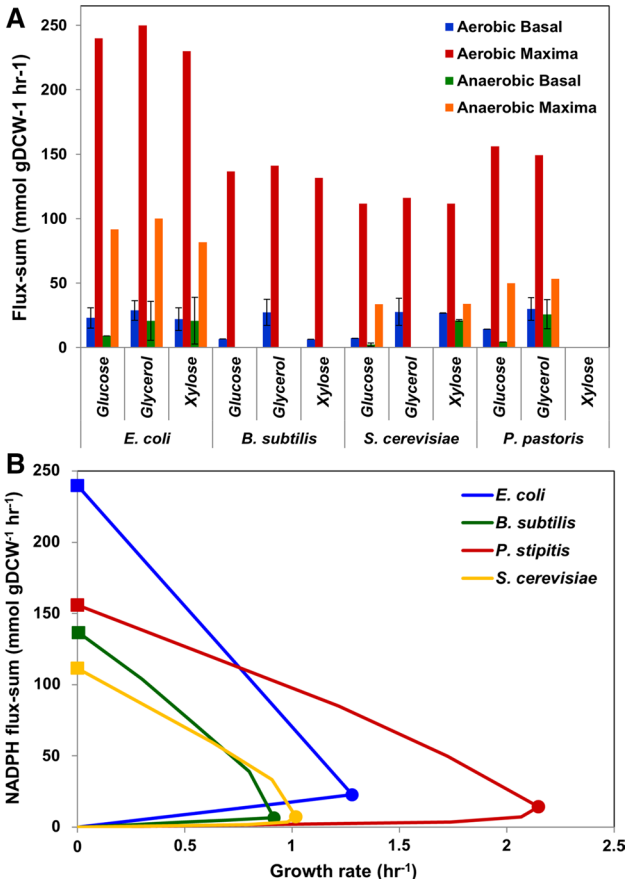


Fig. 3 **a** Basal and maximal NADPH flux-sums of the four microbes. It should be noted that since some organisms do not grow in certain conditions, we did not compute the flux-sum maxima for those conditions and they do not appear. **b** Relationship between NADPH flux-sum and growth rate under aerobic conditions while growing in glucose. Basal and maximal flux-sums are indicated with *closed circles* and *squares,* respectively

dependent on the corresponding biosynthetic pathways. Therefore, in order to evaluate the maximum theoretical flux-sum from the metabolic network without considering any cellular objective, we computed its flux-sum maxima [9]. In all the cases, the NADPH flux-sum maxima is much higher than the basal flux-sum (more than 5 times), indicating that these are "partially utilized" metabolites which are not turned over at their full capacity during maximum cell growth. This evaluation characterizes the metabolic network flexibility toward NADPH synthesis, and thus, allows us to design more targeted approaches, such as gene deletion/amplification, to improve the NADPH biosynthesis of a particular microbe which has substantially low levels than its maximum limit while growing in the cell culture.

Notably, *E. coli* has very high NADPH flux-sum maxima despite having comparable basal flux-sums with other organisms, suggesting that it probably has the most flexible central metabolism for NADPH synthesis (Fig. 3a). Further examination of corresponding intracellular fluxes revealed that it used the membrane-bound transdehydrogenase (TDH) enzyme to produce the whole of NADPH. It should be noted that this TDH is unique to *E. coli* and is not present in any of the other three organisms examined. Moreover, this enzyme is reported to provide about 40 % of total NADPH in *E. coli* while growing in glucose batch cultures [43]. Collectively, such observations suggest that TDH has a critical role in overall NADPH regeneration ability in *E. coli* where the turnover rates can be further augmented by overexpressing the PPP enzymes, such as G6PDH and GND. Unlike *E. coli*, all other microbes had relatively lower flux-sum maxima and rely heavily on the oxidative PPP fluxes for their NADPH production. Here, it is interesting to note that *B. subtilis* has a comparable flux-sum maxima with that of *S. cerevisiae* and *P. pastoris* despite having substantially low basal flux-sum. Such observations highlight the fact that although *B. subtilis* has superior ability to regenerate NADP(H), probably it synthesizes NADPH at sub-optimal ranges while growing exponentially to maintain the overall redox balance and channel maximum possible carbon flux into the biomass synthetic pathways.

We also analyzed the effect of NADPH turnover on cellular growth from the lowest value, i.e., zero, to the maximum value, i.e., maxima. Figure 3 shows the relationship between NADPH flux-sum and biomass production of all the four microbes under aerobic and anaerobic conditions. It could be observed that when NADPH synthesis is completely shut down from the network, the cells were unable to grow, indicating that it is an "essential metabolite" [17].

