**Thermodynamic analysis of the pathway for ethanol production from cellobiose in *Clostridium thermocellum***

Satyakam Dash1 Daniel G. Olson2, Siu Hung Joshua Chan3, and **Costas D. Maranas**1\*

1Department of Chemical Engineering, The Pennsylvania State University, University Park, University Park, Pennsylvania, USA

2Thayer School of Engineering at Dartmouth College, Hanover, New Hampshire, USA

3Department of Chemical and Biological Engineering, Colorado State University, Fort Collins, Colorado, USA

E-mail: Satyakam Dash: sud25@psu.edu, Daniel G. Olson: Daniel.G.Olson@dartmouth.edu, Siu Hung Joshua Chan: joshua.chan@colostate.edu, and Costas D. Maranas: [costas@psu.edu](mailto:costas@psu.edu)

\*Corresponding author:

Costas D. Maranas (costas@psu.edu)

The Pennsylvania State University

126 Land and Water Research Building

University Park, PA 16802

Phone: 814-863-9958

**Abstract**

**Background**

*Clostridium thermocellum* is an ideal candidate for consolidated bioprocessing by carrying out both cellulose degradation and fermentation. However, despite significant efforts the so far achieved ethanol titer has remained below industrially required targets. Several studies have analyzed the impact of increasing ethanol concertation on *C. thermocellum*’s membrane integrity, cofactor pool ratios, and altered enzyme regulation. In this study, we explore whether thermodynamic bottlenecks are responsible for throttling back the ethanol production pathway at high ethanol concentrations.

**Results**

We use the max-min driving force (MDF) algorithm to identify the range of allowable metabolite concentrations maintaining negative free energy of change for all reaction steps in the pathway as a function of ethanol concentration in *C. thermocellum*. In this study, we evaluate the thermodynamic consistency of the metabolomic dataset from the work by Tian et al 2017 [1] and then use it as constraint to demonstrate the thermodynamic favorability of pyruvate phosphate dikinase (PPDK) over malate shunt due to limiting CO2 concentrations and identify glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as the key thermodynamic bottleneck under high external ethanol concentrations. The MDF approach was also deployed in a prospective mode to identify pathway modifications alleviating the thermodynamic bottleneck by analyzing EFMs with combinations of ten genetic interventions which modify the reactions cofactor association.

**Conclusions**

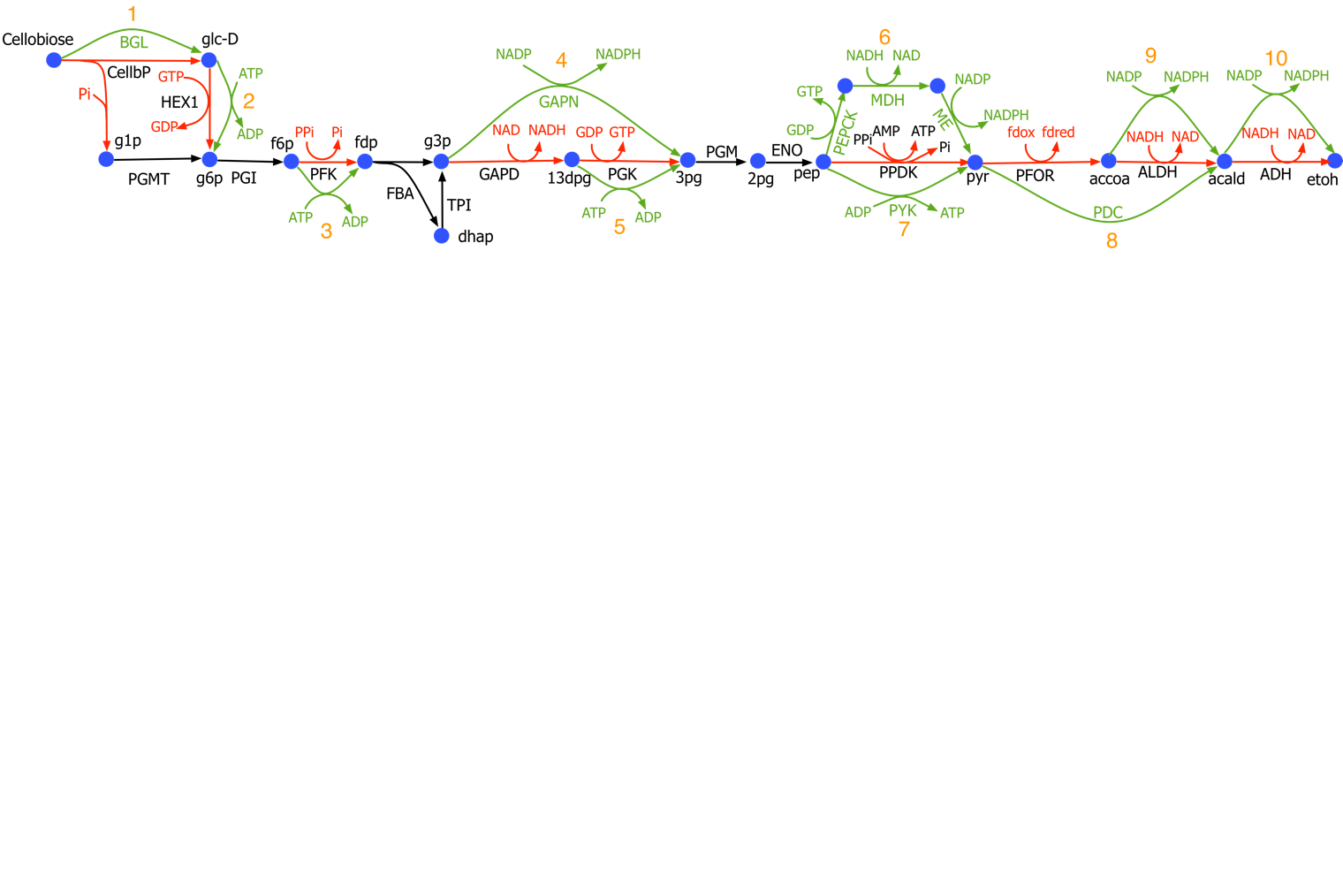
EFMs with NADPH dependent glyceraldehyde-3-phosphate dehydrogenase (GAPN) are able to overcome thermodynamic bottlenecks at high ethanol concentrations and have a high MDF, and those with malate shunt had a significantly low MDF due to limiting CO2 concentrations under physiological conditions. ATP linked phosphofructokinase (PFK-ATP) and NADPH linked alcohol dehydrogenase (ADH-NADPH) emerged as the best genetic intervention which generates ATP while maintaining a high pathway MDF. This study shows that pathway thermodynamics can inform our understanding of metabolic phenotypes and propose testable solutions for restoring pathway thermodynamic feasibility.