Table 1 CMA identified CSE targets and the % increase of NADP(H) flux-sum and growth in the mutant strains with respect to the wild type

Organism	Substrate	Aerobic				Anaerobic			
		CSE Target[a]	NADPH flux-sum (% increase)		Growth rate (% increase)	CSE Target	NADPH flux-sum (% increase)		Growth rate (% increase)
			Basal	Maxima			Basal	Maxima	
E. coli	Glucose	GAPD	41.20	13.88	5.73	GAPD	263.80	36.36	9.58
		PDH	5.51	11.11	5.51	PDH	8.19	27.27	6.97
	Glycerol	GAPD	172.26	13.33	0	GAPD	96.84	33.21	0
		PGCD	54.84	0	0	GLYCD	6.82	0	0
	Xylose	GAPD	47.57	14.49	6.17	GAPD	309.02	40.81	9.58
		PDH	7.44	11.59	6.16	PDH	8.19	30.61	6.97
B. subtilis	Glucose	GAPD	384.45	19.50	2.94	NA	NA	NA	NA
		PDH	249.57	5.36	2.94	NA	NA	NA	NA
	Glycerol	PDH	573.09	22.83	0	NA	NA	NA	NA
		PGCD	504.53	0	0	NA	NA	NA	NA
	Xylose	GAPD	405.98	20.25	2.94	NA	NA	NA	NA
		PDH	262.36	5.57	2.94	NA	NA	NA	NA
S. cerevisiae	Glucose	GAPD	356.69	23.86	0.22	ADHE	557.25	0	8.11
		MDH	266.65	8.15	0.22	GAPD	329.28	98.19	8.11
	Glycerol	GAPD	132.90	27.75	0	NA	NA	NA	NA
		MDH	42.06	6.58	0	NA	NA	NA	NA
	Xylose	GAPD	6.31	23.88	9.04	GAPD	16.65	98.19	16.66
		XYLTD	2.25	17.91	8.83	–[b]	–	–	–
P. pastoris	Glucose	GAPD	71.03	4.71	5.11	GAPD	598.50	50	7.90
		MDH	11.16	12.7	5.11	ADHE	568.57	0	7.90
	Glycerol	GAPD	69.74	9.37	0	GAPD	1180.15	56.13	0
		PGCD	3.15	0	0	PGCD	561.73	0.02	0
	Xylose	NA	NA	NA	NA	NA	NA	NA	NA
		NA	NA	NA	NA	NA	NA	NA	NA

[a] Enzyme abbreviations are same as Fig. 2

[b] Apart from GAPD, CSE of none other enzymes improved NADPH turnover

Moreover, when the flux-sum is intensified to full capacity from the basal value, our simulations again resulted in zero cell growth, demonstrating that it is a "competitive metabolite" as well [9]. This implies that any targeted channeling of carbon flux toward NADPH synthetic pathways will deprive the cell from producing some of its essential metabolites, thus ceasing its growth. Therefore, it is necessary to maintain the regeneration rate at a sub-optimal level which improves the product synthesis significantly but at the same time does not penalize cell growth severely.

Cofactor engineering enzyme targets to improve NADP(H) regeneration

We next applied the CMA to identify the optimal enzyme targets for CSE to improve NADPH regeneration as it has been earlier shown that cofactor engineering increases product yields without compensating biomass production [20]. Table 1 summarizes the CMA identified top two CSE targets for each of them (see Table S5 in the Supporting Information for more targets). It can be observed along with NADPH flux-sums, as expected, CSE also improved the biomass yield significantly in several cases.

Global CSE target: glyceraldehyde-3-phosphate dehydrogenase

Glyceraldehyde-3-phosphate dehydrogenase (GAPD) was identified as the best enzyme target for CSE to improve the overall NADP(H) in several cases, especially while growing in glucose and xylose. Swapping of GAPD's cofactor almost reverses the NAD(H)/NADP(H) ratio in several cells (results not shown). Since GAPD is a central enzyme in glycolysis, it has an assured carbon flow through it while growing on glucose, glycerol, or xylose and contributes largely for high NAD(H) turnover rates in the wild-type

strains. Therefore, it is expected that the engineering of GAPD's cofactor from NAD(H) to NADP(H), allows the cells to generate large amounts of NADP(H) instead of NAD(H) through glycolysis itself. Here, it should be noted that this strategy has already been practically shown to work effectively in *E. coli* by Martínez et al. who replaced the native GAPD with NADP-specific GAPD from *Clostridium acetobutylicum* [33]. The authors observed a high production rates of the NADP-dependent products without an increase through PPP pathway, suggesting that the NADP(H) required for product synthesis is indeed produced from glycolysis itself. Similarly, a *S. cerevisiae* strain was also constructed with NADP-dependent GAPD to improve the xylose fermentation [47]. Further, the same strain has been patented, reporting that it has significantly higher yield for several products, such as polyhydroxyalkanoates, amino acids, fats, vitamins, and nucleotides [31].