**Background**

Biofuels have emerged as a major alternative to reduce fossil fuel emissions in the recent years [2-4]. One of the most industrially relevant biofuel production strategy relies on the microbial breakdown of cellulosic biomass, release of sugars and subsequent fermentation to alcohols [5]. *Clostridium thermocellum* is a promising candidate as it can natively combine the metabolic outcomes of cellulose degradation and fermentation. However, low ethanol titers [6] which have been linked to ethanol tolerance [1] have so far prevented its industrial adoption. Several studies have delved into problem of ethanol tolerance with focus on changes to *C. thermocellum*’s membrane integrity, varying cofactor pool ratios, and feedback regulations which affect enzyme kinetics [7-9]. An alternate approach to this problem as explored in this study involves identifying any limitations on the fundamental driving force of metabolism i.e. thermodynamics.

Thermodynamic constraints have already been integrated in computational modeling techniques such as flux balance calculations to determine possible reaction directionality and predict feasible metabolite concentration ranges [10]. Kinetic parameterization procedures often seek aid of thermodynamic constraints (i.e., negative free energy of change imperatives) to reduce the search space of feasible kinetic parameters and restrict reaction reversibility [11]. Thermodynamic analysis can be extended beyond single reaction to entire pathways. The thermodynamic feasibility of the operation of an entire pathway can be evaluated using the max-min driving force (MDF) formulation [12]. MDF tests whether an assignment of metabolite concentrations is possible ensuring that the free energy of change for every reaction in the pathway remains negative implying a positive thermodynamic driving force. By design, the solution identified by MDF is the one that maximizes the smallest among all reaction steps driving force (i.e., max-min formulation) akin to the optimization of the worst-case scenario [12]. Because metabolite concentrations are treated as optimization variables, the impact of metabolite pool accumulations and/or depletions on pathway feasibility can be directly assessed [12]. MDF has already been applied to study the causes of growth cessation in *C. thermocellum* under high substrate loading conditions. It revealed that hydrogen and formate accumulations are potential culprits rendering acetyl-CoA formation from pyruvate thermodynamically infeasible, thus arresting cell growth [9].

In this study, we use MDF to study the impact of increasing ethanol concertation on *C. thermocellum*’s metabolism. We restrict our analysis to glycolysis along with ethanol production pathway (Figure 1) and cofactor regeneration system. We first evaluate the thermodynamic feasibility of time varying metabolite concentrations for *C. thermocellum* grown with and without external ethanol addition revealing the inconsistency in dihydroxyacetone phosphate (DHAP) and fructose bisphosphate (FDP) measurements. The dataset (excluding DHAP and FDP) is then used to constrain the wild-type pathway thermodynamics of *C. thermocellum* which shows that high ethanol concentration makes glyceraldehyde-3-phosphate dehydrogenase (GAPDH) the prominent bottleneck due to rising NADH levels rendering further ethanol production infeasible. We also observe that in WT *C. thermocellum* pyruvate formation using pyruvate phosphate dikinase (PPDK) has a higher MDF than malate shunt due to low CO2 concentrations which renders phosphoenolpyruvate carboxykinase (PEPCK) thermodynamically infeasible. We then explore ten plausible metabolic interventions by modifying cofactor dependencies (Table 1) of cellobiose phosphorylase (CellbP), gluckokinase (GLK), phosphofrucktokinase (PFK), glyceraldehyde-3-phosphate dehydrogenase (GAPDH), phosphoglycerate kinase (PGK), pyruvate phosphate dikinase (PPDK), aldehyde dehydrogenase (ALDH), and alcohol dehydrogenase (ADH) to ensure higher driving force and resolve any thermodynamic bottleneck. Several of these genetic interventions have already been explored in *C. thermocellum* with reported increase in ethanol titers [13-15] but in this study, we evaluate the underlying thermodynamic driving force and ethanol production feasibility for all these interventions by replacing the native reaction with its genetic variant. These genetic interventions and their combinations thereof are systematically evaluated by generating unique elementary flux modes (EFMs) which possess them. An EFM here represents a minimal set of reactions under steady state conditions with redox and energy balance while ensuring substrate (cellobiose) consumption and product (ethanol) formation [16]. We evaluate MDFs for 336 EFMs and identified genetic variants of glycolysis with malate shunt to have the least MDF due to constraints by physiological CO2 concentrations (similar to the wild-type case) suggesting the close proximity of PEPCK and CO2 generating reactions to ensure malate shunt operability under wild-type conditions. We also observe that replacing GAPDH and PGK with NADP-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPN) significantly improves the pathway MDF (>4) and overcomes thermodynamic hindrance of GAPDH (by replacing NADH with NADPH) under ethanol stress at the expense of ATP generation. Ultimately, the thermodynamic analysis of all possible genetic interventions reveals that changing the cofactor association of PFK reaction to ATP and ADH reaction to NADPH improves the driving force of ethanol production while allowing for ATP generation to sustain cell growth resulting in an efficient ethanol overproducing strain.

****

**Figure 1:** Glycolysis with ethanol production pathway in C. thermocellum: Reactions and cofactors highlighted in black are present in all cases, those highlighted in red are replaced by those highlighted by green in the corresponding mutants, i.e. we replace 1) cellobiose hydydrolase (CellbP) with betaglucosidase (BGL) 2) GTP/GDP with ATP/ADP as cofactors for gluckokinase 3) PPi/Pi with ATP/ADP as cofactors for PFK, 4) glyceraldehyde dehydrogenase (GAPD) and phosphoglycerate kinase (PGK) with glyceraldehyde-3-phosphate dehydrogenase (GAPN), 5) GTP/GDP with ATP/ADP as cofactors for phosphoglycerate kinase (PGK) , 6) PPDK with PEPCK and malate shunt, 7) PPDK with PYK, 8) pyruvate formate oxidoreductase (PFOR) and aldehyde dehydrogenase (ALDH) with pyruvate decarboxylase (PDC), 9) NADH/NAD with NADPH/NADP as cofactors for aldehyde dehydrogenase, and 10) NADH/NAD with NADPH/NADP as cofactors for alcohol dehydrogenase