It should be emphasized that among all the four organisms analyzed, *B. subtilis* natively has a NADP-dependent GAPD in addition to the NAD-specific GAPD. However, we did not observe any high NADP(H) turnover in wild-type *B. subtilis* because only the NAD-dependent GAPD operates in the glycolytic direction whereas the NADP-dependent GAPD acts on reverse direction as it is captured by the *i*YO844 model. Therefore, it is important to manipulate the regulatory mechanisms which prevent NADP-dependent GAPD from operating in glycolytic direction or to engineer the NAD-dependent GAPD to prefer NADP such that it can generate sufficient amounts of NADP(H) from glycolysis.

Organism-specific and condition-specific CSE targets

Our simulations also unraveled several organism-specific targets, in addition to the global target GAPD. For example, cofactor engineering of pyruvate dehydrogenase (PDH) also significantly improved the NADP(H) turnover than that of wild type in *E. coli* and *B. subtilis*. Similarly, *S. cerevisiae* and *P. pastoris* showed improvements in overall NADP(H) regeneration upon switching the cofactor of malate dehydrogenase (MDH) and alcohol dehydrogenase (ADHE) where the former is suitable for aerobic and the later for anaerobic growth, respectively. Collectively, these observations indicate that these identified enzyme targets could potentially carry high flux irrespective of the carbon source, and thus their cofactor swapping helps improving NADP(H) turnover rates across multiple environmental designs.

Enzyme engineering hotspots for cofactor dependence switch

Once the relevant enzyme targets are identified for cofactor engineering, there are two possible methods to switch the cofactor specificity. First method is to simply replace the native gene which is NAD(H)-specific with a gene that has opposite cofactor. The second technique is to mutate the enzyme by protein engineering techniques such that it prefers the other cofactor than the original one. Among them, altering the cofactor specificity using enzyme engineering approaches is an effective way to circumvent metabolic bottlenecks by increasing the intracellular pool of NADH/NADPH. Therefore, we attempt to suggest possible mutation regions in the amino acid sequence for the CSE targets identified. To do so, first, we identified the determinants for cofactor affinity in these enzymes, and then performed sequence alignments and structural analysis to locate the key loop regions that are crucial for the discrimination between NADH and NADPH. Although the loop regions can be easily identified with multiple sequence alignment within the same enzyme class, they share low sequence identities among different enzyme classes. Therefore, structural information was included, in combination with sequence similarity based method, to identify the key residues in the loop region that determines cofactor specificity.

Four enzymes from those identified in Table 1 have available structural information, and the loop region that interacts with the 2′-OH side of NADH can be readily identified in these structures (Fig. 4a). For the rest without structural data, homology models were built to gain more insights into the cofactor-binding site (Fig. 4b). Apparently, all the enzymes adopt a Rossmann-type fold [41] in their NADH-binding sites, and a highly conserved aspartate or glutamate residue is found to form a hydrogen bond with the 2′-OH of the cofactor, except for *E. coli* ADHE, in which no direct hydrogen bond is spotted. Studies have shown that mutating this residue to smaller residues with a hydroxyl group, such as serine, can compensate for the size change in the binding pocket and the negative charge introduced by the 2′-phosphate group in NADPH, thereby switching the cofactor specificity toward NADPH [4]. Therefore, the conserved glutamate or aspartate in the loop region should be targeted for mutagenesis study, and a replacement by serine may potentially change the cofactor dependence. However, it should be noted that other residues in the loop can also be critical for the binding and orientation of the cofactor to achieve optimal catalytic activity, which, therefore, should also be considered for site saturation mutagenesis to tune the binding affinity toward NADPH.