**Table 1.** Reaction in Wild-type *C. thermocellum* and the corresponding replacements

|  |  |  |  |
| --- | --- | --- | --- |
| Wild-type | | Replaced with | |
| Reaction name | Equation | Reaction name | Equation |
| CBP | pi + cellb <=> glc-D + g1p | BGL | cellb + h2o <=>2 glc-D |
| GLK-GTP | glc-D + gtp <=> g6p + gdp | GLK-ATP | glc-D + atp <=> g6p + adp |
| PFK-PPi | ppi + f6p <=> pi + fdp + h | PFK-ATP | atp + f6p <=> adp + fdp |
| GAPDH | pi + nad + g3p <=> nadh + 13dpg | GAPN | g3p + nadp + h2o <=> 3pg + nadph |
| PGK-GTP | gdp + 13dpg <=> gtp + 3pg |
| PGK-GTP | gdp + 13dpg <=> gtp + 3pg | PGK-ATP | adp + 13dpg <=> atp + 3pg |
| PPDK | amp + ppi + pep <=> atp + pi + pyr + h | PYK | adp + pep <=> pyr + atp |
| PPDK | amp + ppi + pep <=> atp + pi + pyr + h | PEPCK | gdp + co2 + pep <=> gtp + oaa |
| MDH | nadh + oaa <=> nad + mal-L |
| ME | nadp + mal-L <=> nadph + co2 + pyr |
| PFOR | coa + 2 fdxox + pyr <=> h + accoa + 2 fdxrd + co2 | PDC | pyr <=> acald + co2 |
| ALDH-NADH | nadh + accoa <=> nad + coa + acald |
| ALDH-NADH | nadh + accoa <=> nad + coa + acald | ALDH-NADPH | nadph + accoa <=> nadp + coa + acald |
| ADH-NADH | acald + nadh <=> etoh + nad | ADH-NADPH | acald + nadph <=> etoh + nadp |

# Results and discussion

## Metabolite dataset

In this study, we have used metabolomic dataset from the work by Tian et al 2017 [1] . This dataset represents intracellular metabolites collected from wild type (WT) *C. thermocellum* growing with and without (control) the presence of added ethanol for two replicates each at three different timepoints. The ethanol concentration is steadily increased to reach a maximum concentration (i.e. at the final timepoint) of 40 g/L, which has been estimated as the minimum titer for commercial viability [6]. The depletion of extracellular cellobiose pool and increase in fermentation product (such as lactate, acetate, ethanol) pools (Supplementary table FFF), clearly indicates that cellobiose was converted to ethanol during all three timepoints for the no-ethanol control and for the first two timepoints for the ethanol added samples implying thermodynamic feasibility of ethanol production. However, for the case of samples with added ethanol, we observe reduced consumption of cellobiose and the lack of increase in ethanol pool for the final timepoint indicating that the system is thermodynamically infeasible for ethanol production. Thus, we evaluate the thermodynamic feasibility of the measured metabolite concentrations (excluding cofactors) by imposing the measured metabolite levels as constraints one at a time on our WT pathway and calculating the pathway MDF and checking for thermodynamic feasibility (i.e. positive MDF) for all but the final timepoint of ethanol added case. For each measured metabolite, we include an uncertainty factor of 1.5 (see materials and methods for description) based on the variability across sample replicates. The results indicate that DHAP and FDP measurements are thermodynamically inconsistent resulting in negative MDF for all timepoints (Supplementary table GGG). DHAP is difficult to measure by LC/MS because it has the same M/z ratio as G3P, and the two metabolites are interconverted by the TPI reaction, potentially allowing for rapid equilibration during quenching. Because of these uncertainties, we suspect the measurement error for DHAP may be higher. The set of reactions that allow for substrate-level phosphorylation (SLP) in EMP glycolysis (FBA, TPI, GAPDH and PGK) are generally known to be thermodynamic bottlenecks [12, 17]. In *E. coli*, this pathway is feasible due to the high (>10 mM) intracellular concentration of FDP [18], however in *C. thermocellum*, the intracellular concentration of FDP is 2-3 orders of magnitude lower (10-100 µM), suggesting that glycolysis may be able to function in *C. thermocellum* despite much lower levels of intracellular FDP than are commonly observed in model organisms. We anticipate that future studies of intracellular metabolite concentrations in *C. thermocellum* may provide insight into this question by accurately measuring both the FDP concentration and NAD/NADH ratio.