Discussion

NADP(H) is an essential redox cofactor, very similar to NAD(H), and is specifically required in high quantities for the synthesis of various plant-based natural products

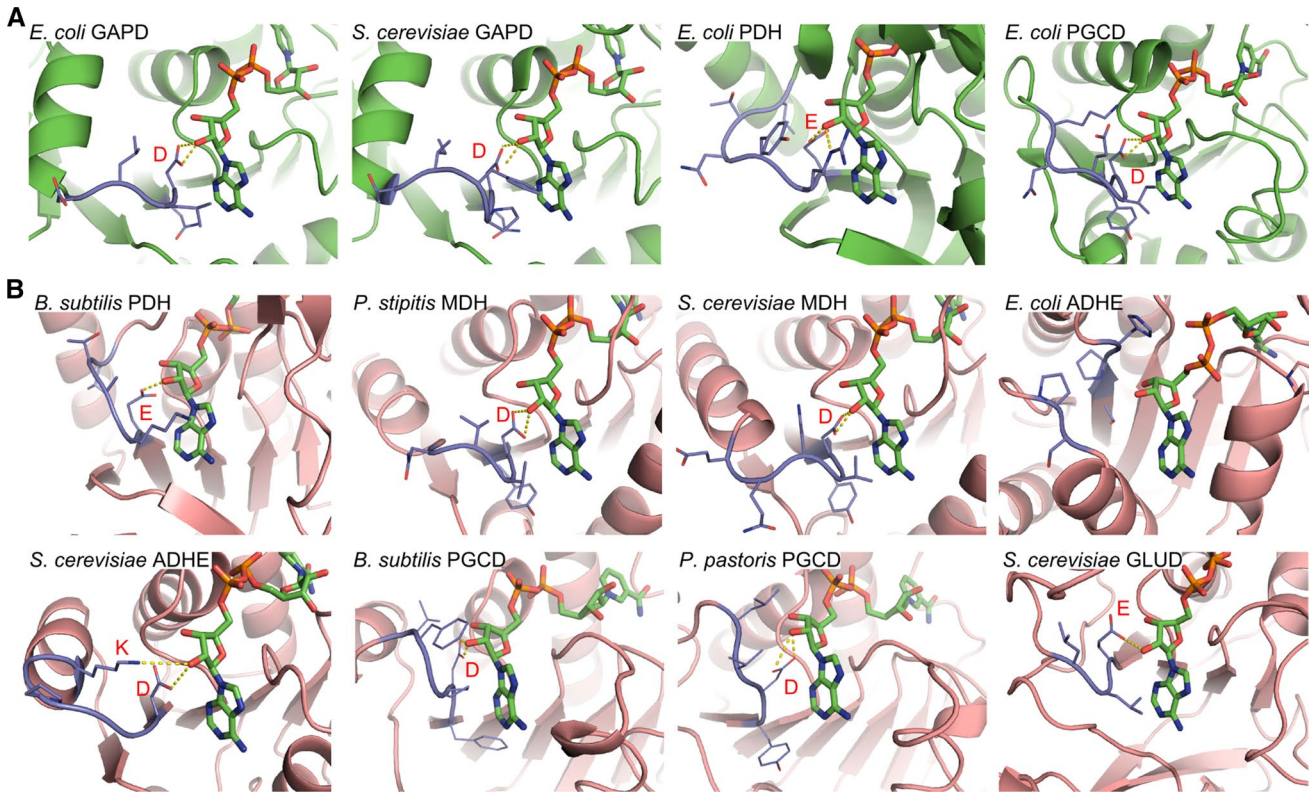


Fig. 4 Close-up views of the NADH-binding sites in enzymes with available crystal structures (**a**) and protein homology models (**b**). The cofactor binding site in homology models are speculated from the geometry and orientation of NADH in structural templates or their close homologs. The proteins are shown in cartoon representation with the loop region colored in slate, while the cofactors are shown in stick representation with the following color scheme: *red* for oxygen, *blue* for nitrogen, *green* for carbon, *orange* for phosphorus. Hydrogen bonds are shown in *dotted lines*

and chiral pharmaceuticals. Although various strategies have been proposed to improve cofactor regeneration, still the overall redox balancing in commonly used industrial microbes and the effects of using different carbon sources to manipulate cofactor turnover rates have not been assessed thoroughly. Therefore, using the CBM technique, in this study, we addressed this issue by first examining the overall NADP(H) and NAD(H) turnover rates in four of the commonly used industrial microbes, *E. coli*, *S. cerevisiae*, *B. subtilis*, and *P. pastoris*. Among them, *E. coli*, with its unique *PntAB* gene, showed remarkably high ability to produce NADPH. All other organisms, in contrast, relied heavily on oxidative PPP fluxes for NADPH biosynthesis, since they lack the *PntAB* gene. *S. cerevisiae* showed a very high NAD(H)/NADP(H) ratio (6.6 in aerobic and 26.6 in anaerobic) while growing on glucose due to negligible fluxes via oxidative PPP. Similarly, a very low NADPH biosynthesis was observed in *B. subtilis* across all carbon sources due to the lack of sufficient NADP-dependent enzymes in the central metabolism. Interestingly, glycerol identified as the best carbon source for achieving a high NADPH regeneration as the initial step of glycerol breakdown in *E. coli*, *P. pastoris*, and *S. cerevisiae* had a NADP-dependent enzyme and thus did not rely much on oxidative PPP for NADPH synthesis. Here, it should be noted that such observations are in very good agreement with the previous studies which showed that glycerol would be a better carbon source than glucose and xylose while producing terpenoids from *E. coli* and *S. cerevisiae* [16]. Furthermore, our simulations also indicated that aerobicity also has a greater NADP(H) production as the NADP(H)/NAD(H) ratios were much lower in anaerobic conditions due to the excessive fluxes via NAD(H)-dependent fermentative pathways.