**Glycolysis in *C. thermocellum***

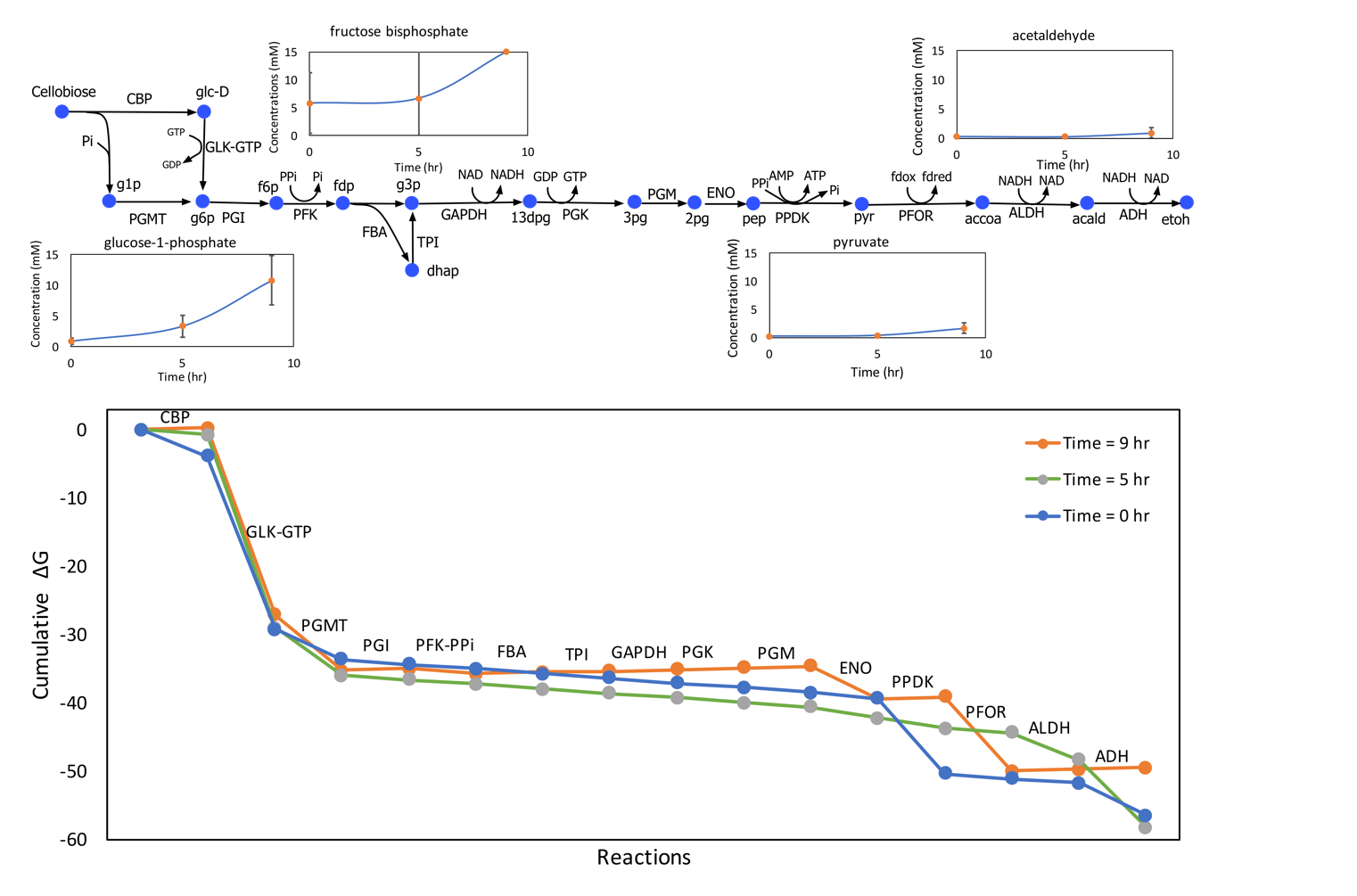
The time-varying metabolomic dataset (excluding DHAP and FDP) as described in the previous section with the energy cofactors (ATP/ADP, GTP/GDP, and PPi/Pi) allowed a ratio greater than 10:1 [19] and the redox cofactors (NADH/NAD+, NADPH/NADP+, Fd(red)/Fd(ox)) ratios varying between 1:100 and 100:1 [19] were imposed as constraint to evaluate the wild type *C. thermocellum* pathway MDF at all the time points in presence and absence of externally added ethanol as shown in Table 2. The analysis shows g6p and f6p measurements constrain the pathway thermodynamics (Supplementary table XX) but the overall ethanol production pathway is feasible for all cases except the final timepoint of ethanol addition. For the final timepoint of ethanol addition, increase in ethanol concertation causes NADH accumulation which causes the reactions associated with NADH such as GAPDH, ALDH and ADH to have lower driving force with GAPDH emerging as the prominent thermodynamic bottleneck (i.e. change in thermodynamics of GAPDH will have maximum impact on pathway MDF) which is consistent with experimental observations [1].

**Table 2.** Pathway MDF of wild-type *C. thermocellum* for all samples and time points with and without added ethanol. Negative MDF values indicate thermodynamic infeasibility at the final timepoints of ethanol addition

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Time point (hr) | Externally added ethanol | | No added ethanol | |
| Sample 1 | Sample 2 | Sample 3 | Sample 4 |
| 0 | 1.08 | 0.69 | 0.78 | 0.50 |
| 5 | 0.03 | 0.66 | 1.17 | 0.85 |
| 9 | -0.22 | -0.23 | 1.05 | 0.70 |

## PPDK vs malate shunt

*C. thermocellum* does not possess the conventional pyruvate kinase (PYK) enzyme, instead it uses pyruvate phosphate dikinase (PPDK) and phosphoenolpyruvate carboxykinase (PEPCK) with the malate shunt as two alternate pathways to generate pyruvate from phosphoenolpyruvate (see Figure 1). The metabolic flux allocation between these two pathways is essential to characterize wild-type metabolism and predict the impact of genetic perturbations on metabolism using kinetic models. Several studies have demonstrated that the flux ratio between PPDK and malate shunt is tilted heavily towards PPDK [8, 13] suggesting the usefulness of malate shunt only for its transhydrogenase activity for anabolic processes [20]. In this study, we evaluate the driving force for the two pathways using wild-type measurements under varying ethanol concentrations. The results (shown in Figure 2) clearly indicate that the thermodynamic driving force is consistently positive for PPDK for ethanol concentrations up to 0.8M in contrast to the malate shunt which remains thermodynamically infeasible for the entire range of ethanol concentrations (0-1M). This is because the malate shunt requires a high intracellular CO2 concentration (>0.1M) to ensure thermodynamic feasibility, but the CO2 concentration bounds are restricted (<0.01mM) under on standard experimental conditions [12]. This alludes to a localized increase of CO2 concentration in WT *C. thermocellum* for PEPCK to be feasible and also indicates close proximity of PEPCK to CO2 generating reactions such as pyruvate: ferredoxin oxidoreductase (PFOR), malate dehydrogenase (MDH) which should be probed experimentally. We also observe accumulation of upper glycolysis metabolites especially sugar phosphates (also observed in ethanol stress studies [1, 21]) as a consequence of GAPDH being a thermodynamic bottleneck due to an increase in NADH pool under ethanol stress (Figure 2).



**Figure 2:** Glycolysis with ethanol production pathway in wild-type C. thermocellum with PPDK showing the variation in metabolite concentrations of key metabolites across time points. The external ethanol concentration increases with increasing time points. Sugar phosphates (e.g. G1P) in upper glycolysis are getting accumulated under ethanol stress.