In this study, we also identified several enzyme targets whose cofactor switching from NAD(H) to NADP(H) or vice versa can increase the NADP(H) turnover significantly. Interestingly, cofactor engineering increased the NADP(H) basal flux-sums by up to 20-folds (Table 1). In order to better understand these observations, we computed the NADP(H) flux-sum maxima of the mutants and observed a significant increase in them as well. Such increase in NADP(H) flux-sum maxima along with the

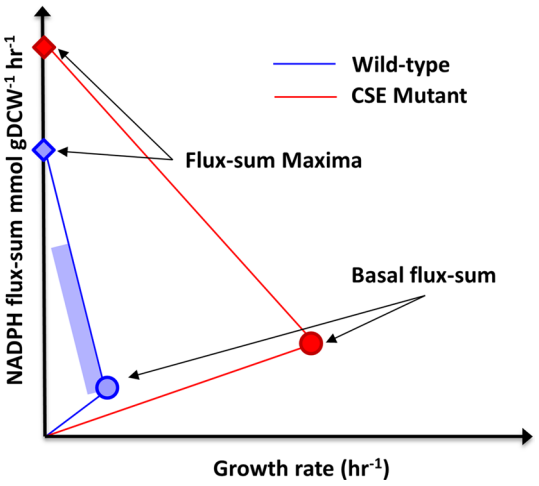


Fig. 5 Comparison of cofactor engineering and gene deletion/overexpression for NADPH improvement. In the deletion/overexpression mutants, the NADPH flux-sum increase along the *shaded region*. Note that these mutants will have growth rate lesser than that of wild type as NADPH is a competitive metabolite. On the other hand, cofactor engineering increases both NADPH synthesis and growth rate due to an overall expansion of flux space

increase in biomass yield indicates that the overall solution space is expanded in the cofactor modified strains. It should be noted that such characteristics of the metabolic solution space emerging from cofactor engineering provide significant advantages than other metabolic engineering strategies, such as gene knockout (KO) or overexpression. As shown in Fig. 5, gene KO or amplification forcefully redirects the carbon flux to NADPH biosynthetic pathways to improve the NADPH basal flux-sum toward its maxima. However, it also should be noted that such improvement in cofactor levels can be obtained only at the expense of biomass as we have earlier identified that NADP(H) is a partially utilized, competitive metabolite from flux-sum analysis. Cofactor engineering, on the other hand, expands the flux solution space with a much higher flux-sum value where the basal NADPH flux-sum of CSE mutants is very closer to the wild-type maxima.

Among several targets identified for CSE, GAPD was the topmost candidate across all organisms and environmental conditions analyzed. It should be noted these results fully support the previous findings of King and Fiest [19], where the authors have showed that the cofactor swapping of this enzyme increased the theoretical yields of several products in *E. coli* and *S. cerevisiae* substantially. Since GAPD is a central enzyme in glycolysis, most of the carbon sources need to be broken down via this enzyme to reach the pyruvate node from which it is distributed to various other pathways involved in biomass and byproduct synthesis. Cofactor engineering of GAPD almost reverses the NAD(H)/NADP(H) ratios in most strains [20]. Therefore, establishing stable strains of GAPD-cofactor-modified microbes is necessary for exploiting them later to produce NADP(H)-dependent products. In this regard, as mentioned earlier, strains bearing NADP(H)-GAPD has been constructed for *E. coli* [33] and *S. cerevisiae* [46], and have already shown a substantial improvement in the NADP(H)-dependent product yields.

As mentioned earlier, altering the cofactor specificity via enzyme engineering is one of the promising techniques as availability of protein sequences and structures allows the identification of key determinants of NAD(P)H binding. A Rossmann fold is commonly found in oxidoreductase protein structures and a conserved βαβαβ structural motif is generally identified as the key for the recognition of these nucleotide-containing cofactors, which interact with surrounding residues via sophisticated non-covalent interactions [41]. The binding sites that distinguish NADH and NADPH are mainly located in a loop region near the 2′ position of the cofactor ribose ring (2′-phosphate in NADPH and 2′-OH in NADH) [27]. Recently, Brinkmann-Chen et al. identified a subset of residues in a loop region of the Rossmann fold in the ketol-acid reductoisomerase (KARI) family, and proposed a general guide for reversing the cofactor preference from NADPH to NADH, which was demonstrated by their successful engineering of three different KARIs [4]. However, this rule may not be generally applicable since the natural sequences in the cofactor binding motifs among different enzyme classes could be hypervariable, which prevents the rational engineering of cofactor dependence. To address this, we combined sequential and structural data and identified potential hotspot residues in the loop regions of the target enzymes that determine the cofactor specificity (Table 2). Mutating these residues may potentially switch the cofactor dependence from NADH to NADPH in a more rational way than using directed evolution approaches. However, enzyme activity can possibly be impaired even if the cofactor preference is altered. In this case, random mutations need be introduced followed by screening or selection for wild-type-like catalysis efficiency with desired cofactor dependence.

Here, we generally assessed the cofactor regeneration rates in wild-type strains which do not produce any non-native products requiring large amounts of NADP(H) for their biosynthesis. However, it is very much possible that the results presented herein for each organism can change significantly depending on the product being produced and corresponding NADP(H) requirements. Therefore, in order to examine the validity of current results, we performed constraint-based flux simulations of cells which produce two different NADP(H)-dependent products:

Table 2 Cofactor-binding motifs of CMA-identified CSE enzyme targets

Enzyme	Source organism	Uniprot Acc. No.	PDB ID	Binding motif[a]
Glyceraldehyde-3-phosphate dehydrogenase (GAPD)	*E. coli*	P0A9B2	1GAD	31 N**D**LLDA
	B. subtilis	P09124	–	33 N**D**LTDA
	P. pastoris	Q92263	–	33 N**D**PFIA
	S. cerevisiae	P00358	–	32 N**D**PFIS
	S. cerevisiae	P00359	3PYM	32 N**D**PFIT
	S. cerevisiae	P00360	–	32 N**D**PFIS
Pyruvate dehydrogenase (PDH)	*E. coli*	P0A9P0	4JDR	36 **E**RYNTL
	B. subtilis	P21880	–	39 **E**KATL
Malate dehydrogenase (MDH)	*P. pastoris*	F2QY33	–	33 Y**D**VVN
	S. cerevisiae	P22133	–	56 Y**D**VNQE
Alcohol dehydrogenase (ADHE)	*E. coli*	P0A9Q7	–	137 SPH; 173 PS
	S. cerevisiae	P00331	–	202 **D**GGPG**K**
Phosphoglycerate dehydrogenase (PGCD)	*E. coli*	P0A9T0	1YBA	180 Y**D**IENK
	B. subtilis	P35136	–	167 F**D**PFL
	P. pastoris	F2QSD3	–	228 Y**D**VL

[a] The conserved residues D or E are in bold. The first residue in the binding motif is numbered according to the protein sequence retrieved from Uniprot

shikimate, a chiral pharmaceutical intermediate which requires one mole of NADPH per mole for its biosynthesis, and lycopene, a natural plant-based food supplement which requires 4 mol of NADPH. The product exchange reaction was maximized during these simulations while simultaneously containing the biomass at 10 % of the wild-type's growth rate. Overall, the results of these additional simulations completely confirmed our earlier findings (1) glycerol as best carbon source, (2) *E. coli* as better expression host, (3) Aerobiosis offer better product synthesis than anaerobic conditions, and (4) GAPD CSE mutants potentially have better product yields than wild type (see Table S6 and S7 in the Supporting Information).

Even though the current study has provided various insights regarding the NADPH balancing in various microbial metabolic networks, several limitations still exist. The current analysis has been performed in a theoretical manner using the stoichiometry accounted in the GEMs. Several model parameters, such as the accuracy of biomass equations, constraints applied during simulations, reaction directionalities, and the cofactor specificity of several enzymes accounted in the model will significantly influence the internal flux distributions, and thus the cofactor turnover rates. For example, our flux analysis indicated a low oxidative PPP flux in *B. subtilis* than *E. coli* whereas earlier experimental reports suggest otherwise [49]. On the other hand, another experimental study showed high intercellular $NADP^+$ concentrations in *E. coli* than *B. subtilis*, despite a low pool of NADPH [14]. Therefore, in order to clarify such ambiguities, the presented method can be further improved by incorporating additional constraints from C^{13} analysis which is not included here due to the differences in base cases and unavailability of data in certain organisms. Nevertheless, despite these limitations, the present work represents the first comprehensive analysis on NADPH balancing in commonly used industrial microbes and demonstrates the advantage of CSE over other metabolic engineering strategies for strain improvement.