## Analysis of possible metabolic engineering interventions in *C. thermocellum*

MDF analysis of wild-type *C. thermocellum* in the previous section showed that cofactor pools (especially NADH) plays a major role in driving the thermodynamic feasibility of ethanol production in *C. thermocellum*. Thus, we can infer that perturbations in cofactor pools can be used to our advantage to increase ethanol production by ensuring a larger MDF for the pathway. To this end, we have systematically evaluated pathway modification strategies by genetic interventions which modify the cofactor association of a single or multiple reactions in *C. thermocellum* including: beta-glucosidase (BGL), ATP-linked phosphofructokinase (ATP-PFK), NADP-dependent glyceraldehyde 3-phosphate dehydrogenase (GAPN), pyruvate decarboxylase (PDC), NADPH-linked ALDH (ALDH-NADPH), and NADPH-linked ADH (ADH-NADPH) (see Figure 1 and Table 1). A genetic intervention often implies multiple changes. For example, changing the cofactor preference of the ADH reaction from NADH to NADPH implies a second change somewhere in the pathway to provide a source of NADPH. To ensure that all of these secondary changes were accounted for, we generate elementary flux modes (EFMs) for all combinations of reactions that allow conversion of one molecule of cellobiose into four molecules of ethanol while allowing for only unique biotransformations (i.e. prevent reactions which differ only in cofactors to appear in the same EFM). For each of the resulting 336 EFMs (Supplementary Table XX), we calculate the ATP generated and the pathway MDF. We quantify the difference between our wild-type EFM and any mutant EFM by calculating their Hamming Distance which measures the number of reaction changes in a mutant compared to our wild-type EFM. A high hamming distance indicates that significant strain design effort would be invested in generating the mutant strain. All the 336 EFMs are analyzed at 1M ethanol concentration while allowing other metabolite concentrations to vary within physiological range (1 μM - 0.02 mM), the energy cofactors (ATP/ADP, GTP/GDP, and PPi/Pi) were allowed to assume a ratio greater than 10:1 and the redox cofactors (NADH/NAD+, NADPH/NADP+, Fd(red)/Fd(ox)) ratios were allowed to vary between 1:100 and 100:1. Figure 3 summarizes the MDF values, ATP generation, and hamming distance of all the EFMs.

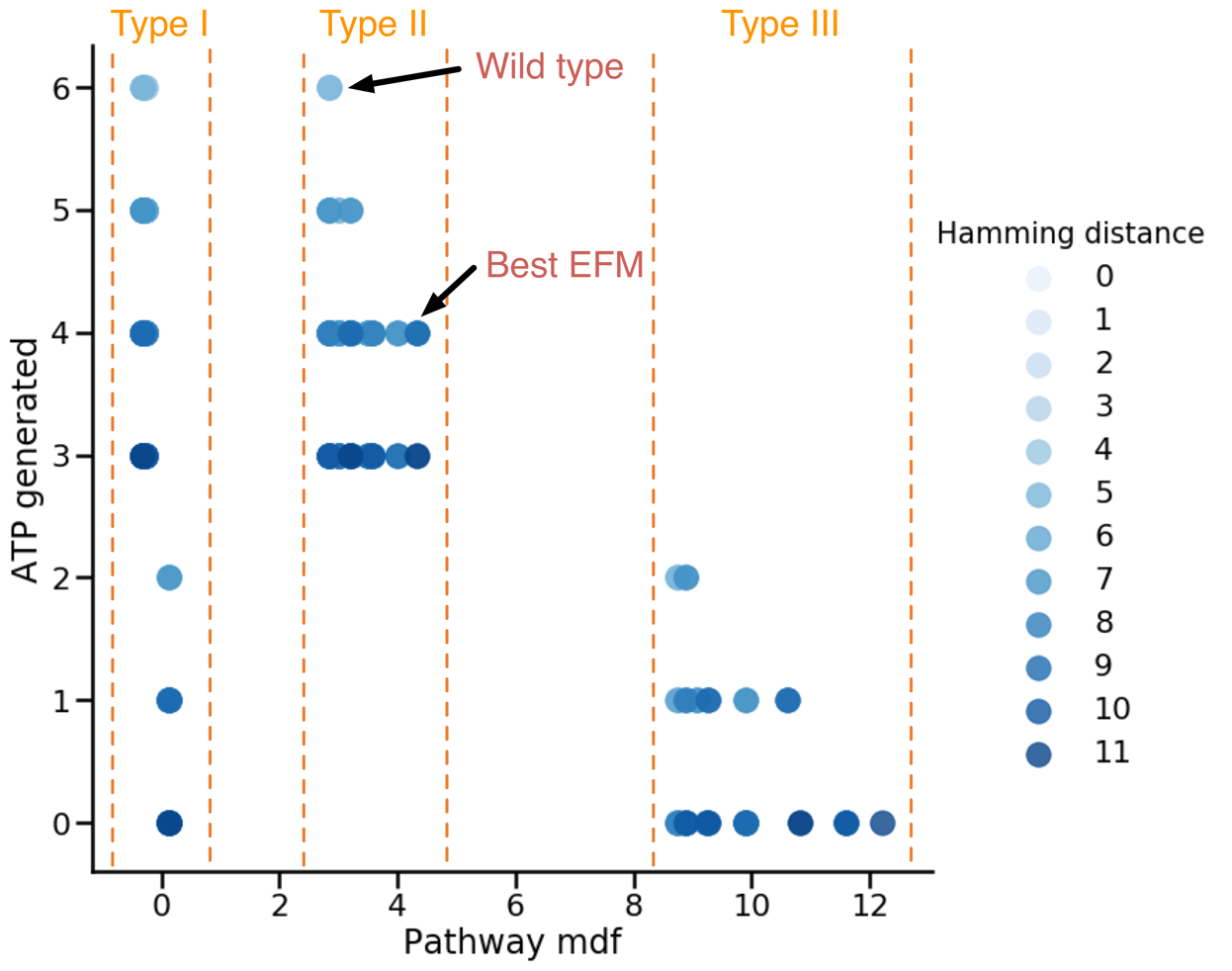
****

Figure 3: MDF at 1M ethanol concentration, ATP generated, and Hamming distance of all 336 EFMs. The EFMs are generated using all possible combinations of cofactor modifications as list in Table 1. The shade of blue represents the value of hamming distance of any given EFM calculated from the wild-type EFM. The EFMs are clustered in three different regions based on their MDF values.

There are three distinct clusters of EFMs which can be seen in Figure 3 based on pathway MDFs, i.e. type I EFMs with low MDF (<1), type II EFMs with intermediate MDF (3-5) but high ATP generation (>2), and type III EFMs with high MDF (>8) but low ATP generation (<3). It is interesting to note that all the type I EFMs possessed the malate shunt to generate pyruvate which led to thermodynamic infeasibility except those EFMs which additionally contain the GAPN reaction. The physiologically allowable CO2 concertation restricted the PEPCK reaction lowering the MDF for all these EFMs close to zero indicating poor thermodynamic feasibility similar to the case with WT *C. thermocellum* as discussed in the previous section. All the type III EFMs possessed the GAPN reaction which replaces GAPDH and PGK reactions from wild-type *C. thermocellum*. Studies have shown that such a genetic intervention can improve ethanol yield in yeast and lysine yield in *Corynebacterium glutamicum* [22, 23]. However, introduction of GAPN also reduces the ATP generated by the pathway which is necessary for cell growth. The EFM with highest MDF completely removed NADH association with the ethanol production by using PDC and NADPH linked ADH. PDC replaces PFOR and ALDH-NADH and hence rids the pathway of redox regeneration systems which entail generation of NADPH or NADH from reduced ferredoxin. The driving force advantage gained by GAPN and PDC can be also understood by the fact that they reduce either the energy generation (GTP in case of PGK) or the redox generation (reduced ferredoxin by PFOR) of the native pathway which negatively impacts pathway MDF. It is noteworthy that studies have shown a higher NADPH/NADP ratio compared to NADH/NAD ratio in *C. thermocellum* and thus NADPH linked reactions would tend to have a stronger driving force in engineered strains [24].