Acknowledgments This work was supported by the National University of Singapore, the National Research Foundation of Singapore (NRF2013-THE001-035), Biomedical Research Council of A*STAR (Agency for Science, Technology and Research), Singapore, and a grant from the Next-Generation BioGreen 21 Program (SSAC, No. PJ01109405), Rural Development Administration, Republic of Korea.

Compliance with ethical standards

No conflicts of interests declared. This study does not involve any human or animal subjects.

References

1. Ahn J, Chung BKS, Lee D-Y et al (2011) NADPH-dependent pgi-gene knockout *Escherichia coli* metabolism producing shikimate on different carbon sources. FEMS Microbiol Lett 324:10–16
2. Bastian S, Liu X, Meyerowitz JT et al (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*. Metab Eng 13:345–352
3. Bordbar A, Monk JM, King ZA, Palsson BO (2014) Constraint-based models predict metabolic and associated cellular functions. Nat Rev Genet 15:107–120
4. Brinkmann-Chen S, Flock T, Cahn JKB et al (2013) General approach to reversing ketol-acid reductoisomerase

cofactor dependence from NADPH to NADH. Proc Natl Acad Sci 110:10946–10951
5. Chemler JA, Fowler ZL, McHugh KP, Koffas MAG (2010) Improving NADPH availability for natural product biosynthesis in *Escherichia coli* by metabolic engineering. Metab Eng 12:96–104
6. Chen Y, Nielsen J (2013) Advances in metabolic pathway and strain engineering paving the way for sustainable production of chemical building blocks. Curr Opin Biotechnol 24:965–972
7. Chin JW, Khankal R, Monroe CA et al (2009) Analysis of NADPH supply during xylitol production by engineered *Escherichia coli*. Biotechnol Bioeng 102:209–220
8. Christen S, Sauer U (2011) Intracellular characterization of aerobic glucose metabolism in seven yeast species by ^{13}C flux analysis and metabolomics. FEMS Yeast Res 11:263–272
9. Chung BKS, Lee D-Y (2009) Flux-sum analysis: a metabolite-centric approach for understanding the metabolic network. BMC Syst Biol 3:117
10. Chung BKS, Selvarasu S, Andrea C et al (2010) Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast Pichia pastoris for strain improvement. Microb Cell Fact 9:50
11. Chung BKS, Lakshmanan M, Klement M et al (2013) Genome-scale in silico modeling and analysis for designing synthetic terpenoid-producing microbial cell factories. Chem Eng Sci 103:100–108
12. Durate NC, Palsson BØ, Fu P (2004) Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*. BMC Genom 5:63
13. Feist AM, Palsson BØ (2011) The biomass objective function. Curr Opin Microbiol 13:344–349
14. Fuhrer T, Sauer U (2009) Different biochemical mechanisms ensure network-wide balancing of reducing equivalents in microbial metabolism. J Bacteriol 191:2112–2121
15. Ghosh A, Zhao H, Price ND (2011) Genome-scale consequences of cofactor balancing in engineered pentose utilization pathways in *Saccharomyces cerevisiae*. PLoS One 6:e27316
16. Gruchattka E, Hädicke O, Klamt S et al (2013) In silico profiling of *Escherichia coli* and *Saccharomyces cerevisiae* as terpenoid factories. Microb Cell Fact 12:84
17. Kim P-J, Lee D-Y, Kim TY et al (2007) Metabolite essentiality elucidates robustness of *Escherichia coli* metabolism. Proc Natl Acad Sci 104:13638–13642
18. King ZA, Feist AM (2013) Optimizing cofactor specificity of oxidoreductase enzymes for the generation of microbial production strains—optswap. Ind Biotechnol 9:236–246
19. King ZA, Feist AM (2014) Optimal cofactor swapping can increase the theoretical yield for chemical production in *Escherichia coli* and *Saccharomyces cerevisiae*. Metab Eng 24:117–128
20. Lakshmanan M, Chung BK-S, Liu C et al (2013) Cofactor modification analysis: a computational framework to identify cofactor specificity engineering targets for strain improvement. J Bioinform Comput Biol 11:1343006
21. Lakshmanan M, Koh G, Chung BKS, Lee D-Y (2014) Software applications for flux balance analysis. Brief Bioinform 15:108–122
22. Lee HC, Kim JS, Jang W, Kim SY (2010) High NADPH/NADP+ ratio improves thymidine production by a metabolically engineered *Escherichia coli* strain. J Biotechnol 149:24–32