The most useful genetic interventions are part of the type II EFMs which retain ATP generation along with intermittent pathway MDF as shown in Figure 3. All the high MDF type II EFMs possess the ATP linked PFK reaction and the NADPH linked ADH reaction. The best EFM has a minimum hamming distance of 2 indicating PFK-ATP and ADH-NADPH as the only changes from native pathway. The ATP-PFK provides a stronger driving force at the cost of ATP generation (PPi is equivalent to 0.5 ATP [25]) while the NADPH linked ADH reaction decouples the impact of rise in ethanol concentrations from NADH associated native reactions. Studies have also shown that the cofactor specificity of alcohol dehydrogenase changes to NADPH from NADH in high ethanol yielding strains of *C. thermocellum* [26]. We can thus conclude that ATP-PFK and NADPH-ADH are the most beneficial genetic intervention which can help us achieve high ethanol titers while sustaining cell growth.

**Conclusions**

This study analyzes wild-type *C. thermocellum* glycolysis using experimentally measured concertation datasets to understand the impact of increasing ethanol concertation. GAPDH was revealed as the major thermodynamic bottleneck at high ethanol concentrations which renders ethanol production infeasible. EFMs of plausible genetic interventions showed the infeasibility of malate shunt under physiological CO2 concentrations and the benefits of having the GAPN reaction which generates high MDF by resolving the GAPDH bottleneck but with reduced ATP generation at high ethanol concentrations. The best genetic intervention which retains ATP generation with a high driving force is the combination of ATP linked PFK and NADPH linked ADH reaction which has also been shown to occur in high ethanol yielding *C. thermocellum* strains [26].

Conventional studies have explored high yield target phenotypes by manipulating networks to improve desired fluxes. This study shows that a fundamental analysis of pathway thermodynamics can augment our arguments for a given genetic perturbation by providing a mechanistic detail of the mutant phenotype and help us weed out thermodynamically infeasible designs in our search for overproducing strains. However, we should be careful about the simplified assumptions which were essential for this study but would play vital role in real world scenarios. Cellulose degradation has not been considered in this study, but there are several challenges associated with this step including enzyme kinetics, cellulosome availability, and most of all recalcitrance. Enzyme kinetics are also essential to accurately estimate the impact of changes in metabolite and regulator pools on reaction fluxes through allosteric and feedback regulations. In this study, we also assumed the intracellular concentration of ethanol to be the same as the extracellular concentration, since the cytoplasmic membrane is not thought to be a barrier to ethanol diffusion [27], however, this assumption may not hold true for other potential biofuels such as butanol and would require intracellular measurements for accurate quantification. Despite all these simplifications, our results show that thermodynamic pathway analysis remains a vital tool to prune potential strain design strategies.

# Materials and methods

## Metabolite quantification (Texts from Dan)

A single 200 ml culture was grown to an OD600 of 0.1, the culture was split in half. Starting at 2 hours (post-split), ethanol was added to one culture at a rate of about 9 g/L/h. No ethanol was added to the other culture. At three timepoints (T=2.0h, 3.8h and 5.9h), each culture was sampled twice for intracellular metabolites using previously described protocols [1, 24, 28], and once for extracellular metabolites (supplemental table BBB). The raw data was re-processed with El-Maven 0.5.0 and quantified using external standards (supplemental figure AAA). Since the response was not linear over the full range of the standards (0.1 µM to 100 µM), quantification was performed by piecewise linear interpolation (supplemental figure DDD). Based on four measurements of our standard curves, we used an uncertainty factor of 1.5 during our simulations. One significant change that we observed from the absolute quantification is that the energy charge (ATP or GTP, calculated using Equation (1)) values which had previously been calculated to be around 0.9, were actually much lower, in the range of 0 to 0.15, which is much lower than expected [28], and thus datasets with 0 energy charge were excluded from analysis.

(1)

In addition, NADH was not observed for the no ethanol control cultures. Since NADH and NAD+ form a cofactor pair, both metabolites were excluded from analysis. Since metabolites are typically diluted during the quenching and extraction process, to determine the *intracellular* concentration of metabolites, the measured concentration was adjusted using Equation (2).

(2)

The intracellular volume is assumed to be 3.9 µL for 1 ml of a culture at an OD600 density of 1 [29]. The concentration factor was calculated separately for each extraction, based on the measured OD600 and sample volume. The target cell density (OD600 x ml) was 2.0, but varied from 1.6 to 2.1 (supplemental table CCC). For ethanol, the intracellular concentration was assumed to be the same as the extracellular concentration, since the cytoplasmic membrane is not thought to be a barrier to ethanol diffusion [27].

## Assessing the thermodynamic feasibility of a pathway

The thermodynamic feasibility of a given pathway is assessed using the max-min driving force (MDF) formulation [12]. The MDF formulation identifies a set of metabolite concentrations that ensure the lowest free energy changes for all the reactions in a pathway. The MDF problem minimizes the maximum of a pathway by optimizing over the concentrations of all metabolites in the pathway. The formulation is given by:

|  |  |  |
| --- | --- | --- |
|  | (-MDF) | (3) |
|  |  | (4) |
|  |  | (5) |
|  |  |  |

where *I*is the set of all metabolites and *J*is the set of all reactions in a given pathway, is the concentration of metabolite , is the gas constant, is the temperature and the matrix refers to the stoichiometric matrix of the pathway with . Constraint (4) relates the Gibbs free energy of reaction () with the standard Gibbs free energy of reaction () and the mass action ratio. The pathway with a positive objective function (i.e. negative MDF) indicates that it is thermodynamically infeasible within the given physiological concentration (and ratio) ranges. The MDF problem is solved using Gurobi Optimizer v6.5.1 solver and Python script modified from the Equilibrator-API Python package [30].

For performing the max-min driving-force (MDF) analysis, maximum and minimum metabolite concentration have to be established for non-measured metabolites. Noor et al [12] proposed a range of 1 µM to 10 mM, based largely on the work of Bennet et al [18]. Based on our measurements, we decided to keep the lower default concentration at 1 µM, but raise the upper default concentration from 10 mM to 20 mM. Although Noor et al fixed the ratios of several cofactor pairs, we have relaxed these constraints, since the values have not been experimentally determined for *C. thermocellum*. Several cofactors were excluded from analysis due to measurement problems (see Metabolite quantification section). To avoid biasing the results by including only a few measured cofactors, we opted to ignore measured values for all cofactors (ATP, ADP, AMP, GTP, GDP, NAD+. NADH, NADP+ and NADPH). Concentration ranges for energy cofactors (ATP/ADP, ATP/AMP, GTP/GDP) were set to allow ratios greater than 10:1 and redox cofactors (NADH/NAD+, NADPH/NADP+, Fd(red)/Fd(ox)) to vary between 1:100 and 100:1. Default bounds for each metabolite is described in Supplementary table EEE.

## EFM evaluation

We implement the algorithm to generate k-shortest EFM of a network to systematically evaluate all the EFMs associated with our network [31]. The formulation is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |
|  |  | (7) |
|  |  | (8) |
|  |  | (9) |
|  |  | (10) |
|  |  | (11) |
|  |  | (12) |

where *I* is the set of all metabolites and *J* is the set of all reactions, is the set of exchange reactions, the rate of reactions is represented by and represents the stoichiometric coefficient of the metabolite *i* in reaction *j* reactions in a given pathway, *K* represents the set of previously found solutions and indicates the optimal value of the binary variable in the *kth* solution . The binary variable (equation 10) assumes a value of 1 if the reaction *j* is included in the EFM and 0 otherwise. The smallest EFM is found by minimizing the sum of all these binary variables as shown by constraint 6. Steady state of the network is ensured by enforcing no net accumulation or consumption of metabolites as shown by constraint 7. The network is decomposed to ensure that all reaction fluxes are positive (constraint 11) and the binary variable was linked to reaction rates by constraint 8. The binary variables corresponding to exchange reactions () are fixed by constraint 9 to ensure that all EFMs consume the substrate (cellobiose) and generate the product (ethanol). We finally use integer cuts (constraint 12) to generate all possible EFMs associated with the network. The hamming distance between the wild-type and any given EFM is calculated by calculating the number of wild-type reaction which are absent in the mutant EFM. The ATP generated by any given EFM is calculated by adding the fluxes of the two ATP hydrolysis reaction in the models (Table 3).

Table 3: List of all possible reactions in any given EFM.

|  |  |
| --- | --- |
| **Reaction Name** | **Equation** |
| ATPase1 | h2o + atp <=> adp + pi |
| CBP | pi + cellb <=> glc-D + g1p |
| BGL | cellb + h2o <=>2 glc-D |
| GLK-GTP | glc-D + gtp <=> g6p + gdp |
| GLK-ATP | glc-D + atp <=> g6p + adp |
| PGMT | g1p <=> g6p |
| PGI | g6p <=> f6p |
| PFK-PPi | ppi + f6p <=> pi + fdp + h |
| PFK-ATP | atp + f6p <=> adp + fdp |
| FBA | fdp <=> dhap + g3p |
| TPI | dhap <=> g3p |
| GAPDH | pi + nad + g3p <=> nadh + 13dpg |
| GAPN | g3p + nadp + h2o <=> 3pg + nadph |
| PGK-ATP | adp + 13dpg <=> atp + 3pg |
| PGK-GTP | gdp + 13dpg <=> gtp + 3pg |
| PGM | 3pg <=> 2pg |
| ENO | 2pg <=> pep + h2o |
| PYK | adp + pep <=> pyr + atp |
| PPDK | amp + ppi + pep <=> atp + pi + pyr + h |
| PEPCK | gdp + co2 + pep <=> gtp + oaa |
| MDH | nadh + oaa <=> nad + mal-L |
| ME | nadp + mal-L <=> nadph + co2 + pyr |
| PFOR | coa + 2 fdxox + pyr <=> h + accoa + 2 fdxrd + co2 |
| RNF\_ATPase | 2 fdxrd + 1 nad + 0.25 adp + 0.25 pi <=> 2 fdxox + 1 nadh + 0.25 atp + 0.25 h2o |
| RNF\_PPiase | 2 fdxrd + 1 nad + 1 h + 1 pi <=> 2 fdxox + 1 nadh + 0.5 ppi + 0.5 h2o |
| NFN | 2 fdxrd + nadh + 2 nadp = 2 fdxox + nad + 2 nadph |
| PDC | pyr <=> acald + co2 |
| ALDH-NADH | nadh + accoa <=> nad + coa + acald |
| ALDH-NADPH | nadph + accoa <=> nadp + coa + acald |
| ADH-NADH | acald + nadh <=> etoh + nad |
| ADH-NADPH | acald + nadph <=> etoh + nadp |
| Bifur\_Hyd | 2 fdxrd + nadh + 2 h = nad + 2 h2 + 2 fdxox |
| NDK | gtp + adp <=> atp + gdp |
| Gly-cyc | atp + pi <=> adp + ppi |
| NADPH-FNOR | 2 fdxrd + nadp <=> 2 fdxox + nadph |
| NADH-FNOR | 2 fdxrd + nad <=> 2 fdxox + nadh |
| ATPase2 | atp + h2o <=> amp + ppi |

**References**

1. Tian L, Perot SJ, Stevenson D, Jacobson T, Lanahan AA, Amador-Noguez D, Olson DG, Lynd LR: **Metabolome analysis reveals a role for glyceraldehyde 3-phosphate dehydrogenase in the inhibition of C. thermocellum by ethanol**. *Biotechnol Biofuels* 2017, **10**:276.

2. Singh AP, Agarwal AK, Agarwal RA, Dhar A, Shukla MK: **Introduction of Alternative Fuels**. *Energy Env Sustain* 2018:3-6.

3. Azad AK, Rasul MG, Khan MMK, Sharma SC, Hazrat MA: **Prospect of biofuels as an alternative transport fuel in Australia**. *Renew Sust Energ Rev* 2015, **43**:331-351.