23. Lee JW, Kim TY, Jang Y-S et al (2011) Systems metabolic engineering for chemicals and materials. Trends Biotechnol 29:370–378
24. Lee WH, Park JB, Park K et al (2007) Enhanced production of ε-caprolactone by overexpression of NADPH-regenerating glucose 6-phosphate dehydrogenase in recombinant Escherichia coli harboring cyclohexanone monooxygenase gene. Appl Microbiol Biotechnol 76:329–338
25. Lee WH, Kim MD, Jin YS, Seo JH (2013) Engineering of NADPH regenerators in *Escherichia coli* for enhanced biotransformation. Appl Microbiol Biotechnol 97:2761–2772
26. Lee SY, Lee D-Y, Kim TY (2005) Systems biotechnology for strain improvement. Trends Biotechnol 23:349–358
27. Lesk AM (1995) NAD-binding domains of dehydrogenases. Curr Opin Struct Biol 5:775–783
28. Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol 10:291–305
29. Li ZJ, Cai L, Wu Q, Chen GQ (2009) Overexpression of NAD kinase in recombinant *Escherichia coli* harboring the *phbCAB* operon improves poly(3-hydroxybutyrate) production. Appl Microbiol Biotechnol 83:939–947
30. Lim S-J, Jung Y-M, Shin H-D, Lee Y-H (2002) Amplification of the NADPH-related genes *zwf* and *gnd* for the oddball biosynthesis of PHB in an *E. coli* transformant harboring a cloned *phbCAB* operon. J Biosci Bioeng 93:543–549
31. Londesborough J, Penttilae M, Richard P, Verho R (2003) Fungal micro-organism having an increased ability to carryout biotechnological process(es). WO2003/038067 A1
32. Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng 5:264–276
33. Martínez I, Zhu J, Lin H et al (2008) Replacing *Escherichia coli* NAD-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with a NADP-dependent enzyme from *Clostridium acetobutylicum* facilitates NADPH dependent pathways. Metab Eng 10:352–359
34. Medema MH, van Raaphorst R, Takano E, Breitling R (2012) Computational tools for the synthetic design of biochemical pathways. Nat Rev Microbiol 10:191–202
35. Mo ML, Palsson BO, Herrgård MJ (2009) Connecting extracellular metabolomic measurements to intracellular flux states in yeast. BMC Syst Biol 3:37
36. Monk J, Nogales J, Palsson BO (2014) Optimizing genome-scale network reconstructions. Nat Biotechnol 32:447–452
37. Oh Y-K, Palsson BO, Park SM et al (2007) Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. J Biol Chem 282:28791–28799
38. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nat Biotechnol 28:245–248
39. Price ND, Famili I, Beard DA, Palsson BØ (2002) Extreme pathways and Kirchhoff's second law. Biophys J 83(5):2879–2882
40. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol 4:R54
41. Rossmann MG, Moras D, Olsen KW (1974) Chemical and biological evolution of nucleotide-binding protein. Nature 250:194–199
42. San K-Y, Bennett GN, Berríos-Rivera SJ et al (2002) Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in *Escherichia coli*. Metab Eng 4:182–192
43. Sauer U, Canonaco F, Heri S et al (2004) The soluble and membrane-bound transhydrogenases *UdhA* and *PntAB* have divergent functions in NADPH metabolism of *Escherichia coli*. J Biol Chem 279:6613–6619
44. Seo JH, Lee WH, Chin YW et al (2011) Enhanced production of GDP-L-fucose by overexpression of NADPH regenerator in recombinant *Escherichia coli*. Appl Microbiol Biotechnol 91:967–976
45. Sievers F, Wilm A, Dineen D et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539

46. Söding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 33(Web Server issue):W244–W248
47. Verho R, Londesborough J, Penttilä M, Richard P (2003) Engineering redox cofactor regeneration for improved pentose fermentation in *Saccharomyces cerevisiae*. Appl Environ Microbiol 69:5892–5897
48. Wang Y, San KY, Bennett GN (2013) Cofactor engineering for advancing chemical biotechnology. Curr Opin Biotechnol 24:994–999
49. Zamboni N, Fischer E, Laudert D et al (2004) The *Bacillus subtilis yqjI* gene encodes the NADP+-dependent 6-P-gluconate dehydrogenase in the pentose phosphate pathway. J Bacteriol 186:4528–4534