4. Tzimas E, Soria A, Peteves S: **The introduction of alternative fuels in the European transport sector: techno-economic barriers and perspectives**. *JRC Petten and JRC Seville EUR* 2004, **21173**.

5. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS: **Microbial cellulose utilization: fundamentals and biotechnology**. *Microbiol Mol Biol Rev* 2002, **66**(3):506-577, table of contents.

6. Dien BS, Cotta MA, Jeffries TW: **Bacteria engineered for fuel ethanol production: current status**. *Appl Microbiol Biotechnol* 2003, **63**(3):258-266.

7. Demain AL, Newcomb M, Wu JH: **Cellulase, clostridia, and ethanol**. *Microbiol Mol Biol Rev* 2005, **69**(1):124-154.

8. Woodruff LB, Boyle NR, Gill RT: **Engineering improved ethanol production in Escherichia coli with a genome-wide approach**. *Metab Eng* 2013, **17**:1-11.

9. Thompson RA, Trinh CT: **Overflow metabolism and growth cessation in Clostridium thermocellum DSM1313 during high cellulose loading fermentations**. *Biotechnology and bioengineering* 2017, **114**(11):2592-2604.

10. Ataman M, Hatzimanikatis V: **Heading in the right direction: thermodynamics-based network analysis and pathway engineering**. *Curr Opin Biotechnol* 2015, **36**:176-182.

11. Dash S, Khodayari A, Zhou J, Holwerda EK, Olson DG, Lynd LR, Maranas CD: **Development of a core Clostridium thermocellum kinetic metabolic model consistent with multiple genetic perturbations**. *Biotechnol Biofuels* 2017, **10**:108.

12. Noor E, Bar-Even A, Flamholz A, Reznik E, Liebermeister W, Milo R: **Pathway thermodynamics highlights kinetic obstacles in central metabolism**. *PLoS Comput Biol* 2014, **10**(2):e1003483.

13. Tian L, Perot SJ, Hon S, Zhou J, Liang X, Bouvier JT, Guss AM, Olson DG, Lynd LR: **Enhanced ethanol formation by Clostridium thermocellum via pyruvate decarboxylase**. *Microb Cell Fact* 2017, **16**(1):171.

14. Olson DG, Sparling R, Lynd LR: **Ethanol production by engineered thermophiles**. *Curr Opin Biotechnol* 2015, **33**:130-141.

15. Lo J, Zheng T, Hon S, Olson DG, Lynd LR: **The bifunctional alcohol and aldehyde dehydrogenase gene, adhE, is necessary for ethanol production in Clostridium thermocellum and Thermoanaerobacterium saccharolyticum**. *J Bacteriol* 2015, **197**(8):1386-1393.

16. Klamt S, Regensburger G, Gerstl MP, Jungreuthmayer C, Schuster S, Mahadevan R, Zanghellini J, Muller S: **From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints**. *Plos Computational Biology* 2017, **13**(4).

17. Stephanopoulos G: **Metabolic engineering**. *Biotechnology and bioengineering* 1998, **58**(2-3):119-120.

18. Bennett GN, San KY: **Engineering E. coli Central Metabolism for Enhanced Primary Metabolite Production**. *Systems Biology and Biotechnology of Escherichia Coli* 2009:351-376.

19. Milo R, Jorgensen P, Moran U, Weber G, Springer M: **BioNumbers--the database of key numbers in molecular and cell biology**. *Nucleic Acids Res* 2010, **38**(Database issue):D750-753.

20. Olson DG, Horl M, Fuhrer T, Cui J, Zhou J, Maloney MI, Amador-Noguez D, Tian L, Sauer U, Lynd LR: **Glycolysis without pyruvate kinase in Clostridium thermocellum**. *Metab Eng* 2017, **39**:169-180.

21. Yang S, Giannone RJ, Dice L, Yang ZK, Engle NL, Tschaplinski TJ, Hettich RL, Brown SD: **Clostridium thermocellum ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress**. *BMC Genomics* 2012, **13**:336.

22. Guo ZP, Zhang L, Ding ZY, Wang ZX, Shi GY: **Improving ethanol productivity by modification of glycolytic redox factor generation in glycerol-3-phosphate dehydrogenase mutants of an industrial ethanol yeast**. *J Ind Microbiol Biotechnol* 2011, **38**(8):935-943.

23. Takeno S, Hori K, Ohtani S, Mimura A, Mitsuhashi S, Ikeda M: **L-Lysine production independent of the oxidative pentose phosphate pathway by Corynebacterium glutamicum with the Streptococcus mutans gapN gene**. *Metab Eng* 2016, **37**:1-10.

24. Beri D, Olson DG, Holwerda EK, Lynd LR: **Nicotinamide cofactor ratios in engineered strains of Clostridium thermocellum and Thermoanaerobacterium saccharolyticum**. *Fems Microbiol Lett* 2016, **363**(11).

25. Zhou J, Olson DG, Argyros DA, Deng Y, van Gulik WM, van Dijken JP, Lynd LR: **Atypical glycolysis in Clostridium thermocellum**. *Appl Environ Microbiol* 2013, **79**(9):3000-3008.

26. Zheng TY, Olson DG, Tian L, Bomble YJ, Himmel ME, Lo J, Hon S, Shaw AJ, van Dijken JP, Lynd LR: **Cofactor Specificity of the Bifunctional Alcohol and Aldehyde Dehydrogenase (AdhE) in Wild-Type and Mutant Clostridium thermocellum and Thermoanaerobacterium saccharolyticum**. *Journal of Bacteriology* 2015, **197**(15):2610-2619.

27. Shinoda W: **Permeability across lipid membranes**. *Bba-Biomembranes* 2016, **1858**(10):2254-2265.

28. Rabinowitz JD, Kimball E: **Acidic acetonitrile for cellular metabolome extraction from Escherichia coli**. *Analytical Chemistry* 2007, **79**(16):6167-6173.

29. Volkmer B, Heinemann M: **Condition-Dependent Cell Volume and Concentration of Escherichia coli to Facilitate Data Conversion for Systems Biology Modeling**. *Plos One* 2011, **6**(7).

30. Noor E, Haraldsdottir HS, Milo R, Fleming RM: **Consistent estimation of Gibbs energy using component contributions**. *PLoS computational biology* 2013, **9**(7):e1003098.

31. de Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ: **Computing the shortest elementary flux modes in genome-scale metabolic networks**. *Bioinformatics* 2009, **25**(23):3158-3165.